Understanding and Improving Human Data Relations



Alex Bowyer

Supervisors: Prof. David Kirk

Dr. Josephine Go Jefferies

Dr.-Ing. Jan Smeddinck

Prof. Rob Wilson, PhD

School of Computing

Newcastle University

This thesis is submitted for the degree of

Doctor of Philosophy

June 2023

Table of Contents

TABLE OF CONTENTS ..... **II**

ABSTRACT ..... **XII**

DEDICATION ..... **XIII**

IN MEMORIAM ..... **XIV**

ACKNOWLEDGEMENTS ..... **XV**

LISTS OF TABLES AND FIGURES ..... **XVIII**

Lists of Tables by Chapter ..... **xviii**

Tables in Chapter 3 xviii

Tables in Chapter 4 xviii

Tables in Chapter 5 xviii

Tables in Chapter 7 xix

Tables in 'Additional Reference Information' xix

Lists of Figures by Chapter ..... **xix**

Figures in Chapter 1 xix

Figures in Chapter 2 xix

Figures in Chapter 3 xix

Figures in Chapter 4 xx

Figures in Chapter 5 xx

Figures in Chapter 7 xx

Figures in Section IV xx

Figures in Chapter 8 xx

Figures in Chapter 9 xxi

Figures in Chapter 10 xxii

Figures in 'Additional Reference Information' xxii

SECTION I: INTRODUCTION, BACKGROUND AND APPROACH **23**

CHAPTER 1. INTRODUCTION 25

1.1 Background and Motivation for this Research 25
1.1.1 Personal Motivation and Context 27
1.1.2 Research Objectives and Purpose 28

1.2 Contributions of the Research 31
1.2.1 An Understanding of what People want from Personal Data 31
1.2.2 Methodologies for Participatory Work around Personal Data 33
1.2.3 Best Practices and Design Guidelines for Systems and Processes involving Personal Data 34
1.2.4 Principles for Generative Action towards Better Data Relations 37
1.2.5 A Detailed and Actionable Research Agenda and Strategy for Empowerment & Systemic Change 38

1.3 Publications Arising From and Connected to This Research 39
1.3.1 Pilot Study 39
1.3.2 Primary Case Studies 39
1.3.3 Workshop Papers, Presentations & Interviews 42
1.3.4 Secondary Publications 43

1.4 The Structure of this Thesis 44

CHAPTER 2. LITERATURE REVIEW 49

2.1 Understanding Data and its Role in Today's World 49
2.1.1 What is Data? 49
2.1.2 Personal Data & the Rise of Data-centrism 52
2.1.3 Legislating to Protect Personal Data: The Story So Far 54
2.1.4 The Need for Practical and Effective Personal Data Access 56
2.1.5 Research Gap: The Human Experience of Personal Data 58

2.2 A History of Personal Data Interaction 59
2.2.1 Computers as General-Purpose Information Tools 59
2.2.2 Personal Information Management 60
2.2.3 Self Informatics (SI) 65
2.2.4 Overwhelmingly Complex Digital Lives 68
2.2.5 Research Gap: The Data Beyond the Individual 70

2.3 An Alternative to Data-centrism: Human-centricity 71
2.3.1 Human-Computer Interaction (HCI) Foundations 71
2.3.2 Data Transcendence, Context and Human-Data Interaction (HDI) 72

2.3.3 Human-centred Design: A Sociotechnical Challenge 73
2.3.4 The Pursuit of Human-centred Personal Data Ecosystems (MyData and beyond) 75

2.4 Summation: Defining the Research Question 78

CHAPTER 3. METHODOLOGY 80

3.1 Forming a Research Paradigm: Ontology & Epistemology 80

3.2 Research Approach 82
3.2.1 Participatory Co-Design & Experience-centred Design 82
3.2.2 A Two-Track Research Approach: Learning and Practice 86

3.3 Research Objectives 92
3.3.1 Research Question 1 (RQ1) 92
3.3.2 Research Question 2 (RQ2) 93
3.3.3 Practical Embedded Research: Going beyond the original RQs 94

3.4 Methodologies Employed in Case Studies 95
3.4.1 Sensitisation of Researcher and Participants 96
3.4.2 Discussion and Exploration with Stimuli 100
3.4.3 Participatory Co-Design of Possible Solutions 105
3.4.4 Data Analysis 109
3.4.5 Recruitment 112
3.4.6 Ethics 116

3.5 Summation: Towards an Understanding of Better Data Relations 117

SECTION II: UNDERSTANDING ATTITUDES & DESIRES THROUGH PARTICIPATORY CASE STUDIES 118

CHAPTER 4. CASE STUDY ONE: ACCESS TO CIVIC DATA IN EARLY HELP 122

4.1 Context: Data Use in Early Help 123
4.1.1 Data-centric Family Intervention in UK Social Care 123
4.1.2 Current Practice: Early Help Case Records as a Source of Truth 124
4.1.3 Rethinking the Role of Data in Early Help Support Relationships 126

4.2 Activities and Methodology for this Context 127
4.2.1 Embedded Research and Development Placement in CHC's SILVER Project 127

4.2.2 Pilot: Understanding Family Civic Data Study 127
4.2.3 Case Study One Workshops: Data Interaction in Early Help 128

4.3 Preliminary Explorations of Family Civic Data: Families' and Support Workers' Perspectives 130
4.3.1 What is Family Civic Data? 131
4.3.2 What is the Family Perspective on Their Civic Data and its Use? 131
4.3.3 What is the Staff Perspective on Family Civic Data and its Use? 132
4.3.4 Shared Data Interaction: An Equitable and Mutually Beneficial Data Use Model 133
4.3.5 Sensitising Participants and Discovering Shared Values 133

4.4 Thematic Findings 137
4.4.1 Themes & Subthemes 137
4.4.2 Theme 1: Meaningful Data Interaction for Families 143
4.4.3 Theme 2: Giving a Voice to the Family 147
4.4.4 Theme 3: Earning Trust through Transparency 152

4.5 Discussion 157
4.5.1 The Value of Involving People with Their Data 157
4.5.2 Effective Data Access Requires Human Interaction 160
4.5.3 Shifting the Locus of Decision Making Through Shared Data Interaction 162

4.6 Summation of Case Study One 168

CHAPTER 5. CASE STUDY TWO: THE HUMAN EXPERIENCE OF GDPR 172

5.1 Context: Accessing Your Personal Data Using Your GDPR Rights 173
5.1.1 The Current Need for Data Access 173
5.1.2 Current GDPR Research and its Limitations 175
5.1.3 Human-Data Interaction: Towards a Human-centric Personal Data Ecosystem 177

5.2 Overview of Activities for this Context 178
5.2.1 Formative Research in Personal Data and Digital Life 179
5.2.2 Case Study Two: The Human Experience of GDPR 180

5.3 Methodology for the Case Study Interviews 181
5.3.1 Interview Timeline and Stages 181
5.3.2 Interview Structure and Questions Asked 185

5.4 GDPR Request Outcomes 188
5.4.1 Interview 1: GDPR Target Selection 188

5.4.2 Interview 2: Privacy Policy Review and Goal Setting 192
5.4.3 Interview 3: Reviewing the GDPR Response 194
5.4.4 Perceived Power and Trust 198

5.5 Thematic Findings 203
5.5.1 Themes & Subthemes 203
5.5.2 Theme 1: Many Companies are Evasive and People are "Still in the Dark" 207
5.5.3 Theme 2: People Struggle to Understand, Use and Control Their Data 210
5.5.4 Theme 3: Poor GDPR Handling Can Damage the Forced & Fragile Trust Relationship 215

5.6 Discussion 222
5.6.1 Implications for Policymakers 222
5.6.2 Implications for Data Holders 224
5.6.3 Implications for Individuals 226

5.7 Summation of Case Study Two 228

SECTION III: DEFINING A NEW RESEARCH AGENDA: HUMAN DATA RELATIONS 231

CHAPTER 6. DISCUSSION - AN UNDERSTANDING OF WHAT PEOPLE WANT IN DATA RELATIONS 233

6.1 Answering RQ1: What do people want in *direct* data relations? 234
6.1.1 Visible Data 234
6.1.2 Understandable Data 239
6.1.3 Useable Data 244

6.2 Answering RQ2: What do people want in *indirect* data relations? 247
6.2.1 Process Transparency 248
6.2.2 Individual Oversight 252
6.2.3 Involvement in Decision Making 257

6.3 Summation: Empowering Individuals with Better Data Relations 261
6.3.1 Empowerment 261
6.3.2 Disempowerment 262
6.3.3 Towards A Better Society 263

CHAPTER 7. DEFINING A NEW RESEARCH AGENDA: HUMAN DATA RELATIONS 265

7.1 Foundations for Practical Exploration of Data Wants 265

7.2 'Human Data Relations': A Definition 268

7.3 The Role of Personal Data 271

7.4 The Two Distinct Motivations for Better Data Relations 273

7.4.1 Life Information Utilisation (LIU) 274

7.4.2 Personal Data Ecosystem Control (PDEC) 275

7.5 Four Objectives for Human Data Relations 275

7.6 Better Human Data Relations as a Recursive Public 277

7.7 Summation: HDR—A Landscape Ready to Explore 279

SECTION IV: OPERATIONALISING THE RESEARCH AGENDA THROUGH DESIGN & PRACTICE .. 281

CHAPTER 8. MAPPING THE HUMAN DATA RELATIONS LANDSCAPE 293

8.1 Obstacles to the HDR Objective of Data Awareness & Understanding 296

8.1.1 Invisible, Inaccessible or Unrelatable Data 296

8.1.2 The Personal Data Diaspora 296

8.2 Obstacles to the HDR Objective of Data Useability[11] 297

8.2.1 Immobile, Inaccessible or Unmalleable Data 297

8.3 Obstacles to the HDR Objective of Ecosystem Awareness & Understanding 298

8.3.1 Complex and Invisible Personal Data Ecosystems 298

8.3.2 A Lack of Metadata 299

8.4 Obstacles to the HDR Objective of Ecosystem Negotiability 301

8.4.1 Hegemony through Data Holding 301

8.4.2 The Active Diminishing of User Agency 304

8.4.3 The Intractable Data Self 307

8.5 Obstacles to the HDR Objective of Effective, Commercially-Viable and Desirable Systems 307

8.5.1 A Lack of Individual Demand 308

8.5.2 Closed, Insular and Introspective Practices 308

8.5.3 A Lack of Organisational Investment in HDR 309

8.5.4 A Lack of Interoperability 310

8.5.5 Insufficient Machine Understanding of Human Information 310

8.6 Summation of Chapter 8: From Obstacles to Opportunities 311

CHAPTER 9. PRACTICAL APPROACHES TO IMPROVE HDR 313

9.1 Additional Background: Theories of Change (ToC) 315

9.2 Additional Background on Personal Data Stores: Databox, Solid and the Hub-of-All-Things 318

9.3 Approach 1 to Improving HDR: Discovery-Driven Activism 321

9.3.1 Obtaining Your Personal Data 322

9.3.2 Data Flow Auditing 323

9.3.3 How Activists Pursue Change 324

9.3.4 Data Access & Ecosystem Understanding Services 325

9.4 Approach 2 to Improving HDR: Building the Human-centric Future 327

9.4.1 Life Interfaces 330

9.4.2 Life Partitioning 335

9.4.3 Entity Extraction From Pooled Life Data 339

9.4.4 Life Information Manipulation Capabilities 345

9.4.5 Digital Self Curation & Inclusive Data Flows 351

9.5 Approach 3 to Improving HDR: Defending User Autonomy and Hacking the Information Landscape 353

9.5.1 A Deceptive Landscape 354

9.5.2 The Battle at The Seams 356

9.5.3 Regulating to Free the Information Landscape 357

9.6 Approach 4 to Improving HDR: Winning Hearts and Minds: Teaching, Championing and Selling the Vision 359

9.6.1 HDR Education & HDR Literacy 360

9.6.2 Demonstrating the Business Value of Better HDR 363

9.6.3 Proving the Viability of Human-centric Approaches 363

9.7 Summation of Chapter 9, and of Section IV 368

SECTION V: CONCLUSIONS AND OUTLOOK 371

CHAPTER 10. THESIS CONCLUSION 373

10.1 Principles for Generative Action 374

10.1.1 Principle 1: Life Information makes Data Relatable 375
10.1.2 Principle 2: Personal Data Needs to be United and Unified 377
10.1.3 Principle 3: Data must be Transformed into a Versatile Material 379
10.1.4 Principle 4: Ecosystem Information is an Antidote to Digital Life Complexity 380
10.1.5 Principle 5: We Must Know Data's Provenance 383
10.1.6 Principle 6: Data Holders use Four Levers of Infrastructural Power 384
10.1.7 Principle 7: Human-centred Information Systems must serve Human Values, Relieve Pain and Deliver New Life Capabilities 387
10.1.8 Principle 8: We Need to Teach Computers To Understand Human Information 394
10.1.9 Principle 9: Individual GDPR requests can Compel Companies to Change Data Practices. 395
10.1.10 Principle 10: Collectives can Compare and Unify their Data and use their Pooled Knowledge to Demand Change. 397
10.1.11 Principle 11: Automated Entity Identification could Enhance Machine Understanding and Unburden Life Interface Users. 398
10.1.12 Principle 12: The 'Seams' of Digital Services need to be Identified, Exploited and Protected. 401
10.1.13 Principle 13: We could Demonstrate Business Benefits of Transparency and Human-centricity .. 404

10.2 Pragmatic Reflections on Pursuing Better HDR 406

10.3 Limitations and Future Work 413

10.4 Personal Reflection 419

10.5 Legacy of This Thesis 422

BIBLIOGRAPHY 426

ADDITIONAL REFERENCE INFORMATION 483

ARI2. ADDITIONAL REFERENCE INFORMATION FOR CHAPTER 2 484

ARI2.1 Data Protection Terminology and a Legal Definition of Personal Data 484

ARI3. ADDITIONAL REFERENCE INFORMATION FOR CHAPTER 3 486

ARI3.1 The Private Data Viewing Monitor 486

ARI4. ADDITIONAL REFERENCE INFORMATION FOR CHAPTER 4 487

ARI4.1 Family Civic Data Categories **487**

ARI4.2 Sentence Ranking - List of Sentences and Analysis Approach **488**

ARI4.3 Storyboarding Action Cards **490**

ARI4.4 Notation for Quotations in Chapter 4 ..... **496**

ARI4.5 Additional Participant Quotations **496**

Quotes from Families-Only Workshop [A] 496

Quotes from Staff-Only Workshop [B] 498

Quotes from Combined Parents and Staff Workshop [C] 507

ARI5. ADDITIONAL REFERENCE INFORMATION FOR CHAPTER 5 ..... **509**

ARI5.1 GDPR Data Analysis Approach **509**

ARI5.2 Best and Worst Companies for GDPR Handling **521**

ARI7. ADDITIONAL REFERENCE INFORMATION FOR CHAPTER 7 ..... **524**

ARI7.1 A Prototype Entity Extractor and Time-Event Extractor **524**

ARI9. ADDITIONAL REFERENCE INFORMATION FOR CHAPTER 9 ..... **526**

ARI9.1 How I compelled Spotify to improve their GDPR return **526**

APPENDICES **528**

Appendix A. The Pilot Study: Published CHI 2018 Paper **528**

Appendix B. Ethics Approvals **542**

INDEX OF KEY IDEAS, INSIGHTS AND CONTRIBUTIONS **551**

The Six Wants in Data Relations 551

HDR Objectives 551

HDR Obstacles 551

The HDR Principles for Generative Action 552

HDR Approaches 553

Thesis Contributions 553

Major Concepts of This Thesis 554

GLOSSARY OF PRE-EXISTING ABBREVIATIONS, NAMES AND TERMS 557

Abstract

PCs, smartphones, and cloud computing have transformed the world: In our daily lives, we interact with many businesses and public services who (often to reduce costs) rely on data collection and processing rather than face-to-face user interactions to inform their decisions. This creates an *imbalance of power* between those who hold personal data and the individuals about whom data is stored, who cannot easily see their data or examine how it is used. This *Digital Civics* PhD research explores, from an individualist perspective, the lived experience of this imbalanced and data-centric world. Through two qualitative case studies across public and private sectors, it discovers desires for visible, understandable and useable data, and for transparent relations with data holders that enable oversight and involvement. Case Study One focuses on *Early Help* social care: Through four workshops with supported families and social workers, perspectives on civic data use are understood. *Shared data interaction* is explored as a means to empower individuals while maintaining an effective care relationship. Case Study Two is a three-month study exploring 10 participants' experience of using *GDPR data access rights* to view their own data, resulting in insights into individual needs and the challenges of data-centric service relationships, and recommendations for improvement of policies and practices. Expanding upon prior work in *Personal Information Management*, *Human Data Interaction* and *MyData* personal data ecosystems, and informed by parallel industrial experience, a new research agenda for improving ***Human Data Relations (HDR)*** is established. The thesis also explores the *practical pursuit* of this agenda. Through the author's expert participation in projects at BBC R&D and Hestia.ai/Sitra, *the landscape for future research and innovation* is mapped out. 13 *Principles for Generative Action* are described, as well as objectives, obstacles, designerly insights and both conformist and activist strategies for change.

Dedication

For my children Rosie, Joey, and Zach; my nephew Elliott; and my nieces Amy and Lyla. My wish is that that you and your generation might soon experience a future where technology can truly help people and empower them to thrive, and where personal data drives human flourishing more than corporate profit. I hope that this research can contribute to a better future for you all.

In memoriam

In memory of my auntie Jenny Arnott (1938-2021) — a passionate traveller, art & history enthusiast and kind-hearted lady who always strived to share her passions and bring happiness to others.

Acknowledgements

No-one really knows how hard a PhD is until they are already well beyond the point of no return. It is demanding, challenging and often thankless and lonely work. You spend countless hours striving to find meaning among masses of data or iterating endlessly to concisely express complex, nebulous and elusive ideas. This endeavour is made even harder when you are a mature student with financial and parental responsibilities. At times, especially during the final unfunded writing-up period, the impacts upon my life and those around me were huge and unreasonable. For this reason, the greatest thanks of all go to my wife Joni Bowyer, who stood by me throughout, picking up the slack again and again where I could not. She endured the uncertainties and impacts upon income, time, and divided attention that this unforgiving work thrust upon our family. She also provided practical help on countless occasions with everything from poster layout to time management to editing and grammatical advice. I love you forever, Joni. Thank you.

The next person I want to thank is Jack Holt, who dedicated many weeks and months of his life to collaborate with me to analyse of mountains of participant data from Case Study Two and to co-write the paper (Bowyer, Holt, *et al.*, 2022) with me. I have absolutely no doubt that without his dedication, it would have been impossible to complete and publish the GDPR study in any reasonable timeframe.

I would like to thank, in reverse chronological order, my supervisors and all the other faculty and staff who have supported me on this six-year journey:

- Dave Kirk, for both detailed and high-level advice in bringing my thesis to conclusion during the final year;
- Jan Smeddinck, for his calm, pragmatic encouragement and thorough feedback and on drafts and plans through the latter half of my research;
- Rob Wilson, for always reminding me to stay grounded in the data, and for his sage advice in matters sociotechnical and philosophical;
- Josephine Go Jefferies, for much-needed scrutiny and challenges to my writing and many detailed chapter draft reviews;
- Rachel Pattinson, for being the best CDT manager a postgraduate could hope for;

- Patrick Olivier, Pete Wright and Dave Kirk, for their continuing commitment to help me find ways to make the PhD financially viable through the finding and accommodating of external paid work;
- Kyle Montague, for his valuable input on study design in the early stages of my PhD, and for advocating to protect my independence and integrity as a researcher during problematic negotiations with a partner organisation;
- Phil Lord, for assistance and advocacy with those same issues at a crucial time;
- Madeline Balaam, for inspiration on participatory methods right at the start, and for helping me develop a paper writing style;
- Rob Comber, Simon Bowen, Matt Wood and other lecturers who taught me valuable *Digital Civics* and research skills during the MRes, that helped shape me into the researcher I am; and not forgetting
- Paul, Nicola, Sara, James, Glau, Fion, Alex, and all the other School admin staff, for countless random assists.

I have been fortunate to take this journey with dozens of other researchers. I would like to especially thank:

- Tom Maskell, for his help with wrestling with concepts around data access and involvement, and companionship and Digital Civics discussions on many bus commutes in the early years;
- Sunil Rodger, for valuable moral support, writing camaraderie, and practical advice for the last two years; and
- Stuart Wheater, for myriad data discussions and tactical discussions that helped me get through a difficult period earlier in the PhD and which often went beyond project business.

I would also like to thank, in no particular order:

- Rebecca Nicholson, Sean Peacock, Jen Manuel, Rosie Bellini, Megan Venn-Wycherley, Kieran Cutting, James Hodge, Hazel Dixon, Seb Prost, Sara Armouch and all the other *Digital Civics* PhDs whose company and mutual support I enjoyed during dozens of writing sessions and discussions;

- (Matt) Marshall, for sharing his powerful script framework for markdown-based thesis writing and document generation (Marshall, 2020), and supporting me in adapting it, making thesis development so much easier;
- Tom Nappey, Laura Pinzon Cardona and Rosie Daglish for their assistance with graphic design;
- Louis Goffe, Debbie Smart, Kat Jackson, Liam Spencer, Ruth McGovern and Kyle Montague for giving their time to help run workshops with participants in the pilot study and in Case Study One;
- Paul-Olivier Dehaye, Mike Martin, Soheil Human, Jasmine Cox, Peter Wells, Ian Forrester, Rhianne Jones, Tim Broom, Suzanne Clarke, Hannes Ricklefs, Max Leonard, Chris Gameson, Euijin Hwang, Louis Goffe, Toby Lowe, James Nicholson, Alan Mitchell, Iain Henderson, StJohn Deakins, Dalia Al-Shahrabi, Anna Scott, Sarah Knowles, and Michael Jelly, all of whom I have had the pleasure to work with, ruminate with or learn from;
- Chris Lintott and everyone in the Oxford University/Zooniverse team for giving me the smoothest possible start in postgraduate academia before I joined Open Lab;
- Kellie Morrissey, Aare Puussaar, Andy Dow, Anja McCarthy, Zander Wilson, Raghda Zahran, Michael Jelly, Jay Rainey, David Williams, Paula Williams, Ben Wright, Paul Whittles, Jon Bowyer, and all the other lovely people who encouraged me and supported me and my family along the way on this journey; and
- the research participants themselves, who shall remain nameless but without which this work could not exist.

Finally, I would like to thank my parents, Jim and Rosi Bowyer. You have always been there for me, whenever I needed you. Thank you for your unquestioning love and support in an ever-changing world.

This PhD was funded by the Engineering and Physical Sciences Research Council (EPSRC)'s Centre for Doctoral Training in Digital Civics at Newcastle University in the UK (EP/L016176/1).

Lists of Tables and Figures

Lists of Tables by Chapter

Tables in Chapter 3

- 3.1 - Context One (Civic Data & Early Help): Participants involved in Research Activities leading into Case Study One.
- 3.2 - Context Two (Digital Life): Participants Involved in Digital Life Research Activities Leading into Case Study Two.

Tables in Chapter 4

- 4.1 - Case Study One Group Design Workshops
- 4.2 - Theme 1 - Meaningful Data Interaction for Families: Subthemes & Participant Quotes
- 4.3 - Theme 2 - Giving a Voice to the Family: Subthemes & Participant Quotes
- 4.4 - Theme 3 - Earning Families' Trust Through Transparency: Subthemes & Participant Quotes

Tables in Chapter 5

- 5.1 - Types of Data Holding Organisation Targeted for GDPR Requests by Study Participants
- 5.2 - Types of Personal Data Potentially Accessible from Data Holders via GDPR Rights
- 5.3 - Presence and Quality Assessments of GDPR Responses by Data Type (as Percentages)
- 5.4 - Participants' Hopes, Imagined Data Uses and Goals for GDPR, as well as Resultant Outcomes
- 5.5 - Theme 1 - Insufficient Transparency: Subthemes & Participant Quotes
- 5.6 - Theme 2 - Confusing & Unuseable Data: Subthemes & Participant Quotes
- 5.7 - Theme 3 - Fragile Relationships: Subthemes & Participant Quotes

Tables in Chapter 7

- 7.1 - Eight Lenses on Personal Data

Tables in 'Additional Reference Information'

- ARI4.1 - Example Categories of Family Civic Data
- ARI5.1 - Best and Worst Data Holders for GDPR, according to Participants' Judgements

Lists of Figures by Chapter

Figures in Chapter 1

- 1.1 - Poster Presentation of Case Study One at CHI 2019
- 1.2 - The Structure of This Thesis

Figures in Chapter 2

- 2.1 - The Wisdom Curve: Making Data into Meaningful Information
- 2.2 - Li *et al.*'s Stage-based Model of Personal Informatics Systems

Figures in Chapter 3

- 3.1 - The Action Research Cycle, as Applied in this PhD
- 3.2 - The Two-Track Research Approach of this PhD, and its Feedback Cycles
- 3.3 - Family Facts — What is Data?
- 3.4 - Walls of Data — Sensitising Participants to the World of Commercially-held Data and GDPR
- 3.5 - Sentence Ranking — Bringing Support Workers and Families to a Shared Problem Space
- 3.6 - Family Civic Data Cards — Things to Think With
- 3.7 - Personal Data Examples — Making Data Relatable
- 3.8 - Home Interviewing: Card Sorting with a Family in Their Living Room
- 3.9 - Ideation Decks — Combining Random Design Ingredients to Generate New Ideas
- 3.10 - Group Poster Design — A Participant-designed Poster to Advertise Features of Imagined Data Interface Products

- 3.11 - Storyboarding Action Cards — A Collaboratively-constructed Narrative Created through Discussion from a Palette of Possible Parent and Staff Actions
- 3.12 - Thematic Analysis of Qualitative Data using *Quirkos* for Case Study One
- 3.13 - Spreadsheet-based Quantitative Analysis of Interview Data for Case Study Two
- 3.14 - Pilot Study Recruitment Poster

Figures in Chapter 4

- 4.1 - Participants' Shared Values Deduced from Sentence Rankings Data
- 4.2 - Current Model of Data Interaction, and Proposed Model of Shared Data Interaction

Figures in Chapter 5

- 5.1 - A Journey Map of Each Participant's Study Progression
- 5.2 - An Example Life Sketch from Interview 1, with Data Handling Companies in Red, Data Types in Blue, and Feelings in Green
- 5.3 - Sankey Overview of Participants' GDPR Requests
- 5.4 - Longitudinal Distribution of Net Changes in Participants' Perceived Power and Trust Scores
- 5.5 - Perceived Power Balance between Individual and Data Holder at Different Stages of the GDPR/Study Process
- 5.6 - Participants' Perceived Trust in Provider at Different Stages of the GDPR/Study Process

Figures in Chapter 7

- 7.1 - The Two Motivations for HDR: Controlling Your Personal Data Ecosystem and Utilising Your Information About Your Life
- 7.2 - Mapping the Six Wants into Objectives for the HDR Opportunity Landscape

Figures in Section IV

- IV.1 - Screenshot from a Web Augmented version of the *Just Eat* Website

Figures in Chapter 8

- 8. 1 - Obstacles and Resulting Insights in the HDR Opportunity Landscape

- 8. 2 - Some of the Many Aspects of Metadata that Might Exist About a Datapoint or Dataset
- 8. 3 - The Panopticon Structure of the Illinois State Penitentiary

Figures in Chapter 9

- 9. 1 - Theory of Change [ToC]: The Four Dimensions of Change
- 9. 2 - HDR Approach 1: Discovery-Driven Activism
- 9. 3 - HDR Approach 2: Building the Human-centric Future
- 9. 4 - Conceptual Model for a Personal Data Store System
- 9. 5 - High Level Data Types
- 9. 6 - Life Information Modelled as Happenings
- 9. 7 - A Simple PDS Life Information Presentation Model
- 9. 8 - Mock-up of Life Information Presented in a PDS Interface
- 9. 9 - Life Partitioning Analogy using a Cluedo™ board
- 9.10 - Mock-up: Browsing by Areas of Life
- 9.11 - Facebook's *world2vec* Model, Semantically Modelling Human Information from Social Media Posts on Facebook
- 9.12 - Identifying the Attributes of Data
- 9.13 - Determining the Nature of a Piece of Data
- 9.14 - Attributes of Data
- 9.15 - Actions One Might Perform on Life Information
- 9.16 - Questions One Might Ask of Life Information
- 9.17 - Example Taxonomies for Life Information Navigation
- 9.18 - HDR Approach 3: Defending User Autonomy and Hacking the Information Landscape
- 9.19 - The Modern 'Black Box' View of Technology
- 9.20 - HDR Approach 4: Winning Hearts and Minds: Teaching, Championing and Selling the Vision
- 9.21 - SILVER Health Data Viewing Interface
- 9.22 - Summary of Generalised Change Strategies for Pursuing Better HDR, Using the ToC Model

Figures in Chapter 10

- 10. 1 - Life Concept Modelling
- 10. 2 - Mock-up of a Unified TV Viewing History Interface
- 10. 3 - SubsCrab: An Example Application for Ecosystem Detection and Visualisation
- 10. 4 - The Four Levers of Infrastructural Power
- 10. 5 - Human Values (from BBC R&D research)
- 10. 6 - A Contact-and-Calendar-centric PDS Approach
- 10. 7 - The Scattered Data Relating to a Vacation
- 10. 8 - Mock-up of a Unified Interface for a Vacation
- 10. 9 - Annotating Data with Semantic Context
- 10.10 - Identifying Entity Associations in Data

Figures in 'Additional Reference Information'

- ARI3.1 - Private Data Viewing Monitor with Viewing Glasses
- ARI4.1 - Extract of Sample Scenario Storyboarding Exercise walkthrough
- ARI4.2 - Example Backing Mat for Storyboard Decks
- ARI5.1 - Screenshot from *Quirkos* During Coding Process
- ARI5.2 - Screenshot from *Quirkos* at End of Coding Process
- ARI5.3 - Screenshot from *Workflowy* During Theme Construction
- ARI7.1 - Prototype Entity Extractor and Time-Event Extractor

SECTION I:

INTRODUCTION, BACKGROUND & APPROACH

Introduction to Section I

This thesis is divided into five sections:

- Section I: Introduction, Background and Approach
- Section II: Understanding Attitudes and Desires through Participatory Case Studies
- Section III: Defining A New Research Agenda: Human Data Relations
- Section IV: Operationalising the Research Agenda through Design & Practice
- Section V: Conclusions and Outlook

This first section, Section I, provides a thorough background and explanation to the thesis through three chapters:

- Chapter 1 introduces the research, explaining the background, context and objectives, then summarises the contributions and publications generated, before outlining the structure of the rest of the thesis.
- Chapter 2 contains a literature review, laying out the foundational ideas upon which this research is built; drawing from information theory, data access legislation, personal data interaction system design, and the emergent human-centric paradigm of personal data ecosystems.
- Chapter 3 describes the two-track research-and-practice approach of the thesis, from its philosophical foundations to its overall research objectives, and details the specific participatory techniques, methodologies and data analysis approaches used in the research.

Chapter 1. Introduction

"My data is everywhere, and I am nowhere."

—Imogen Heap (musician and digital rights advocate), speaking at MyData 2019, encapsulating a core problem with modern digital life.

1.1 Background and Motivation for this Research

We live in an increasingly data-centric world, where our direct and indirect interactions with computer systems depend upon the collection, storage and use of personal data about individuals. Motivated to reduce costly human interaction and scale to serve more customers, organisations capture and represent individuals as data and rely increasingly on the interpretation of those datapoints to make decisions—decisions that affect our everyday lives in myriad ways, from determining eligibility to access particular services, benefits or products or targeting advertisements or recommendations to influencing our decisions and behaviour. Data about people has become extremely valuable. It is 'the new oil' (Toonders, 2014) driving a model of *surveillance capitalism* (Zuboff, 2019) that collects data and exploits it for profit.

This multi-party use of data has resulted in a splintering (Lemley, 2021) of our *digital selves* across hundreds of different organisations' computer systems, creating 'a chaos of multiplicity' (Bødker, 2015) where data becomes trapped (Abiteboul, André and Kaplan, 2015) and hard to manage. Current data practices cause harm to individuals through anxiety, distraction and a sense of being overwhelmed (Fu *et al.*, 2020; Timely, 2020). Data use can harm society too, due to the ease with which people's attention and beliefs are manipulated, risking radicalisation and a loss of democratic freedom (Thompson, 2011; Chan, 2019). It is generally accepted that:

> *"There is **a power imbalance** in the amount of information about individuals held by industry and governments, and the lack of knowledge and ability of the same individuals to control the use of that information."*—World Economic Forum (Hoffman, 2014b)

To address this power imbalance, people need a more effective relationship with their own data. This has been conceptualised as a lack of *agency* (the ability to act for oneself), *negotiability* (the ability to exert influence over data use within the system as things change), and *legibility* (the ability to understand one's data and its implications) (Mortier *et al.*, 2014). A need is growing for the conscious design of human relationships with data, as these problems cannot be solved through interface design alone but only through the reconfiguration of service relationships that involve data. Currently, users of today's digital services typically experience a *point of severance* (Luger and Rodden, 2013): they are forced to sacrifice their data and are subsequently cut out of the loop. Without understanding how they are currently seen through data, people face risks of unfair treatment, or physical or psychological harm (Bowyer *et al.*, 2018; Crossley, 2022). Left unexplored and unchallenged, the situation will not improve, as the datafication of society grows, and user agency continues to diminish.

This thesis focuses on that power imbalance and seeks to design improvements to information interaction interfaces, processes and service relationships that can improve individual agency, legibility and negotiability. This PhD tackles this design challenge through a hybrid approach that uncovers individual needs through participatory co-design workshops and qualitative interviews, alongside grounded innovation and development work carried out by the author while embedded in industrial and academic project settings, focusing on operationalising and refining the emerging design objectives through user-centred system design, conceptual modelling, and prototyping.

The two parallel tracks of this research [Figures 1.2, 3.2] are described separately in Section II, which describes Case Studies exploring how people relate to data, and Section IV, which embraces the commercial realities and lived experiences of today's data-centric world to present the author's grounded learnings from projects that seek better data relations. These tracks ran concurrently, resulting in a perpetual action research feedback cycle [Figure 3.15] where participant insights and theoretical models informed practical system and process designs while real-world learnings about societal challenges and practical obstacles refined the emerging models.

Section III combines Case Study findings into a unified theory that captures six core human desires for data relations as future research goals for empowering individuals through their

personal data. A new research agenda **Human Data Relations (HDR)** is formalised, which incorporates Human Data Interaction (HDI) (Mortier *et al.*, 2013, 2014) theory. This specifies four objectives for improved HDI and data relations in practice. Section IV therefore serves as an operationalisation of that agenda, documenting practical obstacles as well as four distinct approaches to the pursuit of HDR (incorporating both conformist design and *adversarial design* (DiSalvo, 2012) thinking). Section V crystallises these novel contributions to knowledge into thirteen Principles to support the generation of actions by future researchers and innovators and designers who wish to pursue the empowerment of individuals and the rebalancing of power over data. It equips them with ideas and understandings they will need to bring about better Human Data Relations at a societal level.

1.1.1 Personal Motivation and Context

This PhD and this thesis represent the culmination of my lifelong passion to help people get more value from our computers. My experience and expectations of computers were shaped by the 1980s home computing revolution, which taught me and a generation of young people that the computer was a machine to program, a tool to be exploited, mastered and bent to your will. Then, in my formative years approaching the turn of the millennium, I lived through the birth of the public Internet and marvelled at the ability for computers to connect people across the world, empower individuals as creators, innovators and broadcasters, level the playing field and transform the way people interact. In my software engineering career, I gradually transitioned from back-end to front-end development and ultimately to User Experience (UX), driven to take a more active role in building software features that directly benefit users and improve their lives. Keenly tracking and embracing the Web 2.0 revolution while observing the digitisation and disruption of so many industries, I became fascinated with the ways in which humans were shaping computer systems which in turn were shaping our habits and our society, phenomena I explored in writing on the Human 2.0 blog which I co-founded (Bowyer and Croll, 2009-2011).

But then, having seen Internet-era computing give us new capabilities, and knowing the potential of computers to become tools for positive change in society, I bore witness to a changing world, and the balkanisation and commercialisation of the once-free Internet. As companies adapted to the Information Age and shifted to data-driven, cloud-centric

business models, our ability to harness computers for our own ends began to slip away. While immersed in the start-up community in Montréal, Canada, I became frustrated at this loss of potential. Driven to explore the reasons for, and implications of, this loss of agency, and the possibilities for more human-centric computing, I published several essays and presentations (Bowyer, 2009, 2010, 2011, 2012a, 2012b, 2013) which collectively form the seed from which this thesis grew.

By 2014, it was beyond doubt to me that the software industry had lost its way, prioritising business goals over user agency, reducing features and creating technology designed with in-built limits to our capabilities and interfaces optimised to corral users to behave in certain ways. Web 2.0's revolutionary potential of a 'people's internet' had been squashed and withered away in the shadow of new data giants Google, Facebook, Apple and Amazon, who reshaped and usurped Internet, Web and smartphone technologies for profit, at great cost to human wellbeing. Against a backdrop of a social media revolution which was literally breaking society and democracy to drive profit (Tufekci, 2017; Hall, Tinati and Jennings, 2018), I took the leap to escape corporate IT. I sought to research, design and build a better digital future where computers could be made useful again. This led me to the *Digital Civics* CDT programme (Open Lab, 2014), where I was finally able to work full-time on what I consider the most important problem of our age—**Understanding and Improving Human Data Relations**.

1.1.2 Research Objectives and Purpose

The aim of this thesis is to establish how people relate to personal data, how they understand and use it, and what they want from it and its holders in order to thrive and pursue their goals; and from there, to begin to explore how those needs might be met in practice.

The thesis is informed by a constructivist ontology and a pragmatist, individualist epistemology [3.1], and employs a multi-disciplinary *Digital Civics* (Vlachokyriakos *et al.*, 2016) approach to examine two key Research Questions (RQ):

RQ1. What is the human experience of personal data, and what do people want from their data? [3.3.1]

RQ2. What role does data play in people's service relationships, and how could relationships involving data be improved? [3.3.2]

This thesis assumes that if the asymmetry over data access and use between individuals and organisations holding data is to be addressed then a greater understanding of current data use issues is needed by all parties; thus the production of knowledge and insights is a vital first step towards the design of a more balanced model of data use that can deliver increased agency and negotiability.

After this Introduction chapter, Section I continues with a review of relevant existing literature and research [Chapter 2] to identify a clear baseline and research gap, then an explanation of the research approach and methodological choices to be used [Chapter 3].

Section II reports on the two Case Studies, which invited participants to 'look behind the curtain' of the opaque data-centric organisations they interact with. These enabled participants to consider more deeply the collection, storage and use of their personal data by service organisations. A participatory research design was employed, collecting interview transcripts to enable qualitative analysis and identify themes that inform a descriptive model of human-centred data empowerment needs. The focus was upon examining current practices, identifying attitudes to those practices, and imagining alternative designs and approaches for data use by service providers and the participants themselves. The participant groups were:

- **Case Study One**: Supported families, who meet with *Early Help* support workers (whose role to understand and empower those individuals to improve their lives is currently supplemented through access to myriad civic data records including housing, benefits and educational data) [Chapter 4]; and
- **Case Study Two**: Ordinary British and European citizens, who gained new legal rights via 2018's *GDPR* legislation [2.1.3], enabling them to request copies of held personal data along with other relevant information from the companies and service providers in their lives [Chapter 5].

Comparing the two case studies reveals commonalities in individual attitudes and understandings, validating each other's findings and allowing the expression of clear insights about people's relationships to personal data that can answer the two RQs. This synthesis and analysis of interview data enabled the generation of a descriptive model to address the research gap, concluding the core academic research of this PhD [Chapter 6].

This research is situated in the HCI discipline, which means that design (both participatory co-design and expert-informed user-centred design [3.2]) forms a key part of the approach to exploring the problem space. Examination of individual attitudes and desires around data in a 'whole life' sense is an under-researched area. Where HCI traditionally focuses on the mechanisms by which humans interact with data, the Case Studies, like the thesis as a whole, adopt a more sociotechnical focus on understanding lived experience.

In order to produce workable designs and theories that could be applied in the real world, the remainder of the thesis takes an even broader perspective, recognising that for participants' desired changes in data relations to be realised requires examination, recognition and confrontation of current technical, legal and commercial realities and the multi-party complexities of modern digital life.

The design work presented in Section IV complements the participatory Studies, providing an additional source of learning: real-world explorations, conducted alongside the qualitative Case Studies, of how to design more effective data relationships. This is done through design, modelling and conjecture, drawing upon the author's direct experiences working part-time in related industrial projects that share this thesis's focus on empowering individuals through data.

To support future researchers, activists and innovators in achieving the vision of a more human-centric future, the pursuit of the identified wants are shaped into a defined new research agenda - **Human Data Relations (HDR)**, which represents a broadening of focus from merely understanding the world towards planning how to change it. Motivations and objectives are identified, and in [Chapter 7] HDR is positioned as a broad activist agenda whose practitioners seek to reconfigure society to their own advantage.

Sections IV and V are deliberately open-ended, mapping out the existing landscape of challenges and possibilities for the HDR agenda, deviating from a traditional thesis

structure in order to offer more actionable insights. Given the scale of the sociotechnical design challenge society faces, this thesis does not carry out 'in the wild' evaluation of particular data interaction approaches or interface designs. Instead, drawing upon direct experience as well as the work of other researchers and innovators in this space, it documents known obstacles [Chapter 8]. It also maps out four possible change trajectories [Chapter 9] that can inform future design and innovation in human-centric system design.

A series of design Principles underpin Chapters 8 and 9, developed by the author based upon the learnings from both the participatory and project-based research tracks. These Principles are documented in [Chapter 10], along with reflections, limitations, a summary of contributions, and ideas for future work. The presented strategies and insights (which should be viewed as speculative rather than definitive), can aid the design and implementation of initiatives that recognise and confront the unique challenges of the status quo in pursuit of improving people's agency and negotiability over and through personal data.

Taken together, the two parallel research tracks of this thesis serve as novel and actionable reference material for future research, activism and innovation, solidly grounded in literary theory, participant experience, and industrial reality.

1.2 Contributions of the Research

The contributions of this thesis are summarised in this section, and described in more detail throughout the thesis, in the referenced chapters and subsections.

1.2.1 An Understanding of what People want from Personal Data

Through the qualitative Case Studies in Chapters 4 and 5 and the discursive analysis in Chapter 6, this thesis establishes a clear picture of what people want from their personal data. In general, people need all of their personal data to be *visible* [6.1.1], *understandable* [6.1.2], and *useable*[11] [6.1.3].

Furthermore, the sociotechnical focus of the research delivers deep qualitative insights into the *relationships* people have with data holding organisations and services, revealing attitudes and emotions including distrust, fear, frustration, confusion, helplessness and curiosity. Specific insights are drawn for the two Case Study contexts:

- In social care [Chapter 4], families want to be given a voice, and an ability to see and meaningfully interact with their data. They want to be able to influence how they are understood through data, and to be able to explain and discuss their data with authorities. The Case Study (and its pilot study [4.2.2; Bowyer *et al.* (2018); Appendix A] validate prior findings that highlight families' difficulties with civic data, while documenting their ideas for improving the use of their civic data and for improving empowerment and trust within the support context.
- In the service provider relationships of everyday life [Chapter 5], individuals want to be able to understand what companies are collecting about them. They want to see that data, understand what inferences and decisions result from its use, and exert control over how it can be used, kept and shared. These desires (which also explain motives for using GDPR rights) provide novel insights into the lived experience of the data-centric world, and in particular of using the GDPR to access and control one's personal data.

Examined together [Chapter 6], the Case Studies reveal clear common themes for people in service relationships that involve personal data; people want *process transparency* [6.2.1] when it comes to processes that use their personal data, *individual oversight* [6.2.2] over how their data is used, and ultimately *involvement in processes and decision making* [6.2.3].

The Case Studies also establish a more granular view of what we mean by personal data. The pilot study for Case Study One [4.2.2; Bowyer *et al.* (2018); Appendix A] establishes the term *family civic data* [ARI4.1], which can encompass relationship information, educational records, welfare details, health data, housing data, and civic and legal records. Case Study Two establishes a model of personal data according to its origin; data can be *volunteered*, *observed*, *derived*, *acquired* or *metadata* [Table 5.2]. This model has been adopted by others: it has been used for three years in design, ideation and strategy planning at BBC R&D, and was adopted throughout Sitra/Hestia.ai's *digipower* investigation [See Section IV Introduction] (for explaining data holding to participants and as a frame for data analysis) (Bowyer, Pidoux, *et al.*, 2022; Pidoux *et al.*, 2022).

Through the design and innovation work in Section III and beyond, a deeper understanding of the relationship people have with personal data is developed, and of what role data plays in people's lives. This forms part of the research agenda summarised in 1.2.5 below.

1.2.2 Methodologies for Participatory Work around Personal Data

A significant component of this PhD (the Case Studies and the preceding formative work and pilot studies) involved selecting, designing and executing methods for engaging with individuals and groups around personal data. At the outset, the challenge of getting laypeople to engage with this potentially dry and boring topic seemed significant. However, building on established methodologies such as card sorting [3.4.2] and ideation decks [3.4.3], a number of novel approaches were developed, which constitute novel contributions of this research:

Data Cards and Family Design Games

This approach was pioneered in the pilot study [4.2.2]; Bowyer *et al.* (2018); [Appendix A]. To serve as *boundary objects* and *things to think with* [3.4.2], a tangible set of *Family Civic Data Cards* [shown in Figure 3.6] were developed to represent the different types of civic data. These were used in a variety of exercises in both the pilot study and in Case Study One. While these are not formally assessed as a methodology in this thesis, these were very relatable to participants and proved highly effective at getting conversations flowing around data. A similar approach was used during the BBC Cornmarket project, where a set of Data Cards derived from the Family Civic Data cards concept were designed by the author and BBC colleague Chris Gameson to represent different types of data in the 'everyday digital life data' context (see Bowyer, 2020b for mockup). These cards were successfully used in a 1,500 participant study by BBC R&D (Sharp, 2021), showing the value of this contribution. As well as the Family Civic Data Cards, a number of other original activities for considering data such as 'Family Facts' [Figure 3.3] and 'Sentence Ranking' [Figure 3.5] were developed and used in the Case Studies. Some of these were published by the author as *Family Design Games* (Bowyer *et al.*, 2018).

Storyboarding Action Cards

For the third workshop of Case Study Two, there was a need to explore interpersonal interactions between different parties in a service relationship, without getting caught up on details of interface layouts or software and hardware capabilities. A novel approach was developed consisting of the use of *storyboarding action cards* [Figure 3.11] with participants to give them a vocabulary to map out interactions at a sociotechnical, functional level. These

cards are documented in ARI4.3 and this approach could be useful to others wishing to undertake participatory co-design work at the relationship or service design level.

A Methodology for Qualitative Interviews that Explore a User's Personal Data

For Case Study Two, an approach was designed that would allow a researcher to accompany an individual participant on a journey of exploration of their own personal data files, retrieved from companies using GDPR, and to carry out a guided 'deep dive' into that data, in order to reveal attitudes to revealed data storage and use practices. This approach is detailed in 5.2.2, and focused on using privacy policy promises and GDPR access rights as a frame of reference from which to assess data response quality and usefulness. The methodology was successful, enabling the production of detailed insights which are presented in Chapter 5. The methodology was not designed with a view to it being used elsewhere, but in fact the author was engaged by Hestia.ai to replicate this methodology (with some expansion and tweaks) as co-leader of the digipower investigation, which aimed to take 15 high profile European politicians, civil servants and journalists through a similar journey of discovery. The digipower investigation took place in 2021 and was also able to apply the methodology very successfully, resulting in the publication of multiple research reports [1.3.4].

A technological accompaniment to the Case Study Two methodology was originally developed by the author for use in the interviews, the *private data viewing monitor* [ARI3.1], but this was not used for the Study due to the infeasibility of face-to-face interviewing during the COVID-19 pandemic in 2020.

1.2.3 Best Practices and Design Guidelines for Systems and Processes involving Personal Data

The findings from the Case Studies allowed clear recommendations to be produced that can be applied by service providers and other stakeholders in the two specific contexts of the Case Studies. These contributions can be applied more broadly to other data-related service interactions.

Shared Data Interaction in the Context of Care

Through Case Study One [Chapter 4], a new human-centric approach and set of practices for support workers working with with families in the Early Help context was developed: To treat family civic data as a resource to be looked at, discussed and updated together on a shared screen or tablet, within the context of a support interaction. This *Shared Data Interaction* approach was considered advantageous by both staff and families, as it would offer families increased negotiability and evidence-based feedback while also provided staff with a more accurate view of families' lives and greater confidence of their consent. The approach was applied in detail to several scenarios as part of the Case Study, and also has the benefit of shifting the *Locus of Decision Making (LDM)* [4.5.3] closer to the supported family.

In pursuit of Shared Data Interaction, 12 desirable best practices (which could be of value to any service organisation focused on individual development) are laid out through the sub-themes of Case Study One [4.4]:

- providing understandable information summarises;
- interacting with data together;
- offering direct and unified data access;
- providing ongoing data access and support;
- avoiding treating people like records;
- checking data together;
- accounting for changes in people's lives;
- enabling individuals and families to contribute data;
- providing granular access controls;
- handling personal data transparently and with respect;
- prioritising the acquisition and demonstration of greater understanding; and
- pro-actively challenging data-centric norms.

Recommendations for Policymakers, Service Providers and Individuals on Data Use

Through Case Study Two [Chapter 5] (which explores relationships with service providers in everyday life - especially those where GDPR rights apply), recommendations for rethinking the use of data (and especially access to it) in such relationships were produced:

- Policymakers are recommended to regulate and provide guidance to improve GDPR response quality, to improve enforcement and protection of data rights, and to redesign future policies to consider ongoing access to and understanding of data rather than one-time file delivery.
- Data holders are presented with evidence that improving data transparency can improve customer loyalty. Treating data as a co-owned resource could increase engagement, and an untapped demand for personal data ecosystems is identified.
- Individuals are advised of the importance of their personal data, given it contains meaningful and potentially valuable information about their lives. Individuals are recommended to exert their data rights to improve their awareness and make more informed choices.

An additional contribution presented is the evidence of the positive and negative impacts of GDPR: Knowledge and experience of poor data handling practices are shown to harm provider trust and perceived individual power, but both trust and individual power can be improved through transparency and inclusive data processes.

A Reframing of Data Literacy for the Sociotechnical Context

Building on existing thinking about *data literacy* as a set of skills one can acquire to increase one's effectiveness in an increasingly technological world (Gurstein, 2011; Knight, 2016; Precisely Editor, 2022), this thesis offers an expanded definition of data literacy for the context of data-driven service relationships [summarised in 9.6.1 and explained throughout Chapter 9], including knowledge and appreciation of:

- the intrinsic value of personal data
- the implications of how one's personal data is used and shared by organisations
- the importance of portable data and of separation between platforms and personal data
- how to exercise one's data rights and assess and act upon responses to data access requests
- the need for a free and fair information landscape
- the ability to identify when one's individual agency is being diminished

All of the guidelines and recommendations that have been summarised in 1.2.3 can guide designers and inform individual and organisational choices, but this thesis also contributes specific Principles which can be used to generate system, process and interface designs for better data relations. The Principles are summarised in the next subsection.

1.2.4 Principles for Generative Action towards Better Data Relations

In Chapters 8 and 9, examples and insights drawn from the author's work with BBC R&D and Hestia.ai, as well as from the work of HDI specialists and activists, are shared. The conclusion of this thesis formalises these learnings and insights into thirteen concrete Principles for improving our relationships with and through data. These Principles, detailed in 10.1, go further than interface design guidelines as they address the sociotechnical reality of how personal data is used today. They could be used to generate designs, interfaces, research projects, activist movements, grassroots campaigns, prototypes, user studies and more and serve as the ultimate actionable outputs from this PhD. The Principles are:

- Principle 1 - Life Information Makes Data Relatable.
- Principle 2 - Data Needs to be United and Unified.
- Principle 3 - Data Must be Transformed into a Versatile Material.
- Principle 4 - Ecosystem Information is an Antidote to Digital Life Complexity.
- Principle 5 - We Must Know Data's Provenance.
- Principle 6 - Data Holders use Four Levers of Infrastructural Power.
- Principle 7 - Human-centred Information Systems Must Serve Human Values, Relieve Pain and Deliver New Life Capabilities.
- Principle 8 - We Need to Teach Computers to Understand Human Information.
- Principle 9 - Individual GDPR Requests can Compel Companies to Change Data Practices.
- Principle 10 - Collectives can Compare and Unify their Data and Use it to Demand Change.
- Principle 11 - Automating the Identification of Entities can enhance Machine Understanding and Unburden Life Interface Users.
- Principle 12 - The 'Seams' of Digital Services need to be Identified, Exploited and Protected.

- Principle 13 - It is Possible (and Necessary) to Demonstrate Business Benefits of Transparency and Human-centricity.

1.2.5 A Detailed and Actionable Research Agenda and Strategy for Empowerment and Systemic Change

In order to bring together the learnings of this research and the above contributions into the most actionable form, this thesis (specifically Section III) outlines a new research agenda, called ***Human Data Relations (HDR)*** [Chapter 7]. HDR extends the existing theories of Human-Data Interaction (HDI) (Mortier *et al.*, 2014), in order to invite and facilitate confrontation of its practical challenges (as well as the challenges of the other desires identified through the Case Studies). This involves discursive and grounded examination of the sociotechnical realities of held data and provider power in today's data-centric world. HDR can be considered as a framework to operationalise HDI, and its practitioners as a *recursive public* [7.6].

Drawing on the information theory background presented in 2.1.1, Chapter 7 differentiates the needs participants expressed from data (the stored artefact) and from information (the set of facts and opinions encoded within that data), modelling two distinct concepts of *life information* (from which people gain insights and extract value) and *ecosystem information* (from which people improve awareness and control over the use of their data). Combining these differing motives with the Case Study themes from Chapter 6 yields four clear objectives that the HDR research agenda aims to deliver for people:

- *data awareness and understanding*,
- *data useability*[11],
- *data ecosystem awareness and understanding*, and
- *data ecosystem negotiability*.

The practical pursuit of these objectives form the basis of Section IV of the thesis, with Chapter 8 mapping out the HDR landscape, identifying obstacles and insights that might help overcome them. Chapter 9 provides details of four identified trajectories for those who wish to improve Human Data Relations by pursuing systemic changes that can empower individuals with and through their personal data:

- *Discovery-Driven Activism* [9.3],
- *Building the Human-centric Future* [9.4],
- *Defending User Autonomy and Hacking the Information Landscape* [9.5], and
- *Teaching, Championing and Selling the HDR Vision* [9.6].

1.3 Publications Arising From and Connected to This Research

1.3.1 Pilot Study

My Doctoral Training programme at Open Lab began with a Masters in Research in Digital Civics. For my MRes project[1], I conducted a pilot study, interviewing and exploring issues around data with families who had experience of social care services. During the first months of this PhD, I conducted new analysis of previously collected data, resulting in the synthesis into a full first-author paper published and presented at CHI 2018:

- *Understanding the Family Perspective on the Storage Sharing and Handling of Family Civic Data* (Bowyer *et al.*, 2018)

This study is given a special status in this thesis; while the research was carried out prior to this PhD and thus does not form a core part of this thesis, it plays a critical role as a pilot study for Case Study One and its findings and insights are built upon in Chapters 4 and 6 and in later discussions. The paper is included in full in Appendix A.

1.3.2 Primary Case Studies

Publications from Case Study One

The work exploring shared data interaction in Early Help carried out in Case Study One has been initially published as an Extended Abstract at CHI 2019:

- *Human-Data Interaction in the Context of Care: Co-designing Family Civic Data Interfaces and Practices* (Bowyer *et al.*, 2019)

[1] MRes result awarded: Distinction.

This work was also presented at the conference in the form of a poster, which is shown in Figure 1.1. A journal paper is in prep.

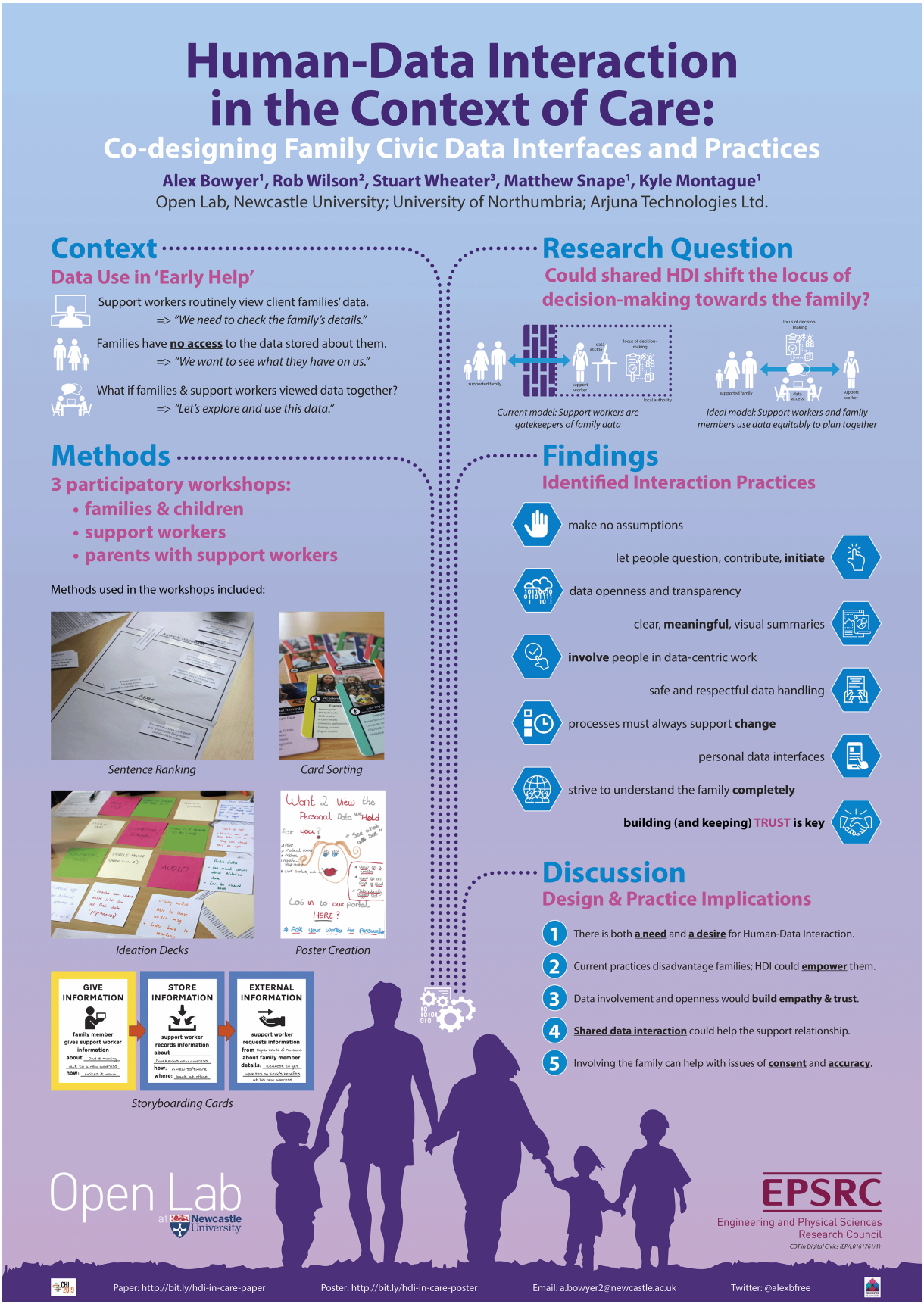


Figure 1.1: Poster Presentation of Case Study One at CHI 2019

Publication from Case Study Two

The work exploring the human experience of GDPR data access carried out in Case Study Two has been published and presented as a full first-author paper at CHI 2022, where it was awarded an *Honourable Mention*:

- *Human-GDPR Interaction: Practical Experiences of Accessing Personal Data* (Bowyer, Holt, *et al.*, 2022).

I carried out all field research myself. Data analysis and paper writing were jointly executed by myself and Jack Holt.

1.3.3 Workshop Papers, Presentations & Interviews

During the PhD, I gave a number of presentations and interviews and published three workshop papers. These outputs included material from, or directly contributed to, this thesis and its arguments.

- *Designing for Human Autonomy: The next challenge that civic HCI must address* (Bowyer, 2017) - a short talk I presented to my peers at Open Lab in January 2017 laying out the landscape of reduced agency and possible avenues for improving humans' relationships to their data that my PhD would explore;
- *Free Data Interfaces: Taking Human-Data Interaction to the Next Level* (Bowyer, 2018b) - a CHI 2018 workshop paper formalising my pre-PhD design thinking and outlining a vision for unconstrained and useful data interaction interfaces;
- *A Grand Vision for Post-Capitalist HCI: Digital Life Assistants* (Bowyer, 2018a) - a CHI 2018 workshop paper where I imagined a form of digital computer assistant that is far more helpful and human-data-centric than the digital voice assistants of today;
- *Personal Data Innovation: Sharing Power to Build Trust* - A guest interview on PDS company digi.me's blog in November 2019;
- *Personal Data Use: A Human-centric Perspective* (Bowyer, 2020c) - a lecture about my research that I was invited to give to undergraduate students at both Northumbria University and Newcastle University in early 2020, just prior to the pandemic;

- *My Thesis in 3 Minutes: Understanding and Designing Human Data Relations* (Bowyer, 2021b) - my entry into Newcastle University's 3-minute thesis competition in April 2021, for which I was co-winner of the people's choice prize;
- *Human-Data Interaction has two purposes: Personal Data Control and Life Information Exploration* (Bowyer, 2021a) - a workshop paper I presented at CHI 2021, introducing my model of the two motivating factors for interacting with personal data;
- *Conversation with Alex Bowyer: Understanding and Improving Human Data Relations* (Bowyer and Bourgeois, 2023) - a podcast interview in January 2023 with the Data-centric Design Lab at Delft University of Technology about the Case Study Two GDPR paper and my PhD research in general; and
- *The Power of Personal Data: How Companies Exploit Your Data to Shape Your World* (Bowyer, 2022b) and *The Power of Personal Data: Taking Back Control of our Data and Challenging Powerful Corporations* - (Bowyer, 2023) a two-part series of layperson-friendly talks about my research I gave at my local co-working space in September 2022 and January 2023.

1.3.4 Secondary Publications

During the same timeframe as this PhD and as resulting from the practical design and innovation work undertaken through the external projects [Section IV], I contributed to a number of publications:

- As a researcher and developer on the SILVER project [see Section II Introduction and Section IV Introduction], my work contributed towards an internal report to CHC as well as the overall impact report (Connected Health Cities, 2021, pp. 129–130).
- Also for SILVER, I published demonstration videos (Bowyer and Wheater, 2017) of a health data interface prototype developed by myself and Stuart Wheater.
- I was co-author on research relating to improving fast food platform interfaces using adversarial design to encourage healthy eating. This was published at BCS 2021 (Goffe *et al.*, 2021) and in the Interacting with Computers journal (Goffe *et al.*, 2022).
- As a research intern at BBC R&D [See Section IV Introduction], I published an internal design research report (Bowyer, 2020a) into personal data store design, and

wrote and presented a 'stimulus presentation' on personal data design innovation for members the project team, to launch an internal hack week.

- At Hestia.ai, I was a lead author on a research report auditing the data economy (Bowyer, Pidoux, *et al.*, 2022), and co-author on a research report on power mechanisms in the data economy (Pidoux *et al.*, 2022).

1.4 The Structure of this Thesis

The overall structure of this thesis is illustrated in Figure 1.2. Activities from the two tracks of participatory co-design research and grounded project design work are shown at the bottom and top respectively, along with arrows showing how those activities relate to the different parts of the thesis and how some of the chapters relate to each other. The five thesis sections described at the start of Section I above are indicated in distinct columns.

Section I continues from this Introduction chapter with a literature review and a methodology chapter, which also introduces the two Research Questions RQ1 and RQ2. Section II describes the research contexts chosen and details the two Case Studies which constitute the participatory research track of the thesis. In Section III, findings and insights from both Case Studies are synthesised to explain, in answer to RQ1 and RQ2, what people want from data and from data holders. The pursuit of such better relationships is then formalised into a new research agenda, *Human Data Relations (HDR)*, with four specific strategic approaches to change. Section IV describes the grounded industrial/project design work undertaken to operationalise this research agenda, mapping some of the key obstacles to HDR progress before laying out detailed designs and approaches. The thesis concludes in Section V, where findings and insights are distilled into thirteen Principles for generative action, before a reflection upon the research in terms of its limitations and pragmatic challenges, and on the PhD in terms of its legacy and the author's personal experience. Further detail on the content of each section is provided below.

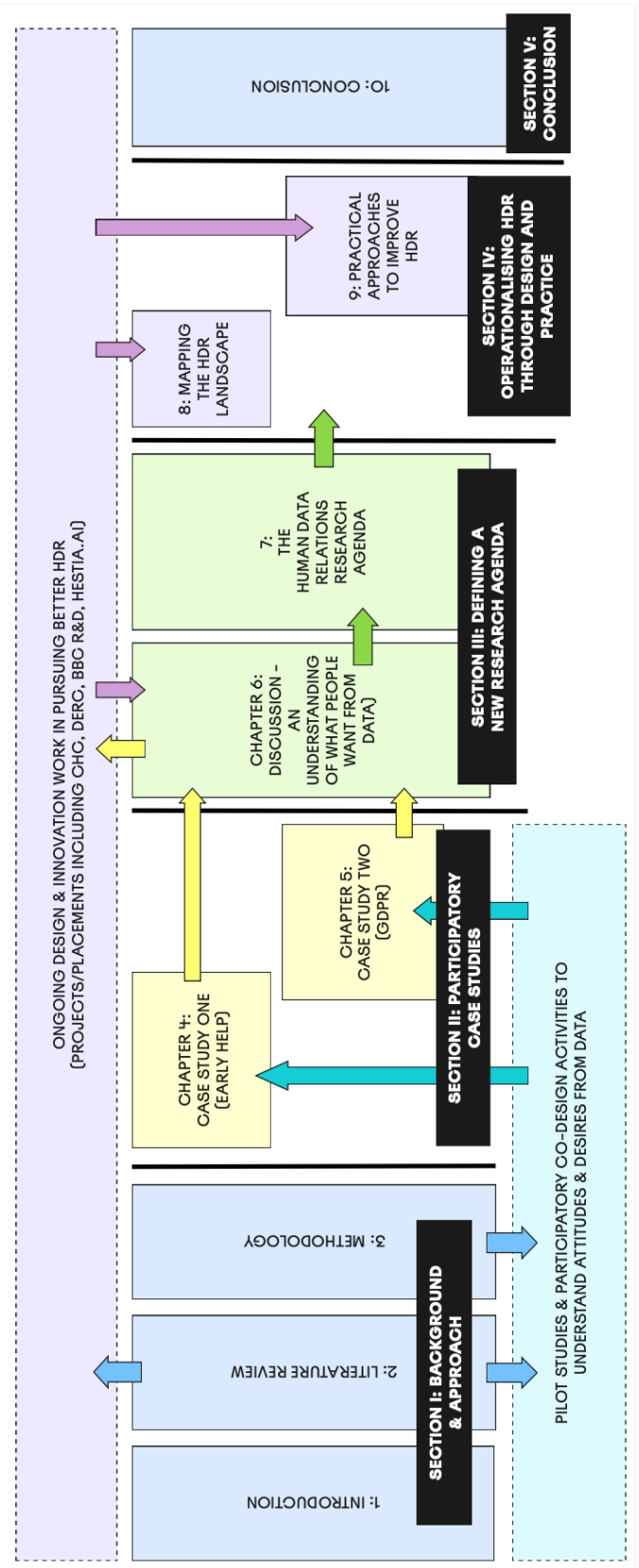


Figure 1.2: The Structure of This Thesis

Remainder of Section I: Introduction, Background and Approach

Chapter 2 contains a literature review. The first part [2.1] examines the difference between data and information, outlines the central role data has taken in our society, why people need effective access to their data and how laws have been introduced to try and deliver this. The second [2.2] serves as history of personal data interaction, from *Personal Information Management* to the emergence of complex digital lives involving relationships with many data-holding providers. Finally, 2.3 charts a path from *HCI* and *HDI* foundations through to the embracing of sociotechnical thinking around data and the current bleeding edge (Collins Dictionary, no date) of human-centred innovation, leading to the primary Research Question (RQ) of this thesis:

"What relationship do people want with their personal data?" [2.4]

Chapter 3 describes the methodology used in this research, explaining first the constructivist ontology and pragmatist, individualist epistemology behind the approach [3.1]. Then, the choice of the dual-track participatory co-design research and embedded design and innovation work in line with a *Digital Civics* agenda is explained [3.2]. The RQ is split into two—RQ1 and RQ2 [3.3]. Finally, the specific methods and techniques adopted in the research are explained and illustrated, including workshop activities, sensitisation, stimuli and recruitment [3.4].

Section II: Understanding Attitudes and Desires through Participatory Case Studies

After an introduction which explains the research contexts for the participatory work and the choices of Case Studies, Section II begins with Chapter 4 which reports on Case Study One. 4.1 is a detailed introduction to the UK's Early Help social care context, including its history of data-centrism which inherently contradicts the empowerment goals of Early Help. This makes it an ideal setting to explore the RQs. After explaining activities and methodology in 4.2, 4.3 introduces prior findings on family and staff perspectives, motivating the *Shared Data Interaction* vision and workshop design. The key findings are presented [4.4] then discussed [4.5] in terms of involving people with their data, effective data access, and shifting the Locus of Decision Making.

Chapter 5 reports on Case Study Two. 5.1 contextualises data access in light of the GDPR. 5.2 summarises the activities and methods adopted, before 5.3 explains the human-centric

approach to this study. Findings are presented in [5.4], which shares quantitative outcomes based on analysis of participants' GDPR requests, interview responses and participant-assigned scores. Key themes from qualitative analysis of interviews and observations follow in [5.5]. The discussion [5.6] then derives GDPR-improving design guidelines for policymakers, data holders and individuals, in line with a human-centric philosophy.

Section III: Defining A New Research Agenda: Human Data Relations

Following its introduction, Section III begins with a discussion chapter, Chapter 6. This synthesises the two Case Studies, and answers RQ1 [in 6.1] and RQ2 [in 6.2] with clear statements about what people want from data—visibility, understanding and useability[11]—and from data holders—transparency, oversight and involvement. 6.3 concludes the chapter in line with this thesis's purpose to address the power imbalance over personal data; it positions these six 'wants' as desirable *empowerment* relative to that perspective, motivating their practical pursuit.

Chapter 7 builds upon that motivation, shifting the narrative of the remainder of thesis towards a practical focus on how human-centric empowerment might be achieved. The thesis' findings are synthesised, drawing on experience from external work and relevant information theory, to formally define a new research agenda called *Human Data Relations (HDR)*, whose practitioners act as a *recursive public* [7.6], pursuing four objectives [7.5] for increased awareness, understanding and negotiability.

Section IV: Operationalising the Research Agenda through Design & Practice

Section IV begins with an introduction to the placements and projects undertaken by the author, through which the parallel track of design and innovation work to operationalise the HDR research agenda was progressed; these contexts provided the grounded learning for the insights presented in this section. Then, Chapter 8 maps out the landscape for the HDR research space, focusing on identifying obstacles to the pursuit of the HDR objectives.

Chapter 9, using a *Theories of Change (ToC)* framing, introduces opportunities for progress, arranged as four different trajectories of change that could be executed to pursue better HDR. These approaches are illustrated with designs and illustrations from the practical work in order to explain possible strategies.

Section V: Conclusions and Outlook

Section V concludes the thesis through Chapter 10. First, in 10.1, a collection of thirteen design Principles for generative action are asserted and illustrated. These Principles emerged through the HDR operationalisation work documented in Section IV. 10.2 provides a pragmatic discussion of the practicalities of pursuing HDR, identifying some additional pitfalls, assumptions and implications. 10.3 takes a critical eye to look back at this body of research, identifying limitations as well as opportunities for future work. To conclude the narrative of this thesis, 10.4 first reflects on this researcher's journey, before 10.5 summarises the legacy and contributions of this body of work, positioning HDR and this thesis as call to arms for activist research and design innovation to tackle the power imbalance around personal data in society.

Chapter 2. Literature Review

> *"We are the sources of surveillance capitalism's crucial surplus: the objects of a technologically advanced and increasingly inescapable raw-material-extraction operation."*
>
> —Shoshana Zuboff
> (Harvard professor, social psychologist and author)

2.1 Understanding Data and its Role in Today's World

2.1.1 What is Data?

Data is an oft-used word that carries multiple meanings. In everyday speech, it might refer to mobile phone bandwidth, a filled application form or a collection of digital files. Even experts have a variety of definitions of data, as well as the related concepts of information and knowledge (Zins, 2015). In this study, we refer to data by its accepted definition as information or knowledge stored in a form suitable for computer processing:

> *"Data is the representation of concepts or other entities, fixed in or on a medium in a form suitable for communication, interpretation, or processing by human beings or by automated systems."*—Wellisch (1996)

This definition is useful, as it includes the fact that both humans and algorithms can use data, and that data is something that needs interpretation.

A note on the word itself: From a strict grammatical stance, *'data'* is a plural of the singular *'datum'*, so it is more correct to write 'the data are correct'. This usage, however, is rapidly declining from use ('Data', no date). Throughout this thesis I use the more widely adopted usage of treating data as a singular mass noun, as in 'the data is correct'.

The concepts of *data* and *information* are closely related, so much so that they are sometimes used interchangeably. Ackoff presented a model for distinguishing data, information, knowledge, understanding/intelligence and wisdom (Ackoff, 1989). He describes data as the physical symbols, effectively the 1s and 0s stored in a computer or the ink marks on a page, which becomes useful when humans or algorithms are able to deduce

facts from those symbols to answer simple questions–at this point it becomes information. Being able to interpret deeper how and why questions allow information to become knowledge and understanding, towards the ultimate goal of wisdom. This is often represented as the *DIKW pyramid* (DIKW being shorthand for the data-information-knowledge-wisdom transformation that occurs as you move up through the layers), the origin of which is unknown (Wallace, 2007). Figure 2.1 builds upon a representation by George Pór (Pór, 1997) of the pyramid as a *wisdom curve*, showing how increasing meaning and value can be obtained from data as deeper questions can be asked of it. This theme of obtaining meaning and value from data is an important concept that will be referred back to.

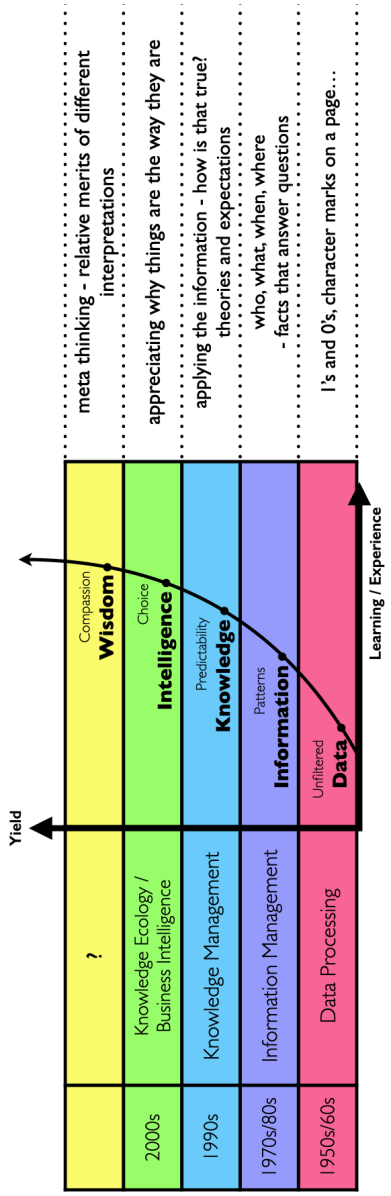


Figure 2.1: The Wisdom Curve: Making Data into Meaningful Information

This model that turning data into information can be thought of as using that data to answer questions is embodied in the definition that:

"Information can be thought of as the resolution of uncertainty"

The exact origin of this definition is unknown ('Information', no date). It is often attributed to mathematician Claude Shannon (Shannon, 1948). From an etymological stance, one who is informed is one who has received knowledge or concepts as a result of what has been communicated to them. Data, therefore, is the material from which information can be

received. It follows also that data contains uncertainty that must be resolved in order for it to become meaningful information.

2.1.2 Personal Data & the Rise of Data-centrism

The earliest computer systems used data to store mathematical and scientific facts. Data processing allowed for previously manual operations to be performed with greater speed and accuracy, most famously the work of Alan Turing and the case of the Enigma code breakers during World War II (Hutton, 2012). This work was the advent of general-purpose computing—machines that could be applied to *any* problem, provided you could reduce that problem to data. Businesses over the following decades began to apply computers to myriad new problem areas in all different fields of work and life, and doing so began the encoding of information about people as data, be it for statistical purposes like censuses or research, or simply to enable the more efficient serving of customers by storing databases of customer records.

The personal computer revolution ('The personal computer revolution', no date) of the late 1970s and 1980s put computers in every office and eventually every home too. It soon became commonplace that each individual would have data stored about them in companies' databases. In the subsequent years three factors have combined to accelerate this trend of storing data about people:

i) Labour costs have remained high. Companies have sought ways to automate their businesses and to implement online services and call centres in place of in-person staff interaction;
ii) Computer processing and storage has become ever cheaper thanks to the advent of cloud computing. Business processes could subsequently be reduced to data processing tasks or entire businesses be moved online; and
iii) The rise of smartphones and web-enabled devices has meant that the public are now ready and willing to conduct much of their daily business online through the web and apps.

These factors have encouraged both commercial and civic providers to centralise their services and to 'go digital' to the greatest degree possible. In doing so they collect ever more data about people (now *service users* or just *users*). Data is now seen as a resource which

can be mined for value, and harnessed for profit and business efficiency - 'the new oil' (Toonders, 2014). Zuboff, in her 2019 book on *surveillance capitalism*, characterises this new digital world as the collection of human behaviour data so that it can be used as free raw material and converted into profit through hyper-personalised advertising and targeting by software platforms (Zuboff, 2019). This philosophy is also known as *data-ism* (Brooks, 2013) and the analysis and exploitation of such data at scale is known as *big data* (Neef, 2015).

As a result of data-ism, the collection of data about people has become an inevitable part of modern life. We live *digital lives* (Various Authors, 2018) where we each interact directly and indirectly with hundreds of digital systems every day–as you shop, socialise, or browse online; as you listen to music or watch TV; as you interact with governments or healthcare services; as you travel, and in many more situations. Every one of those interactions indicates the presence of data about you stored in a company database. Every aspect of our lives involves the input, processing and output of data—either provided by, collected from, or generated about, us. And the digital data we create and consume (whether consciously or not—data sharing is often unwitting (Crabtree and Tolmie, 2018)) has a direct influence on our lived experience. Data affects decisions about what we are entitled to, what opportunities we will be offered, and the advertisements and content recommendations we are shown while we browse.

In 2017, the average American Internet user had 150 online user accounts with different providers (Caruthers, 2018). Data for the UK shows that each individual has to manage an increasing number of service and supply relationships, rising from around 45 in 1997 to around 250 in 2020 (Henderson and B-s Working Group, 2020). As the amount of personal data relating to each of us has increased, the need for individuals to be able to manage this has grown. Unfortunately, the large-scale systems which collect data about us now function as *data traps* (Abiteboul, André and Kaplan, 2015). Data about us is easily gathered but very hard to remove or even to access. This creates a lack of agency for the individuals living in this data-centric world. The World Economic Forum's *Rethinking Personal Data* project recognised the critical role that data, (specifically *personal data*—data created by and about people [ARI2.1]) now holds, and identified that:

> *"An asymmetry of power exists today [...], created by an imbalance in the amount of information about individuals held by industry and governments, and the lack of knowledge and ability of the same individuals to control the use of that information."*—World Economic Forum (Hoffman, 2011, 2013, 2014b, 2014a).

2.1.3 Legislating to Protect Personal Data: The Story So Far

Since as early as 1973, the need to protect individuals' rights over their data has been recognised (US Department of Health Education and Welfare, 1973). The 37-nation organisation OECD in 1980 stated that:

> *"The right of individuals to access and challenge personal data is [...] the most important privacy protection safeguard"*—Organisation for Economic Co-operation and Development (1980)

They also issued recommendations that individuals should be given basic privacy rights, including the right to be informed whether data is stored about them, and the right to an intelligible copy of that data. Over the subsequent decades, lawmakers began to enact laws to deliver these rights to individuals, notably:

- the UK's *Data Protection Act (1984)* (which set up an independent body, the Data Protection Registrar (now the Information Commissioner's Office (ICO)) with which organisations were required to register their usage of personal data);
- Ireland's *Data Protection Act (1988)* (which introduced the concept of a 'duty of care' for data collectors - that they are expected to avoid causing damage or distress to data subjects);
- the EU's *Data Protection Directive (1995)*; and
- the UK's *Data Protection Act (1998)*.

However, such laws were generally found to be ineffective. In 2002, Simon Davies, director of Privacy International, said that the UK's DPA was 'almost useless in limiting the growth of surveillance' (Millar, 2002), and research commissioned but the ICO in 2008 stated that the European Data Protection Directive was 'out of date, bureaucratic and excessively prescriptive' (McCullagh, 2009). During this period, a number of relevant terms have been established through data protection legislation, including a legal definition of personal data and legal terms such as *data controller*, *data processor* and *data subject* [see ARI2.1]. For

simplicity, this thesis will instead use more everyday terms such as *individuals* and *data holders*.

The 2010s saw in-depth discussions on how to reform data protection law across the European Union (Zanfir, 2015), and within its nation states, such as in Sweden (Svantesson, 2012) and Germany (DeSimone, 2010). In 2018, the EU's *General Data Protection Regulation (GDPR)* came into force, carrying with it significant designed-to-hurt fines for non-compliance (Kelly, 2020; Leprince-Ringuet, 2021). This legislation has finally given individuals the practical means to exercise their data rights ('The GDPR: Does it Benefit Consumers in Any Practical Way?', 2020) against a backdrop of massive personal data use across society that had previously rendered data access requests impractical (Cormack, 2016). The GDPR—which gives individuals key rights including rights to timely data access, explanation, erasure and correction (Information Commissioner's Office, 2018)—can be seen as the first serious attempt to rebalance power over data from organisations toward citizens. The GDPR is generally regarded as a landmark piece of legislation and a strong template for individual data protection. Around the world, companies have overhauled their privacy policies and updated their business practices to comply with the GDPR and other similar legislation, such as Japan's *Act on the Protection of Personal Information (2017)*, India's *Personal Data Protection Bill (2019)* and the *2020 California Consumer Protection Act (2020)* (CCPA). In the USA, there has been no national privacy law yet, but the GDPR's influence is being felt in court rulings (Hoofnagle, Sloot and Borgesius, 2019).

Social science researchers have highlighted the critical role that law must play in the protection of individual privacy, and continue to explore the complexities of this challenge (Kosta, Leenes and Kamara, 2022). Data protection regulation can be particularly challenging in a global context where each country's efforts affect those of others. Bennett and Raab, for example, speculate that regulatory interdependence between countries might lead to a search for the highest possible standard of privacy protection, but could also result in competitive deregulation of the personal data space (Bennett and Raab, 2017). In his book *'Surveillance Society'*, David Lyon highlights the inherent complexity of citizen surveillance in that it can easily slip between ideologies of 'care' and of 'control'. This raises important questions about who, ultimately should take responsibility for protecting the individuals in a society (Lyon, 2001). Daniel Solove describes privacy as 'a concept in

disarray' and calls for practitioners and policymakers to better discern four distinct aspects of privacy - *information collection*, *information processing*, *information dissemination*, and *invasions of privacy* (Solove, 2008). Orla Lynskey argues that data protection law is not about privacy, but about granting individuals some individual control over their data (Lynskey, 2015).

Following the Snowden revelations (Gellman, 2013) in 2013, attention and concern over personal data use has grown year on year. In 2018, the Cambridge Analytica scandal ('Facebook–Cambridge Analytica Data Scandal', 2014; Chang, 2018) broke; the personal data of 87 million people, acquired from Facebook, was exploited with the apparent intent of influencing voting outcomes including the UK's 2016 Brexit referendum and the USA's 2017 election of Donald Trump. This, combined with widespread public information campaigns about GDPR, have led to a heightened awareness of personal data rights (European Union Agency for Fundamental Rights, 2020). At the time of writing in 2023, personal data protection laws and individual digital rights remain a rapidly evolving area.

The summary of data protection legislation provided in this section does not attempt to be exhaustive. For additional context and critique on data protection law, how it has evolved and how it applies to today's digital landscape, the publications of David Erdos are recommended (Erdos, 2015, 2016, 2019, 2022).

2.1.4 The Need for Practical and Effective Personal Data Access

The World Economic Forum called in 2011 for a balanced ecosystem around personal data, and identified *transparency* as a key principle needed to achieve this: People need to know what data is captured, how it is captured, how it will be used and analysed and who has access to it. Additionally, people must understand the *value* created by the use of their data and the way in which they are compensated for this (Hoffman, 2011). It is almost impossible for people to assess that value, because they are unaware of most of their data (Spiekermann and Korunovska, 2017). Having *awareness* of your personal data is a critical first step, so that people might assess 'to what extent the bargain is fair' (Larsson, 2018). In this regard, the GDPR can be seen as an important step in the right direction, as it requires data controllers to document their data practices and to provide data copies.

However, it is not sufficient simply to grant data subjects the technical or procedural capabilities to see the stored records about them. Access must be *effective*. Every individual must have the knowledge, skills and structures in place that enable them to achieve their objectives with their personal data (Gurstein, 2003). Gurstein later identified seven aspects that are necessary for access to be effective (Gurstein, 2011) and to avoid a *data divide* of those who can harness their data and those who cannot:

- *Internet*: If data access is via Internet, then issues with affordability, bandwidth, network censorship, or disabilities limiting physical access to Internet devices or terminals would make access ineffective.
- *Computers and software*: Sufficiently powerful computers must be available, for a sufficient amount of time, with sufficiently capable software to perform necessary interpretation or actions.
- *Skills*: If technical skills or knowledge are required to use the software and/or to interpret, analyse or visualise the data, then access is ineffective for the layperson.
- *Content and formatting*: The data should be in an appropriate language and format to allow use at various levels of linguistic and computer literacy.
- *Sensemaking*: Information presentation should be as clear as possible so that people can interpret their data and extract meaningful information from it.
- *Advocacy*: People need support and training to make use of their data and representation if they are to use it at a wider community level.
- *Governance*: There must be financing and appropriate law or policy to support people's desired usage of their data.

Unfortunately, people's ability to derive value from their data, or to assess its value is limited; it is an asset over which we have little control. Our existing data 'resides in isolated silos kept apart by technical incompatibilities, semantic fuzziness, organisational barriers [and] privacy regulations'. This lack of effective data access is detrimental to trust, innovation and growth (Abiteboul, André and Kaplan, 2015).

Beyond these operational concerns over effective access, there are practical limitations affecting people's ability to make use of their data. Where people are given interfaces their data, access is typically via a list or feed combined with a search box. Studies have shown

that people prefer to find information by *orienteering* rather than search–associatively traversing related datapoints (Teevan *et al.*, 2004; Karger and Jones, 2006). Having our documents distributed across multiple platforms, applications and devices makes interrogation and orienteering hard (Krishnan and Jones, 2005).

Abowd and Mynatt highlight that in presenting information about people and their activities, everyday computing needs to recognise that users' activities rarely have a clear beginning or end, are often interrupted, are often concurrent with other activities. *Time*, therefore, is an important factor in finding and interpreting information. *Associative* modelling of information is more useful than hierarchical models, because future usage goals cannot always be anticipated (Abowd and Mynatt, 2000). Recognising these needs, Krishnan and Jones identified that an effective information access system should support giving historical context, finding trends and patterns, time-based contextual retrieval, automatic structuring and multiple perspectives of the information (Krishnan and Jones, 2005). Shneiderman, in the context of considering the effectiveness of interactive information visualisations, identified the need to support seven types of information interaction: *overview*, *pan & zoom*, *focus* (context & distortion), *detail on demand*, *filter*, *relate*, *history* and *extract* (Shneiderman, 1996). These capabilities exist in some data interfaces today. However, no general-purpose personal information access system exists today with all, or even most, of those capabilities. The development and state of the art in the field of *Personal Information Management Systems* is explored in section 2.2 below.

2.1.5 Research Gap: The Human Experience of Personal Data

2.1 described the establishment of the data-centric world in which we live today, the imbalance this creates between individuals and data holders, and the nascent attempts by governments to redress that imbalance through the creation of new laws. It has also shown where research thinking has exceeded the practical data capabilities we have today—identifying factors and capabilities that should be considered when giving people a meaningful relationship with their personal data.

To date, people's relationship with their personal data and the information within it has barely been explored. What mental models do people have around data? What value does it carry to them and what meaningful place does it (or should it) hold in their life? What is it

that makes data meaningful? What do people want from their data? What is it like to live in this data-centric world where your abilities over your data are limited by lack of access to data and a lack of suitable interfaces and technologies to properly manage your digital life? This is one aspect of the research gap this thesis will address—discovering the *human experience of data*.

2.2 A History of Personal Data Interaction

2.2.1 Computers as General-Purpose Information Tools

In the immediate aftermath of the second World War, Dr. Vannevar Bush wrote a landmark article for The Atlantic Monthly in which he envisioned a new scientific agenda for America and the world. By harnessing the new general information-processing capabilities of computers, he believed we could to make the stored knowledge of mankind accessible and useable to all, for the betterment of society. He proposed the *Memex*, a device in which people would store their books, communications and records digitally so that it 'might be consulted with exceeding speed and flexibility'—a personal filing system to serve as 'an enlarged intimate supplement to his memory'. He emphasised the importance of allowing information to be stored in 'associative chains of related materials' so that people would be able to retrieve information in the same way we think of it, traversing related items or ideas (Bush, 1945). During the next three decades, while computer systems were moving out of science labs and being established in workplaces as a means to automate and improve business processes, many researchers began to look beyond usage in business and consider how computers might be used by 'the common man':

- for interpersonal communication (Shannon, 1948);
- to model human thought (Simon and Newell, 1958);
- to augment human intellect (Engelbart, 1962); and
- to store one's personal information in digital files (Nelson, 1965).

Collectively, these constituted a recognition that the computer could be considered as tool that anyone could use for their own purposes. By the end of the home computer revolution ('The personal computer revolution', no date) of the 1970s and 1980s, it seemed that this power had been placed into the hands of ordinary people:

"Having reduced your affairs to software, software can take care of them for you."— Gelernter (1994)

2.2.2 Personal Information Management

Through the examination of people's desk-based working practices, researchers began to understand how people handle information to inform the design of computer information systems. In 1983, Thomas Malone observed that categorisation is hard, and that any system must not only help the user to find information, but also remind the user of things to do. Computers could help through automatic classification, but should also allow both physical and logical 'piles' of information to be arranged by the user (Malone, 1983). *Personal Information Management (PIM)* was first mentioned in 1988 by Mark Lansdale, who identified a need to design information management systems according to the psychology of the people who use them rather than by simulating office practices. By paying attention to how people categorise, recognise and recall information, and labelling information with appropriate attributes, information could be retrieved by its different properties (Lansdale, 1988). PIM includes directly interacting with digital files, webpages and e-mails as well as 'meta-activities' such as finding, arranging, searching, browsing, re-finding, categorising, sensemaking, keeping and discarding personal information. William Jones summarised PIM as 'the art of getting things done in our lives through information' (W. Jones, 2011a).

Driven in part by the pursuit of better *time management* in the late 20th century (characterised by PDAs, palmtops and electronic organisers) (Etzel, 1995) and the focus on personal productivity in the early 2000s (characterised by *GTD (Getting Things Done)* self-help books and to-do list software (Andrews, 2005)) and the continuing challenge of overcoming information overload in an increasingly digital world, PIM has been a thriving field both in research and in practice, with a peak in activity around the mid '00s. Since the 1990s, numerous PIM system designs have emerged, each exhibiting some of the following six traits: Spatial, Semantic, Networked, Temporal, Contextual and Subjective.

Spatial PIM Systems

Spatial PIM systems are based on the idea that people remember 'where' they have put things and that this allows information to be quickly retrieved by associating it with a place (Negroponte and Bolt, 1978), much as people keep current information *'in reach'* on a desk

(Klein *et al.*, 2004). Spatial approaches recognise that *keeping* is a valuable activity in its own right, that informs sensemaking (Marshall and Jones, 2006). Placed information also performs an important *reminding* function (Barreau, 1995; Barreau and Nardi, 1995).

Networked PIM Systems

Building on Bush's ideas of *associative chains* of related materials, *networked* PIM systems focus on the relationships between data. HyperText, as conceived in 1965 (Nelson, 1965) was designed to keep connections between information and allow the computer to understand what linked information *is*. The version of hypertext that underpins the World Wide Web we use today is much weaker than Nelson's HyperText or Berners-Lee's Semantic Web and does not achieve these goals, as the inventors agree (Ross, 2005; Nelson, 2006; Ziogas, 2020). In the absence of connected networks of personal information and with people collecting more information than they discard (Whittaker and Hirschberg, 2001), the 2000s saw software like *Google Desktop Search* ('Google Desktop Search', 2004) and *Infovark* ('Infovark Company Profile', 2007) emerge to try and discover users' data files and unify access to them, with limited impact (Bergman *et al.*, 2008). Around this time, Microsoft invented *WinFS*, a system to re-invent the modern-day operating system to be based upon relational structured data rather than file storage, but sadly it was never released ('WinFS', no date). Paul Dourish *et al.* proposed *Placeless Documents*, which relied on the idea of assigning user-specific properties to documents so that they could be arranged and recalled by their common properties rather than their location (Dourish *et al.*, 2000; Dourish, 2003).

Metadata—information about what the data *is*—is critical to information organisation (Foulonneau and Riley, 2008). One of the more advanced networked PIM systems is the Networked Semantic Desktop, which recognises that critical metadata is lost when files are copied or emailed, and attempts to maintain metadata and traceability by integrating PIM with *peer-to-peer (P2P) technology* (Decker and Frank, 2004). Tags, which emerged as a means to organise data through systems like *del.icio.us* ('Delicious', 2003) and *Flickr* in the 2000s, are still widely used on social media and websites today, and are even available within macOS (Frost, 2019). Tags can be seen as a continuation of attempts to attach metadata to personal data to give it meaning, even though the dream of *folksonomies* has not been fully realised (Abbattista *et al.*, 2007; Terdiman, 2008).

Semantic PIM Systems

Semantic PIM systems (a.k.a. 'The Semantic Desktop') take the idea of metadata even deeper and focus on what the information *means*. The idea is to present an integrated view of a person's stored knowledge by representing their documents, data and messages as URL-addressable semantic web resources (Sauermann, Bernardi and Dengel, 2005). The focus is on both the retrieval of documents and of facts (Schumacher, Sintek and Sauermann, 2008). This implicitly means that the computer must know more about what the data it stores represents, elevating it from number cruncher to something that holds a collection of information about the world. Hendler and Berners-Lee see semantic web technologies as the building blocks for a new age of *social machines* (Hendler and Berners-Lee, 2010), machines that operate in society at an information level. This desire to give computers greater understanding of data has created emergent industries focused on using linguistics and statistics to perform content analysis, text mining and information extraction (Hotho, Nürnberger and Paaß, 2005). It has even been proposed that AI might help computers to understand users' mental models (Nadeem and Sauermann, 2007).

Temporal PIM Systems

While folders have emerged as the dominant means to organise computer files and are effective because they allow you to arrange information according to its meaning to you (Bergman *et al.*, 2012; Bergman, 2013), supporters of *temporal* PIM systems argue they are inadequate as an organising device. Freeman and Gelernter proposed *Lifestreams*, a PIM system based on the principled that storage should be transparent, archiving and compatibility should be automatic, and concise overviews of groups of related information should be available (Freeman and Gelernter, 1996). Central to this system is the idea that personal data can most easily be navigated when viewed as a *timeline*. Almost all data can be associated to a specific time. This maps also onto the idea of relating personal information to human memory (Lansdale and Edmonds, 1992). *TimeSpace* provides another model of a PIM system that organises personal information by both time and the user's own activities, supporting interaction with a 'continuously changing and evolving information space' (Krishnan and Jones, 2005).

Time-based PIM approaches also coincide with a drive to move beyond files as a system of information storage. Gelernter believed we should not have to put effort into organising files, and argued somewhat prophetically that commercial factors have skewed personal data systems design away from the realities of human lives (Steinberg, 1997). In my own 2011 article *Why files need to die*, I mapped out how a personalised timeline could allow better personal information organisation and retrieval (Bowyer, 2011). Echoing this as well as Decker's desire (Decker and Frank, 2004) to maintain an information trail for every piece of information, Siân Lindley *et al.*, having called for time to become a subject of design research in its own right (Odom *et al.*, 2018), explored the concept of the *file biography*. This allows the history of information to be kept as the file is used and changed. File biographies tell a story, and help to reconfigure our thinking away from mindsets around copying, deleting and sharing, to view information as fluid (Lindley *et al.*, 2018). Moving into the world of online information collaboration, *activity streams* can also be seen as a recognition of the importance of tracking data as it changes, and offer new affordances (Hart-Davidson, Zachry and Spinuzzi, 2012).

Contextual PIM Systems

In 1995, Barreau highlighted the importance of *context* to PIM; People need access to different information according to what they are doing (Barreau, 1995). In 2000, Abowd and Mynatt highlighted the importance of paying attention to the user's context in order to offer access to the most relevant information and features, and they suggest context can be identified by considering the *'5 W's' - who, where, what, when and why'* (Abowd and Mynatt, 2000). *Context-aware computing* (Abowd *et al.*, 1999; Eliasson, Cerratto Pargman and Ramberg, 2009) has subsequently emerged as a sub-discipline of research in its own right [Dey (2001); 2.3.2]. Dourish identified that context is both a problem of representation, in that it is information that can be captured and represented, and of interaction, in that it is a relational property between objects or activities. He calls for *embodied interaction*—allowing users to create their own practices and meanings in the course of their PIM system interaction. Context is not objective and predetermined, it arises from the activity (Dourish, 2004). You need different organisations of information in different contexts. This means that PIM systems need to support representing a given set of information in different ways (Lansdale and Edmonds, 1992). More that than, that different information should be shown

according to the current context; different *perspectives* are needed to segment your life. TimeSpace uses *activity workspaces* to achieve this (Krishnan and Jones, 2005), but Karger *et al.*'s Haystack system refines the concept further, introducing the concept of *lenses*. Perspectives change which information records are included, whereas lenses allow you to focus on different attributes of what might be the same or different information (Karger *et al.*, 2005). Using a similar premise, Jilek's *context spaces* system attempted a dynamically reorganising contextual sidebar, but is limited in flexibility as it uses rigid types for specific contexts (Jilek *et al.*, 2018). Lindley observes that different information abstractions are needed for different audiences, from which we can infer that in a multi-user system, no single arrangement of information will suffice because in the same context two people may have different needs (Lindley *et al.*, 2018).

Subjective PIM Systems

This is why the sixth trait of PIM systems is important: *subjectivity*. Information organisation cannot be handled in a deterministic, objective manner. Any PIM system must be tailored to, and adaptable by, the user. Shipman and Marshall found that forcing users into explicit information models or workflows is harmful to user experience, and that interactive systems have to address the challenge of being just explicit enough but still allowing for differences in individual mental models (Shipman and Marshall, 1999). Bergman *et al.* (Bergman, Beyth-Marom and Nachmias, 2003) proposed three principles for subjective PIM, and their 2003 assertion that these principles are not currently well implemented in PIM systems remains true today:

- the *subjective classification principle* - all related items should be classified together regardless of technological format
- the *subjective importance principle* - the subjective importance of information should determine its degree of visual salience and accessibility
- the *subjective context principle* - information should be retrieved and viewed by the user in the same context in which it was previously used

Teevan's take on PIM subjectivity is important:

> *"The user should feel in control of the information"*—Teevan (2001)

She argues that this can be done by 'understanding what *conceptual anchors* the user creates and keeping them constant while the data changes'. With semantic PIM systems, we can see that a successful system (or at least, its designers) must understand a great deal about its users.

2.2.3 Self Informatics (SI)

In the late '00s, researchers and enthusiasts took PIM beyond task management and turned PIM thinking toward the self. In pursuit of Bush's vision of augmenting human memory, Jim Gemmell and Gordon Bell in their *MyLifeBits* project at Microsoft (Gemmell, Bell and Lueder, 2006; Bell and Gemmell, 2009) tried to capture an entire life electronically. This became known as *lifelogging*: gathering as much data about one's daily activity as possible, so that the maximum possible context, detail and understanding can be gained.

In 2007, tech writers Kevin Kelly and Gary Wolf set out a vision for what they called *the Quantified Self*, that is, to achieve increased self-knowledge through self-tracking. This included not only physical metrics such as step counts, heart rates or calories burned, but almost any aspect of one's life that could be numerically recorded in a computer (Kelly and Wolf, 2007). The Quantified Self movement (QSM) is now a world-wide community of enthusiasts who have developed hundreds of tools and techniques for self-tracking/lifelogging and monitoring themselves through data for the purposes of self-improvement. It also has a non-profit organisation aiming to 'advance discovery through increasing access to data' ('About the Quantified Self', no date).

Around 2009, researcher Ian Li began writing about what he called *personal informatics*, noting that it can be difficult to know ourselves due to incomplete self-knowledge, difficulties in monitoring our own behaviours, and being too busy to introspect. He proposes that:

> *"Computers can help: They can store large amounts of data, analyse the data for patterns, visualise the data, and provide feedback at opportune times"*—Li (2009)

Just as QSM has gained traction with enthusiasts in the general public, so personal informatics has grown as an area of research, development and study in academic circles. While QSM and lifelogging focus slightly more on capturing data about oneself and personal informatics focuses slightly more on the mechanisms of integrating and reviewing self-

tracking data, there is so much overlap that all three can be considered the same field. For convenience, I will refer to their union by the shorthand *self informatics* (SI) throughout this thesis. SI can be seen as a distinct advancement from PIM because of its focus on *using* personal information for personal benefit. SI can be seen as the antithesis of corporate data-centric motives [2.1]. Here, data is gathered for the data subject's benefit rather than that of the data-gathering organisation.

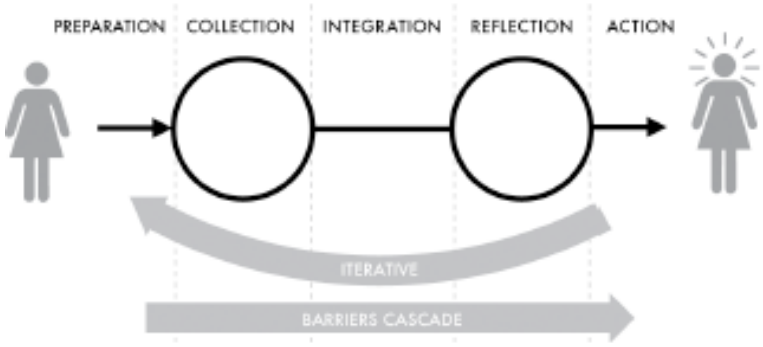


Figure 2.2: Li et al.'s Stage-based Model of Personal Informatics Systems (Li, Forlizzi and Dey, 2010)

Li, Dey and Forlizzi conducted participatory research with SI practitioners and identified five stages of personal informatics systems. These refine William Jones' list (W. Jones, 2011b, p. 72) of the six activities involved in PIM. The five stages, illustrated in Figure 2.2, each of which can be driven by the user, the SI system or both, are:

- *preparation* - motivating oneself and deciding what to collect,
- *collection* - recording or capturing subjective and objective data manually or automatically,
- *integration* - combining, organising and transforming the data so that it can be interpreted as needed,
- *reflection* - reviewing, exploring, interrogating and considering the combined and analysed information, and
- *action* - where, armed with a new understanding of oneself, the data subject may change behaviours or set new goals.

Of these, reflection is perhaps the most important, as the capacity to gain new insight is the motivating reason to engage in SI. *Reflective learning* (Boud, Keogh and Walker, 1985) has

been recognised as a valuable means of knowledge acquisition and improvement in a variety of contexts including education (Dewey, 1938), business (Beck *et al.*, 2001), and research (Lewin, 1946). In the context of the wisdom curve (see Figure 2.1 above), reflection can be seen as asking questions of data in order to acquire knowledge about oneself. Knowledge about oneself (a.k.a. self-insight (Hixon and Swann, 1993)) serves not only to satisfy curiosity (Li, Forlizzi and Dey, 2010) but can improve self-control (O'Donoghue and Rabin, 2001), increase self-awareness (Aslam *et al.*, 2016) and enable positive behaviours such as saving energy (Seligman and Darley, 1976).

Reflection can be facilitated in SI systems by enabling the tracking of subjective factors such as mood or intent as well as objective factors such as health or activity. It can be triggered by means of notifications, or during more direct information exploration by the user as they recall or revisit experiences (Rivera-Pelayo *et al.*, 2012). To aid interpretation of data by SI users, *contextualisation* enhances information with additional facts to ease its comprehension. This can include social, spatial or historical context, subjective or objective metadata or external sources of information (e.g. weather) (Rivera-Pelayo *et al.*, 2012), or external devices (Dey, 2000). There are two phases of reflection. During the initial *discovery* phase, typical questions that SI users ask concern:

- the *history* of data changes;
- understanding the *context* of a datapoint;
- the *factors* that cause a pattern in data; and
- the identification of suitable *goals* to pursue.

During the *maintenance* phase, these goals frame the questions asked, which concern:

- *status* (how well you are doing at meeting your goals) and
- *discrepancies* (examining the difference between actual behaviour and desired behaviours).

In order for a SI user to successfully reach this maintenance phase where they can continue to reflect upon their actions and adjust their goals, they must have been able to successfully navigate each of the 5 phases illustrated in Figure 2.2. If they have not collected the right data, they cannot integrate it. If they have not been able to integrate the collected data in a

meaningful way, they cannot reflect upon it. And so on. Li *et al.* framed this the *barriers cascade* (Li, Forlizzi and Dey, 2010). The pursuit of new ways to overcome these barriers has in effect been the major problem space for all SI approaches. This is especially evident in the QSM (Choe *et al.*, 2014).

While effortless SI is not yet a reality and many barriers still exist, progress in easing the SI journey through the barriers cascade is being made: in 2011, Jones had noted that people often postpone or don't have time for meta-level information management activities (W. Jones, 2011a). By 2019, the increased automation around self-tracking and data collection was judged to have given people more free time and energy for reflection and managing their goals (Feng and Agosto, 2019).

2.2.4 Overwhelmingly Complex Digital Lives

The rise of data-centrism has meant that every aspect of our lives now involves digital service providers and products which process personal data [2.1.2]. Smartphones put computers in everyone's pockets, and cheap cloud computing and an open web allowed every organisation to serve the population digitally through apps and websites. In 2010, broadband access was declared a legal right in Finland ('Finland: Broadband Access Made Legal Right in Landmark Law', 2010). In 2011, the UK Supreme Court declared that Internet access was an 'essential part of everyday living'. Denial of Internet access for criminals such as sex offenders was ruled unlawful (Roche, 2011; Wagner, 2012). Everyone now required access to information and online digital services.

> *"The boundary between real life and online has disappeared"*—Burkeman (2011)

The promise that whatever you want to do 'there's an app for that' had become true (Apple, 2009). During the late '00s and throughout the 2010s data-centric companies disrupted almost every industry, for example: Amazon (shopping & books), Uber (taxis), Netflix (movie rental), Spotify (music), AirBNB (accommodation), Google (email, news & advertising), Facebook (social networking & advertising), Paypal/Revolut/Monzo (banking), match/Tinder (dating), Rightmove (estate agents), Steam (computer games), Just Eat (takeaways), and many more (Levine, 2011; Carter, 2015). As a result, we now produce rich data trails simply by going about our daily lives, and this has become 'the driving force for value creation' online (Symons *et al.*, 2017). More recently as we start the

2020s, the trend has accelerated, with the COVID-19 pandemic driving both information work and many social activities to move online using platforms such as Zoom, Google Docs and miro (O'Donnell, 2020).

Throughout the transition to this information economy, the computing industry has delivered revolutionary new capabilities, but with every provider offering their own apps and websites, the information landscape has become hugely challenging for people to manage; information overload is now a serious problem that has been linked to increased anxiety, impaired critical thinking, exhaustion, and loss of willpower and focus (Hemp, 2009; Tunikova, 2018; Fu *et al.*, 2020). By 2006, our personal information was fragmented and a unified interface was called for:

> *"We must launch multiple applications and perform numerous repetitive searches for relevant information, to say nothing of deciding which applications to look in."*— Karger and Jones (2006)

With today's silo-ed Internet and our myriad devices and apps, this has only got worse. Bergman's subjective principles (see above) imply that our data should be able to move and be referenced freely, but it cannot. Our ability to share and connect data is limited (Crabtree and Tolmie, 2018). Our data is *trapped* (Abiteboul, André and Kaplan, 2015), not only because it is held by organisations without giving us effective access, but also by various practical means such as format incompatibilities, device restrictions, paywalls, and a lack of data portability. We need to free our data, as I expand upon in Bowyer (2018b).

It is clear that general-purpose computing has yet to provide people with the tools to manage their complex digital lives. There have been attempts to create general-purpose interfaces for personal data, typically based around a timeline, such as *AllOfMe.com* ('AllofMe Company Profile', 2007; 'AllOfMe.com Teaser Clip', 2008) in 2008 and *myTimeline* a decade later ('myTimeline', 2018); however, none of these products have reached public availability. To date the closest market-successful tool that people have for general purpose information handling is *Facebook*, given that it can store personal information, handle asynchronous and instant messaging, news, photo sharing, some retail functionality, brand interaction & support, calendaring and event management, and group discussions. However, it is a closed system with no capability for customisation; none of its content is available outside the network and external content cannot be linked or interacted with

except by import; as such it cannot be considered a PIM system. Its own *Timeline* feature, promoted at launch in 2011 as 'the story of your life' and 'a new way to express who you are' (Siegler, 2011) has been retired, along with many other tools designed to make information easier to manage, such as personal news feeds and friend lists (Perez, 2018). This reminds us that Facebook exists primarily to serve its advertisers, rather than the general public, as per the often-repeated saying 'if you're not paying for it, you are the product'. A promising area for the development of interfaces for managing digital lives is the emerging *personal data locker* space, explored more in 2.3.4 below. These offer the promise of *a place for personal data*, as Jones imagined PIM should be (W. Jones, 2011a), though at time of writing are still quite limited. In 2015, Abiteboul noted that:

> *"Everyone should be able to manage their personal data with a personal information management system."*—Abiteboul, André and Kaplan (2015)

As of yet, in any meaningful or holistic way, they cannot, because no general-purpose personal information management system for modern day digital lives exists.

2.2.5 Research Gap: The Data Beyond the Individual

2.2 detailed the ways in which personal information management have developed, and shown that technology has not kept pace with the ever-more-complex needs of the Information Age. Most PIM systems treat data as a static resource to be filed and accessed much like you would a file in a 1970s office. Now that data is almost anything, this model no longer fits. Most digital services operate in isolation from each other, without any shared representation or co-operative understanding of an individual's personal information. Where personal data access is provided, it is limited in usage to the delivery of the specific service on offer, it is treated as a property asset of the provider, and the data is not *participatory*. Katie Shilton wrote that:

> *"Much of the social impact of participatory personal data will depend on how data are captured and organised; who has access; whether individuals consent and participate; and how (or whether) data are curated and preserved."*—Shilton (2011)

We need 'fundamental changes in the way we represent and manipulate data' (Karger and Jones, 2006). We need holistic representations of data that can be subjectively meaningful and which allow for the constant change and evolution of data over time.

Of particular importance is that we recognise that people exist in an interconnected world of relationships - with other individuals, and with organisations, and that the role of data within those relationships needs to be examined. When your data is held by others, managing personal information is not just a matter of arranging your own bookshelves, but rather a multi-party negotiation over representation, ownership, access and consent. Data is a shared resource with multiple users, and only a few researchers have begun to look at people's interactions with data in this context (for example, activity streams (Hart-Davidson, Zachry and Spinuzzi, 2012), social sensemaking (Puussaar, Clear and Wright, 2017), and decentralised file storage (Zichichi, Ferretti and D'Angelo, 2020)). There has been negligible research into the role data plays within human relationships.

This is the second research gap that this thesis aims to address—to look at personal data holistically in the context of your life. How does the holding of personal data by third parties affect people's ability to function in modern life? Do people have meaningful control over their personal data in this multi-party landscape? What practical problems do data-holding organisations current practices cause for people? What role should data take in our complex digital lives?

2.3 An Alternative to Data-centrism: Human-centricity

2.3.1 Human-Computer Interaction (HCI) Foundations

Up until the 1980s, the only reasons to consider the relationship between a human and the computer they were using were ergonomics, comfort and efficiency. People were shielded from the complexities of the machines they were using—the machine did the work (usually a single task) and the human was just the *operator*. In the 1990s, the 'first wave' of what is now known as *Human-Computer Interaction (HCI)* recognised humans as *actors*, often operating in groups, who had tasks to perform either using or assisted by technology (Bannon, 1995). People were now *users* of technology. Design thinking shifted from machine-centric to *user-centric design (UCD)*, motivated by the goal of helping the user to perform their tasks better. In the personal computer revolution of the 1990s, people began to work in complex and varied multi-user situations, and observation and understanding of a user's working environment and *context* provided empathy that enabled better design.

In the 2000s, as smartphones, broadband and *Web 2.0* brought computing into every aspect of our lives, HCI's third wave looked beyond the workplace to consider users as unique humans with emotions and culture; design became about *experiences* (Bødker, 2006) which could span work, mobile and home domains. Computers were no longer just for work; individuals would be free to use computers as they saw fit. This created a 'chaos of multiplicity for HCI in terms of use technologies, use situations, methods and concepts' (Bødker, 2015). Designers would now need to 'embrace people's whole lives' (Bødker, 2006). The blueprint for how this could be achieved can be found in Mark Weiser's seminal 1991 Scientific American article *The Computer for the 21st Century*, in which he envisioned a world where data could be accessed across many different devices, such that interfaces and interactions could be designed around the user's data needs in specific contexts. He recognised the need to put humans, not machines, at the centre of data interaction, and that in order to achieve *calm computing*, technology would need to *disappear into the background* of our lives (Weiser, 1991; Weiser and Brown, 1996).

2.3.2 Data Transcendence, Context and Human-Data Interaction (HDI)

Weiser's vision was significant because it recognised the need for data to transcend the confines of a single machine; to satisfy human needs in different contexts, data needs to be *pervasive* (Saha and Mukherjee, 2003; Krishnan, 2010). From a technical perspective, Weiser's vision has largely been realised, with today's smartphones, tablets and digital whiteboards / smart TVs corresponding directly to his imagined 'tabs', 'pads' and 'boards' respectively. *Ubiquitous computing* now allows environments, vehicles and wearable computing to collect data via sensors—the *Internet of Things (IoT)*, which enables *context-aware computing* (Abowd *et al.*, 1999; Eliasson, Cerratto Pargman and Ramberg, 2009). But what of the interaction perspective? As an answer to this question, the concept of *Human-Data Interaction (HDI)* emerged. This sub-discipline of HCI outlines the vision that the human needs to have a direct, explicit relationship with their own data (Mortier *et al.*, 2013, 2014), and that personal data should be considered an entity in its own right; people do not just need to interact with systems, but with the data itself. This can be seen as an echo of previous calls throughout the decades for a new relationship with our stored knowledge (Bush, 1945; Lansdale, 1988; Rogers, 2006; Hendler and Berners-Lee, 2010; W. Jones, 2011a).

Mortier *et al.* laid out three tenets of HDI. As individuals interacting with our personal data we need:

- *agency* (influence over how our data is used within the system);
- the data needs to be *legible* (i.e. understandable) to us; and
- we need *negotiability* (the ability to flexibly adapt and make use of the data).

HDI has remained a small but important research niche within HCI, and many researchers continue to explore this field today ('Human Data Interaction Project at the Data to AI Lab, MIT', 2015; BBC R&D, 2017; 'HDI Network Plus, University of Glasgow', 2018; 'HDI Lab, Heerlen', 2020), as does this thesis. In order to understand what HDI might mean in practice we can look to Gregory Abowd's 2012 paper which aims to update Weiser's vision. In it, Abowd emphasises the importance of programming for *environments*, building a complete experience for the individual that considers not just the 2D screen they are using, but the entire surroundings and context of their situation. He imagines a hybrid, conjoined experience between people, devices, sensors and the cloud where data storage and processing need not be constrained to the input and output devices we use (Abowd, 2012) and crucially, that the individual within this *'everyday computing'* experience is harnessing technology for their own ends, not just being aided to complete a predetermined task (Abowd and Mynatt, 2000). In essence they are able to program their own environment.

2.3.3 Human-centred Design: A Sociotechnical Challenge

Abowd's vision is a helpful reference point to remind us how far from true human-data interaction we are today. As described above, data is trapped, and very few computing interactions today are designed as a situated experience. Some TV streaming services show a good example of an interaction whose design has taken into account context: instead of typing in long email addresses and passwords, difficult on a TV remote, you can visit a short link from a smartphone or PC where you are already authenticated. But even though there are pockets of research around contextual experiences (for example the work around second screening (T. Jones, 2011; Zúñiga, Garcia-Perdomo and McGregor, 2015)), in general most design work today still focuses on a single interaction surface. In order for technology to disappear into the background so that we might live in a calm, engaged manner, as outlined by Weiser and expanded upon by Yvonne Rogers (Rogers, 2006), a more humane

interface is needed (Raskin, 2000), one which designs for the whole person. Judging the success of a user interaction can no longer be done by assessing task-completion efficiency (Abowd and Mynatt, 2000) but should consider the holistic needs of the individual at that moment in time.

In the 2010s, there was a growing recognition that the world had lurched severely away from such goals. The design of information-consumption interfaces was having a detrimental effect upon people, not just in terms of the psychological impacts of information overload as detailed above in section 2.2.4, but also in terms of the impact on users' attention. This drive to capture users' attention became known as *the attention economy* (Simon, 1971; Croll, 2009; Cogran and Kinsley, 2012; Brynjolfsson and Oh, 2012). Social media technologies like infinite scrolling and smartphone notifications had created 'a culture of perpetual distraction' (Timely, 2020) which 'hijacks people's minds' (Harris, 2016). Zeynep Tufekci posited in a TED talk that:

> *"We are creating a dystopia just to make people click on ads."*—Tufekci (2017)

In 2013, Tristan Harris released a presentation calling on the tech industry to respect users' attention and minimize distraction (Harris, 2013a), which lead to the creation of the *Center for Humane Technology* (Harris, 2013b), a central group in this new movement to design for positive human values and to practice *value-sensitive design* (Friedman and Hendry, 2019). This focus beyond just supporting data interaction to understanding and enhancing the individual's lived experience can be seen as a central guiding tenet of *Human-centred design*.

The design of human-centred personal data interaction is not purely a matter of designing better user interfaces, nor even of designing for the user's physical environment. It is a design challenge that exists at the sociotechnical (Bunge, 1999; Murton, 2011) level. We must consider the social relationships of the individual [2.2.5] as well as the power imbalance that exists between data holders and data subjects [2.1.2]. Andy Crabtree recognised the sociotechnical nature of the HDI challenge in his 2016 paper with Mortier on *The Shifting Locus of Agency and Control* and highlighted particular aspects of this multi-party challenge around personal data, specifically being able to ensure the *privacy* of your data as well as the *accountability* individuals require over data-processing algorithms and

data-handling organisations (Crabtree and Mortier, 2016). These goals are now actively pursued through research into *privacy by design* (Cavoukian, 2010) and *critical algorithm studies* (Gillespie and Seaver, 2016) respectively. In his subsequent work with Peter Tolmie, Crabtree focused on the particular HDI challenges around data-sharing, which must also be designed for (echoing Lindley's work on file biographies mentioned earlier) (Crabtree and Tolmie, 2018). These areas, pursuing a human-centric agenda within a sociotechnical context, are actively researched today, as seen in projects such as Nesta's *DECODE* (Symons *et al.*, 2017) (which focuses on individual empowerment), and UKRI's *not-equal.tech* (Crivellaro *et al.*, 2019) (which focuses on *data justice* (Taylor, 2017)).

2.3.4 The Pursuit of Human-centred Personal Data Ecosystems (MyData and beyond)

During the 2010s, while many were focused on the utility of PIM systems (as described in 2.2.2, hereafter referred to as 'traditional PIM'), some researchers, thought leaders and strategists were developing ideas that can be seen as the first sociotechnical designs for personal data interaction. One of the earliest was Doc Searls, who launched a project called *ProjectVRM* with colleagues at Harvard University around 2008. He envisioned a model he called *Vendor Relationship Management (VRM)*. VRM can be seen as the inverse of *Customer Relationship Management (CRM)*. In CRM, organisations use data to profile and learn more about their customers and get their attention (Searls, 2008). In essence, the VRM vision (expanded in (Searls, 2012)) was to combat the attention economy by turning the world of commerce inside-out. Individuals would publish tightly-controlled personal data about themselves and their needs, and retailers could respond to these individuals with product offers, from which (s)he would then select.

Drawing together the work of Searls and various other innovators in the VRM space, David Siegel catalogued an emerging vision of a personal data interface that could realise VRM thinking, which he called a *Personal Data Locker* (Siegel, 2009, 2010) (Equivalent terms *Personal Data Store (PDS)*, *Personal Data Vault* and *Personal Data Services* are also common). He described 'pull-centric computing', where information is 'pulled' at your request rather than being pushed upon you. The WEF's *Rethinking Personal Data* project describes the potential for individuals to have their own *personal data ecosystem (PDE)* of 'commercial entities, acting as trusted intermediaries, exchanging assets on behalf of the

individual, following a clear set of principles and legally binding contracts'. They describe the PDS as being the technical means to place the individual at the centre of that ecosystem: the PDS provider would be *'an intermediary collecting user data and giving third parties access to this data in line with individual users' specifications'* (Hoffman, 2010). A 2010 report by non-profit *Mydex* helps to contextualise the PDS, explaining that the PDS is a service that helps individuals become 'information managers at the epicentre of a new ecosystem of PIM services'. PDSes, it asserts, will not just give access to data but 'transform relationships between individuals and organisations' (Mydex CIC, 2010). In my view, this is what substantially differentiates the PDE model from traditional PIM systems–it is a response to the sociotechnical need outlined in the previous section. A 2012 report from Ontario's Information Privacy Commissioner notes that the PDE collides with traditional concepts of ownership when it comes to data, that the PDE needs to 'provide a collection of tools and initiatives aimed at facilitating individual control over personal information' wherever it is located; this is another way in which PIM within PDE can be differentiated from traditional PIM (Cavoukian, 2012).

It was against this landscape that *Personal Information Management Services (PIMS[2])* became a business area in its own right, the basis for what is now know as the *Personal Data Economy (PDE)*. PIMS is attempting to create a market for 'tools that help individuals gather, manage and use personal information to make better decisions and manage their lives better'. It is potentially a huge business opportunity (Ctrl-Shift, 2014). In 2016, a global network and non-profit initiative called *MyData* was founded, bringing together researchers, companies and public sector agencies in the PDE space:

[2] The usage of the abbreviation PIMS here is not to be confused with its earlier use to refer to *Personal Information Management Systems* in traditional PIM terminology. Adding further confusion, a 2020 report by the European Data Protection Supervisor of the European Union refers to this newer area of personal data ecosystems or personal information management services as *Personal Information Management Systems (PIMS)*, perhaps as an attempt to reclaim the earlier term (Attoresi and Moraes, 2020). In this thesis, the terminology of Personal Data Ecosystems and of a Personal Data Economy, both referred to by the acronym PDE, is preferred.

> *"The human-centric paradigm is aimed at a fair, sustainable and prosperous digital society, where the sharing of personal data is based on trust, and relationships between individuals and organisations are balanced."*—MyData.org (2018)

An important aspect of MyData is its aim to combine companies' needs for data with individuals' digital human rights. Through analysis of principles of PIMS, VRM and other related spaces ('MyData Comparison of Principles document', 2017), the *MyData declaration* was produced, outlining a detailed vision for the PDE space.

> *"Our common goal is to empower individuals with their personal data, thus helping them and their communities develop knowledge, make informed decisions, and interact more consciously and efficiently with each other as well as with organisations."*—MyData (2017)

MyData now has over 700 parties involved worldwide and provides a focal point to the PDE community. The MyData declaration identifies data controllers' transparency with data and data-handling practices as an essential means for individuals to gain agency and accountability, and puts forward the idea that the individual should be the point of integration of their own personal data ecosystem; in simple terms, 'everything goes through me'. This is the embodiment of the human-centric ideal of individual empowerment but will also be a good way for data controllers to ensure awareness, accuracy and consent.

They also introduce the idea of a *personal data operator* (also known as a *data trust*) which is a key part of the personal data ecosystem–a trusted third party which stores or transfers data on behalf of the data subject, but does not use it themselves. Data trusts introduce the idea of a fiduciary responsibility to keep data safe, and offer a means to give people more control over how their data is used (Delacroix and Montgomery, 2020). An example operator is *digi.me*, which has developed a PDS with a *private sharing* model that allows users to allow subsets of their data to be used by external organisations or apps with strictly controlled parameters (Firth, 2019). The MyData/PDE space is very active currently, with many emerging businesses and start-ups having appeared in the last two to three years. *Citizen.me* (CitizenMe, 2021) positions itself similarly to *digi.me*. Other operators such as *UBDI* ('Whose data is it anyway?', 2019) and *datacy* ('datacy - About Us', no date) are positioned under a different business model which aims to help individuals take control of their personal data for profit. *Open Humans* has a PDS optimised to allow people to share their data for the benefit of research (Price Ball, no date). In 2021, *Ethi*

developed a PDS platform focused on providing individuals with deep insights from their data, and tools to more easily delete personal data from data-holding organisations (Jelly, 2021).

Many of these operators are building for a human-centric world that does not yet fully exist, and Swiss firm *Hestia.ai* [See Section IV Introduction] also operates in this space, with a different focus: helping people to understand the existing data ecosystems that make use of their data (Dehaye, 2018), building out towards MyData ideals with a more pragmatic recognition of the challenges of the status quo.

2.4 Summation: Defining the Research Question

Section 2.3 has shown how the emergent human-centric personal data ecosystem has developed from its roots in HCI, ubicomp and HDI. It is against this backdrop of ongoing research and innovation that this thesis is situated. The call for designs and sociotechnical systems that empower individuals with their personal data arise from the power imbalance (Hoffman, 2014a) that has emerged as a result of the aforementioned datafication of modern life [2.1.2; 2.2.4]. As Chapter 3 will explain, it is the lived individual experience of this power imbalance that must be understood if the situation is to be improved for individuals.

Considering HCI's evolving view of designing for people first as operators and actors, then as users who have experiences [2.3.1], this power shift can be seen as a subsequent, unhelpful shift in design thinking. In the third wave of HCI (Bødker, 2015), user interface design's main consideration was 'what does *the user* want to do?'. However, this is no longer the primary driver. Catalysed by the transition from self-install software products to massive-scale cloud-based Internet services, there has been a perceptible shift over the last decade away from the tenet that the user's needs should come first. Application designs now more reflect the question (considered from the provider's perspective) 'What do *we* want the user to do (or *not* do)?'. Users (people) and their individual needs have been left behind.

While the *MyData* community [2.3.4] have clearly outlined the goals to address this problem (MyData, 2017), much of the focus in this space at present is on technology questions of how to build better PDSs and better PIM interfaces, or on identifying an

effective business model that will facilitate the transition to a PDE. In a capitalist and product-driven market, these are necessary questions, but limit exploration of the wider challenge of reconfiguring society towards human-centric instead of data-centric thinking.

My research, situated at the bleeding edge (Collins Dictionary, no date) of the emerging human-centric PDE space and being non-commercial, is able to take a more purist position. From this stance, maintaining a focus not just upon the technological perspective of human-data interaction but also upon the social implications of data use by service providers, this thesis explores what humans want from their personal data, building on the research gap from 2.1.5. In parallel, it also discovers what people want from those service providers who collect, store and use it in ways beyond individual reach, building on the research gap from 2.2.5.

The primary objective of this PhD research, therefore, is to use participatory research methodologies (detailed in Chapter 3), to advance the existing body of theoretical knowledge around effective data access, information management and human-centric data interaction, by providing an empirically-sound answer to the question:

“What relationship do people want with their personal data?”

Chapter 3. Methodology

"Research is defining the invisible."

—Steven Magee (author and researcher)

The previous chapter described the two research areas this thesis seeks to explore: how people think about data and what they want from it; and the role data takes and should take in people's relationships with organisations. This chapter will explain this PhD's approach to conducting research in this area, detailing the types of methods used, and explaining how the different research activities carried out contribute to the research objectives.

3.1 Forming a Research Paradigm: Ontology & Epistemology

To develop a research paradigm, one must reflect on two things (Guba, 1990):

- *ontology*: your outlook on the nature of reality, and
- *epistemology*: your beliefs on how knowledge of that reality is formed.

It will already be evident to the reader that individual human experience is at the forefront of my thinking. I believe that everyone experiences their own reality, informed by their own concepts and mental models of the world. This ontological stance, known as *constructivism* (Guba, 1990) argues new knowledge is formed by developing one's own mental models in order to explain new experiences, as distinct from the positivist view that there is a single universal reality one can uncover. In today's rapidly evolving technological landscape, people have no choice but to develop new mental models to make sense of new concepts such as *posts*, *feeds*, *link sharing*, *syncing* and *blocking*, in order that we might find value in them.

This idea that reality is constantly renegotiated by the individual is known as *pragmatism* (Campbell, 2011). People's developing conceptions of reality are not purely an intellectual endeavour though; As *Deweyan pragmatism* states, our knowledge and thinking are tested by actions, not just reason, and that this is how we learn. Communication and interaction with others are key parts of that learning. Dewey recognises that every individual is not

solitary, he exists within a society as 'a social being, a citizen, growing and thinking in a vast complex of interactions and relationships' (Dewey and Archambault, 1964). People create systems and meanings through those interpersonal interactions—which they can then use to understand everyday life. This is particularly important in the social world, as unlike the physical, natural world, many concepts are abstract and subject to individual interpretation. This drives my research motivation to understanding how people make sense of their world, and how that changes as a result of their lived experience.

Constructivists tend to believe that people are motivated by a desire to solve problems combined with a confidence to learn (Prawat and Floden, 1994). In this regard, I look beyond traditional constructivism, as I think *individualism* (Lukes, 2020) offers a better explanation. Much though we might not like it to be the case, people are self-interested, pursuing their own individual happiness and wellbeing, and it is this self-interest that drives the pursuit of deeper knowledge and understanding about the world. In essence, everyone wants to improve their own life, and they need to acquire knowledge about how the world works so that they might be able to change it. This path of acquiring and sharing individual understandings about the world, which can then be used to effect change upon the world, is a key driver behind my research. Taken further, this can be seen as *civil libertarianism* (Gulite, 2014), which argues for the supremacy of individual rights and personal freedoms over imposition by authority. The human-centric movement [2.3.4] clearly subscribes to this philosophy, as do I, and this explains why my research moves from simply understanding the world (in Section II), to functioning as an activist trying to change it (in Section IV).

My established ontological stance, then, is that individuals construct concepts, and continually update them through sensory experience, action, and social interaction. They do this in order to maintain a pragmatic knowledge that they can practically exploit to effect change in society and in the world, in order to pursue their own happiness and self-interest.

Looking to epistemology: how can knowledge be acquired? Constructivism suggests that this is best done not through direct observation of the world and explicit testing of hypothesis, but though interacting and communicating with individuals so that we can interpret how they view reality; this is known as an *interpretivist* epistemology (Dudovskiy, 2012). This motivates my choices to favour *qualitative* (understanding perspectives and

collecting spoken/narrative data) rather than *quantitative* (measuring behaviours and collecting numerical data) investigation. By understanding people's views and mental models around data and digital living, I can look for commonalities and develop *theories*—powerful explanations that can be understood and benefitted from by ordinary people—to fill the knowledge gaps in existing research that I have identified [2.4]. Given my pragmatic focus on interpreting people's constructed social realities in terms of practical usefulness (in individualist terms: individual benefit) to them, I will not be deeply analysing their words through language analysis techniques like discourse analysis. I will instead focus on the social, interpersonal level–understanding how people navigate the world of data and data-based relationships. I want to explore how they change their understandings as they pursue their goals in practice, and how they are affected by the systems, relationships and society they exist within. This pragmatic approach and search for commonalities motivates a *mixed methods approach* (UK Health Security Agency, 2020), combining qualitative or quantitative methods as appropriate to the particular sub-question being explored.

Thus, my epistemological stance is an interpretive one, that also recognises that in order to identify commonalities and *shared meanings* (Dudovskiy, 2012), the researcher must employ a mixture of qualitative and quantitative techniques.

3.2 Research Approach

3.2.1 Participatory Co-Design & Experience-centred Design

Moving from general research philosophy to specifics of this PhD, I see the purpose of this research as to formulate theories that can facilitate change–to map out a research and development agenda that might help the world to move from a data-centric [2.1] to a human-centric [2.3] operating paradigm. Learning about people's understandings of their reality informs and enables an inductive research approach where patterns common to multiple people can be identified to form general understandings of what people *want*[9] from data, and from those who hold it. Based on the premise of individual betterment laid out above, these desires serve as justification for the thesis to take a moral stance on personal data: That access to data held about you by others is a fundamental right, and that theory and action that can empower individuals to be able to understand and use their data

for their own purposes is a fundamental good that will benefit individuals and society at large.

As a student of *Digital Civics* (Vlachokyriakos *et al.*, 2016), I believe that research can surface the ways in which current service provisions fail to meet people's needs, and how the world might better empower citizens if it were configured differently with services closer to what they desire. The role of the researcher is to understand the world and to figure out how to change it. It is an accepted view that research cannot be value-free, but in fact I advocate an even more positional stance: the researcher can be an activist, seeking to correct an imbalance in the world through their research. As such, the design elements of this research can be considered as political. In this thesis, especially in parts of Section IV, I embrace *adversarial design* (DiSalvo, 2012). I view this as necessary to counterbalance the strong forces outlined in Chapter 2 and Chapter 8 that are acting against individual interests, and in order to pursue the moral imperative laid out above. By creating space to reveal and confront power relations and influence, we can identify new trajectories for action (DiSalvo, 2010). Therefore, the purpose of the participatory research carried out through the Case Studies is to inform myself as a designer, for both conformist and adversarial or activist design approaches. Acquired insights from the experiences of research participants can help me to develop my own understanding, models, and theories for change.

When designing for people and trying to incorporate their views, there are traditionally two schools of thought: *user-centred design (UCD)* and *participatory co-design (PD)*. In UCD, design is carried out by experts, who have undertaken user research to build up understandings of user needs (Norman and Draper, 1986). This approach places a high value on expertise, but it carries the risk that certain user needs may be overlooked, especially those that are less common (and therefore less likely be present in a designer's concept of 'the average user'). UCD is the most common approach used by technology companies today, not least because commercial motives must be incorporated into designs, and therefore design can never be fully democratised. UCD as implemented in modern software development practice does however recognise the importance of representing the user perspective in the design process, and uses processes such as *focus groups*, *user experience testing*, and *user persona development* to include their perspectives. However, in

a commercial context such perspectives may ultimately be ignored or diluted in favour of expert designs or organisational motives, as observed in 2.4.

Recognition of this inherent problem–that users carry less influence than designers and that this imbalance must be tackled head on–lead to the ideas of co-creation (also known as *co-design*) and PD. PD is based upon the idea that those who will use or be affected by technology have a legitimate reason to be involved in its design (Kensing and Blomberg, 1998). PD is seen as an attempt to design in a more democratic fashion. PD proponents argue that it is not sufficient to study users and go away and design in isolation. Instead, the users and technologists work together in design workshops, with users bringing their lived experiences and perspectives and technologists bringing their expertise on technical and market possibilities and constraints (Bjerknes *et al.*, 1987; Björgvinsson, Ehn and Hillgren, 2010; Smith, Bossen and Kanstrup, 2017). In such a collaboration, a collective, democratic design is created, taking into account all perspectives. In the 2000s, PD grew in popularity across public and private sector organisations, coincident with the growth of internet and social media into its *Web 2.0* phase (Hosch, 2017) which resurfaced the potential to reframe digital technology as something to be harnessed for users' own ends (Jenkins, 2006).

As design approaches, I see merit in both UCD and PD. The participant should play a role as an *informant*–one who can provide critical insights into their own perspective on a design space and help us understand how the world is to them–but also as a *designer*: one who can imagine how they would like the world to be. As we involve the participant, our role as researcher is to elicit the richest possible responses from the participant, by using questions to bring them to consider new possibilities and by giving them stimulating materials to trigger their thinking about what they might like to see changed. The researcher also often needs to *sensitise* the participant to a design space, so that they may properly engage with the questions being posed. Conversely, the researcher cannot arrive at a model or theory unless he has developed *empathy* for the participant's perspective. One of pragmatism's founding philosophers, Peirce, put forward the *pragmatic maxim*, which states that the meaning of anything we experience in the world is understood through the conception of its practical effect, and that theories that are more successful at controlling and predicting our world can be considered closer to the truth (Campbell, 2011). Applying this philosophy to the challenge of design, I find merit in the different, less political, take on

involving users as participants in design exhibited in McCarthy and Wright's *experience-centred design* framework, which identifies processes to improve user empathy including *interpreting*, *reflecting* and *appropriating* (McCarthy and Wright, 2004).

Through this research, I will at times be more participatory, to understand these aspects of user experience or to co-design solutions with participants. I will at other times act more like an expert designer, especially when moving from problem definition to solution-focused design throughout Sections III, IV and V. Taken to the extreme, the PD view is that designs made without the direct involvement of users are invalid, because they inherently no longer represent the desires of those people the designs claim to serve. I oppose this view, because I believe that new ideas will not always arise from participants themselves, especially for this research area–where a more expert-led experience-centred design approach is the most pragmatic way to proceed. By its nature, this research involves thinking about data, information, organisational relations and interaction at a level which the layperson is not accustomed or well-equipped to do. Therefore, while I strive to always include participant viewpoints, I give ultimate precedence in design to my own position of learning that I will acquire through the participatory research as well as embedded design and development work [See Section IV Introduction]. In incorporating both approaches, I will also be a participant in my own research, incorporating my own experiences of living in a data-centric world (and my attempts to challenge it) into my learnings.

It is important to be clear about what constitutes good research in this context; if the outcome of the research is to be my own interpretations and theories, how will we know these are sound? First, it is important to say that this is not about measuring the effectiveness of proposed changes upon the world. There will be no deployment of systems to test the ideas I put forward. This is not because such an activity would not be worthwhile—it would—but simply because by its nature, to develop, build and deploy new data interaction paradigms that function in real life with real personal data at the sociotechnical level would be too large an endeavour for a single researcher (or even a single academic research group) to undertake. Therefore, what I seek in this thesis is not to change the world, but to articulate with the greatest possible clarity discrete theories on how the world should, and could, be changed. Good evidence for the proposed changes will be achieved by ensuring that findings, themes and discussion contributions are backed up

by participant quotes or extant work and literature, and where an idea is suggested or agreed upon by many participants or where it resonates with the practical experiences of myself or of others, that can be seen as adding weight or validation to that idea. However, each person's experience is unique and needs to be put into context; not every insight will be shared by many participants and individual unique insights remain important.

3.2.2 A Two-Track Research Approach: Learning and Practice

The overall research design for this PhD was informed by two key motives:

i) the desire to ensure that both PD (participant-driven) and UCD (expert-driven) design approaches contribute to the learning and models produced by this research, and

ii) the recognition that real-world problems with a lack of agency and negotiability over personal data cannot be fully understood, nor any tangible impact achieved, without embracing the problem space at a sociotechnical level–going beyond understanding the problems, to attempting to actually apply that understanding in practice.

I judged, therefore, that participatory design work needed to be accompanied by practical work in relevant industrial and academic projects that were also attempting to improve people's agency and relationships with data, as well as individual experimentation to familiarise myself with data rights and explore technical possibilities. These grounded efforts would enable learning about the technologies, business processes and market forces affecting the problem space, and would allow theories and models emerging from the theoretical side of the research to be more thoroughly examined from a real-world perspective.

In order to explore the research questions from as many angles as possible, I wanted to design a research approach including three types of activity:

- *participatory co-design activities* - where I discuss and explore experiences, challenges and possible solutions with participants through conversations and design activities
- *self-experimentation activities* - where I carry out experiments, ranging from thought exercises to practical tests of what is possible, to develop ideas and models and explore the problem and solution spaces myself, and

- *embedded research activities* - where I participate as an involved team member, in external organisations' real-world projects that are trying to effect change in the personal data space, in order to learn about the challenges faced and the viability of different approaches, informed by the grounded experience of myself and others. This draws on the concept of embedded research as described in (Cheetham *et al.*, 2018).

In order to simplify the thesis, my self-experimentation activities (2) are not explicitly documented, only mentioned in passing, though they inform the thesis throughout. The Case Studies that make up Section II focus solely on participatory co-design (1), while the embedded research activities (3) are documented in Section IV and inform the latter half of the thesis.

It was an important part of my research design, that these two types of activities happen in parallel throughout the duration of the the PhD, so that theory could inform practice and vice-versa. The Case Studies' work with participants serves to inform an evolving understanding of human needs around data, while the embedded work in external projects involves a conscious shift of mindset from PD to UCD, in essence using the Case Studies' findings as requirements and acting as a testing ground for emerging theories, models and themes coming out of the participatory work.

This cyclical back-and-forth between the theoretical/participatory work and the grounded/embedded projects did not need to follow a formal methodology or discipline, but philosophically takes inspiration from existing concepts that hang on the idea of developing learnings being fed back into an iterative process, allow plans and designs and understandings to be updated. In an academic context, a useful frame for this is *action research* (Lewin, 1946) [3] . In production industries, especially software development, a

[3] It is important to note that while I, like many HCI researchers (Hayes, 2011), draw inspiration from an action research approach, I chose not to follow a formally-defined action research process such as *Participatory Action Research (PAR)*. PAR seeks to make sense of the world through collaborative efforts to transform the world rather than just observing and studying it (Lewin, 1951; Chevalier and Buckles, 2008), as I do, however that is where the similarity ends; I determined that it would be more valuable to explore a

common model for this is *agile development* (Fowler and Highsmith, 2001), where stakeholder feedback is collected at regular intervals. In entrepreneurial business contexts, terms like *fail fast* (Brown, 2015) and *pivot* (Ries, 2011) illustrate the idea that it is crucial to test ideas on real people then adapt quickly based on how that goes.

Whichever term is most relatable to the reader, the common thread is the same: that research or investigation should be done in cycles: you learn something, carry out some action in the world based on your learning, learn from what happened, and repeat. This has become an established approach in HCI research (Hayes, 2011). This PhD deliberately adopts a broad interpretation of these ideas. For me, getting value from action research does not mean that you must test every single idea with an audience for it to be considered valid, but rather that is important to consider that user engagement is not a one-off, but a repeated component that will continue to affect the research and learning path. Each new research activity will draw from past learnings and theories and acquired understandings so far, which are then developed through their exposure to real life in the process of participatory and embedded research activities.

---

variety of different contexts and activities over the PhD's course than to repeatedly loop back to the same participants for iterative feedback of a single set of designs.

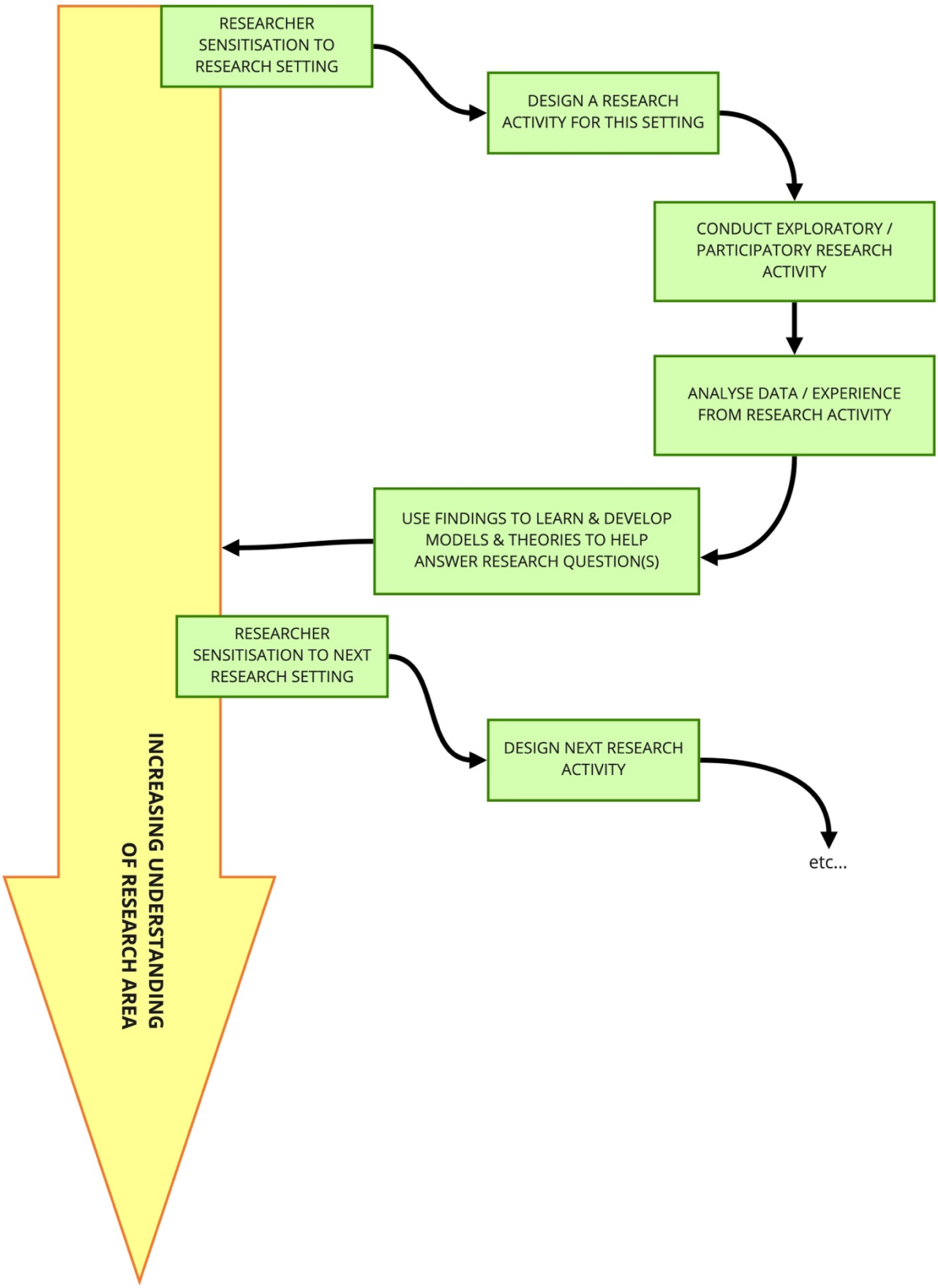


Figure 3.1: The Action Research Cycle, as Applied in this PhD

Figure 3.1 shows in the abstract, what each cycle of learning within this PhD looks like. In each area of life or context I identify as a setting for a research activity, I first carry out initial background reading, experimentation or exploration to familiarise myself with the area, then design a research activity that helps to explore the research questions in that area. After carrying out the planned activity I analyse any data from that activity (or just reflect upon my experience) and then use these findings to update my overall understanding of how we might answer the research questions. I then go on to repeat this cycle, with each subsequent activity, but beginning anew with more developed theories or understandings.

This cycle applies to both participatory activities (the design of which is fully in my control) and, to the greatest degree available to me, to the embedded project activities (whose design or implementation I was at times able to influence). Figure 3.2 shows how the approach I have described here can be considered as two parallel tracks of participatory (theoretical) and embedded (practical) work, where downward arrows represent increasing learning and understanding. The positioning of activities as boxes along these arrows throughout the period of this PhD, indicates that that activity contributes to that area of understanding. Each track's learning is contributed to by multiple activities, which continue to inform each other, both within and across the two tracks. This ongoing cycle of learning and influence between the two tracks is indicated by the angled purple arrows. This diagram represents what was at the outset, a planned approach, but has been updated retrospectively with the actual timeline of activities undertaken, giving a more concrete view of how the research activities progressed in practice. Specific activities are detailed further in Sections II and IV.

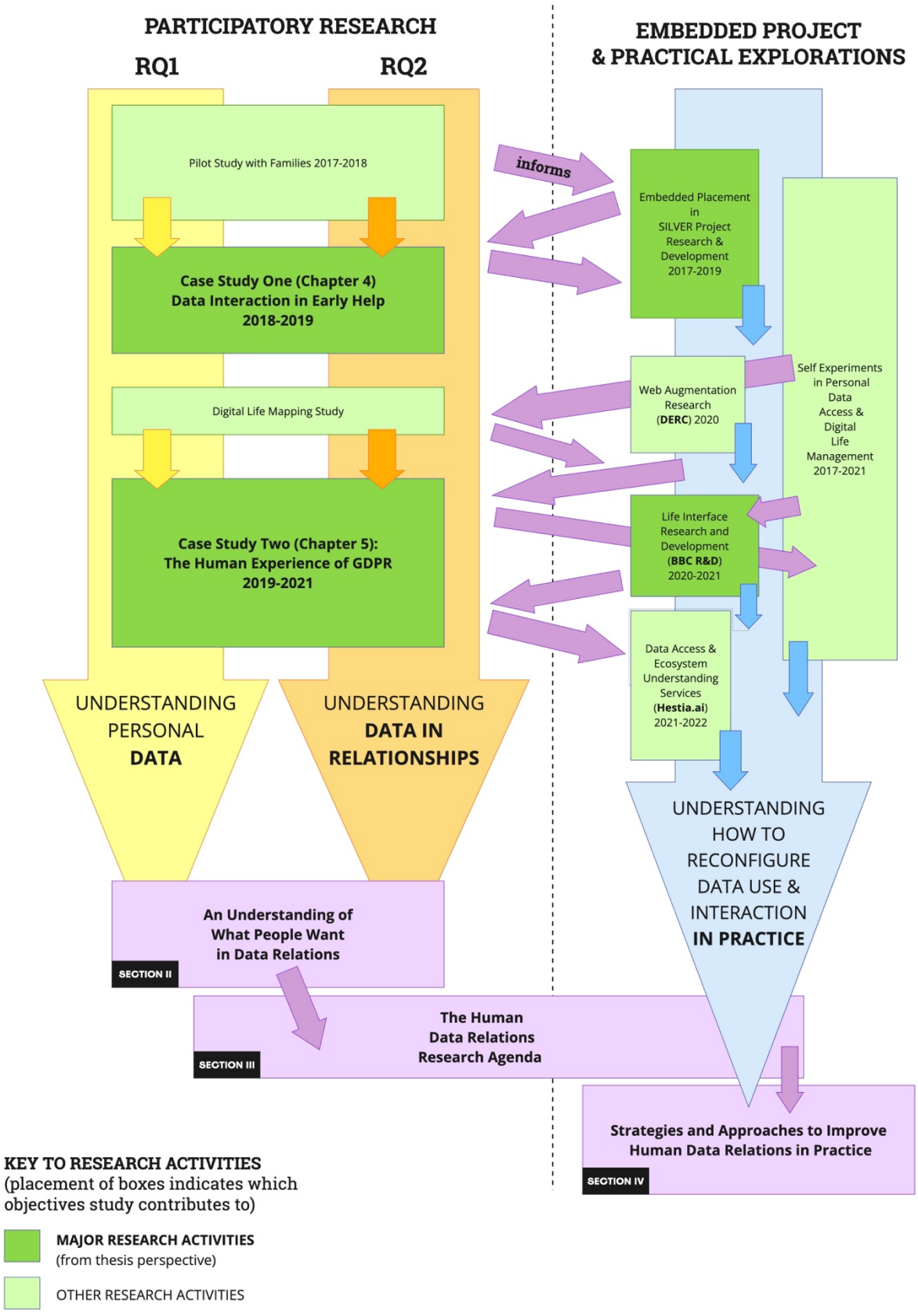


Figure 3.2: The Two-Track Research Approach of this PhD, and its Feedback Cycles

These cycles enable analysis and learning to happen throughout, resulting in a continually updating current understanding that can form a baseline for subsequent research and practice activities. In the next section, I will describe the two specific research objectives that are targeted through the Case Studies, and the further goals of the embedded practical work.

3.3 Research Objectives

The end of chapter 2 introduced the primary research question, which is:

"What relationship do people want with their personal data?"

As identified in 2.1.5 and 2.2.5, there are two distinct research gaps to be explored. The Case Studies, as well as preliminary explorations that feed into them, focus on exploring these two gaps, which are now expressed as two distinct research questions RQ1 [3.3.1] and RQ2 [3.3.2]:

3.3.1 Research Question 1 (RQ1)

What is the human experience of personal data, and what do people want from their data?

Personal data, and its collection and use by commercial and civic organisations, is an established and inevitable part of modern life [2.1], yet the concept of data is abstract and poorly understood. The first objective therefore is to establish a solid understanding of what mental models people have constructed about data. What makes data meaningful to people? Given HDI's belief that everyone needs a relationship with their data, what relationship do people currently have with their data? What *is* data to people? How does that relationship affect them, and what are their unmet desires for improving their relationship to their personal data? What aspects of data cause positive emotions? What problems do people experience with their data? What people want from their data?

A participatory approach will be used to explore these questions: gathering individual perspectives on data, and looking for patterns or trends in those perspectives. The first challenge here will be to find ways to sensitise participants to be able to conduct an informed and productive conversation about the topic of data, which to the layperson may

seem a dry, boring topic. This challenge will be addressed by leading participants into the subject of data using meaningful representations of data as stimulus for conversation, or starting with the individual's own life experience to discover the data in their life, which they are more likely to have opinions and emotions about, rather than talking about the subject in the abstract.

3.3.2 Research Question 2 (RQ2)

What role does data play in people's service relationships, and how could relationships involving data be improved?

Sections 2.2 and 2.3 established that, as of yet, designers of PIM and personal data interfaces have not yet risen to the sociotechnical challenge of today's personal data reality: it is scattered, inaccessible and largely *un-useable*[11]. There is no way for people to view their data holistically, as lamented by Ms. Heap in the quote atop Chapter 1. Nor are there tools to help people manage the many relationships that individuals have with companies, employers, councils, governments and other organisations that rely heavily upon the collection and processing of their personal data. Almost every civic or commercial service we use today handles data about us. We know that the world is data-centric, and that data controllers use data as an asset to inform their decision making, creating a serious imbalance of power (Hoffman, 2010, 2011, 2013, 2014a, 2014b). But what is like to conduct a relationship with an organisation that holds your data? What emotions do people experience? How does it affect their daily life, and what sort of problems do people face as a result of this data-centricity and digital life complexity? If your data is used in ways you do not understand or consent to, how does this affect your outlook on the world?

This is the second strand of research to be explored: to gain an understanding of the data world beyond the individual [2.2.5], so that we can design not just better individual relationships to one's data, but improve people's relationships with organisations that hold and use data. For the purposes of this study, the focus will only be upon service relationships (the primary place where data is held), not social or interpersonal relationships.

To tackle RQ2, participatory research approaches are again appropriate, as the questions relate to the individual mental constructs that people have about their wider digital lives

and relationships. But there is another aspect here, and that is that a relationship involves two parties. Consistent with Dewey's belief in the importance of interaction in creating meaning, the structuralist philosopher Michel Foucault said that 'meaning comes from discourse' (Adams, 2017). In other words, people do not construct their reality in isolation, but rather it is shaped by the social constructs and systems they operate within. Deweyan pragmatism also takes the view that research must seek solutions to real world problems that are generalisable to use in society at large (Dewey and Archambault, 1964; Friedman, 2006). This implies that any such solutions arising from my research must work for all parties. For both these reasons, the participatory research seeks to understand both perspectives: that of the data controller and that of the data subject, and where possible both parties together will be engaged in discourse so that the two parties' worldviews can be brought together to design solutions that could work in practice for all involved.

This second research objective will be tackled in tandem with the first, so that in each research setting we can examine the situation at two levels–to look introspectively at the individual's own relationship in service of RQ1, but also to take a step back and look at the wider social context the individual is operating within, so that we might be better placed to answer RQ2.

3.3.3 Practical Embedded Research: Going beyond the original RQs

While the Case Studies have the potential to uncover theoretical answers to RQ1 and RQ2, the parallel embedded project activities of this PhD presented the author with an opportunity to build a thesis offering more than just a theoretical expression of people's needs. By applying the learnings of the thesis in embedded project contexts, which are described in the introduction to Section IV, the primary research question of what relationship people want with their data can be explored in a real-world, practical sense. The embedded project work presents the opportunity to explore how individuals and groups might be able to take action to help change the world into the one they want, how these desired improved data relationships might be realised. In Section III [7.1], an extended RQ is expressed, which is the focus of the embedded research track of the PhD:

Having understood what relationship people want with their personal data, how might these better data relations be achieved?

The practical exploration of this extended research question is documented in Section IV. Methodologies for the embedded projects were not under the control of this researcher and therefore are not documented in this thesis, except where they form part of or contribute to the insights and learning acquired.

3.4 Methodologies Employed in Case Studies

This section will describe the methods used across the Case Studies within the participatory track, along with rationales as to why they were chosen. Loosely, the methods used can be grouped into five stages, though not every activity involved all stages:

- Sensitisation of Researcher and Participants
- Discussion and Exploration with Stimuli
- Participatory Co-Design of Possible Solutions
- Practical Data Experiments, Interface Design and Prototyping
- Data Analysis and Thematic Synthesis

These five planned stages will be explained using examples and illustrations taken from the actual execution of these methods within the Case Studies. Information about recruitment and ethics follows at the end of this section.

3.4.1 Sensitisation of Researcher and Participants

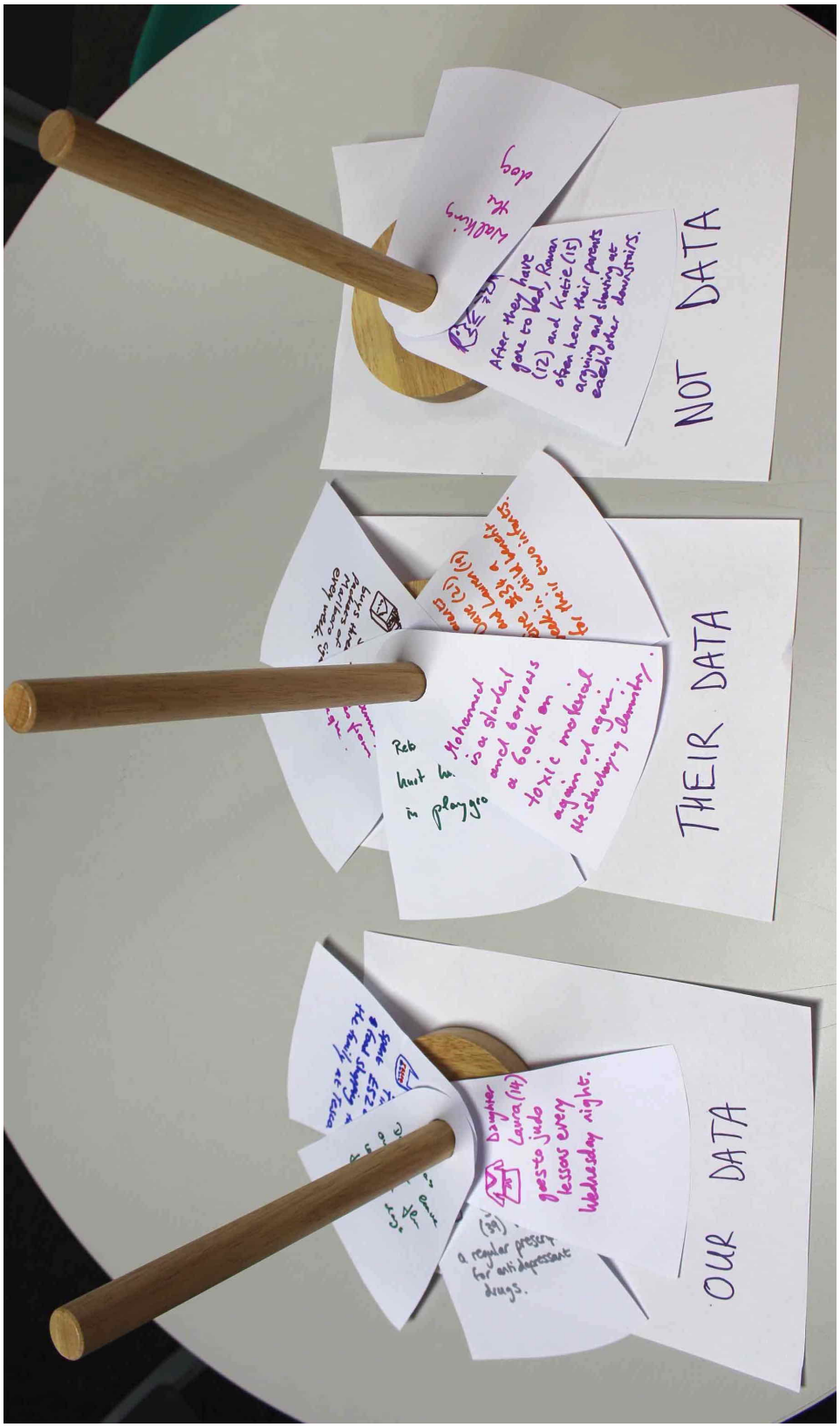


Figure 3.3: Family Facts - What is Data?

Figure 3.4: Walls of Data - Sensitising Participants to the World of Commercially-held Data and GDPR

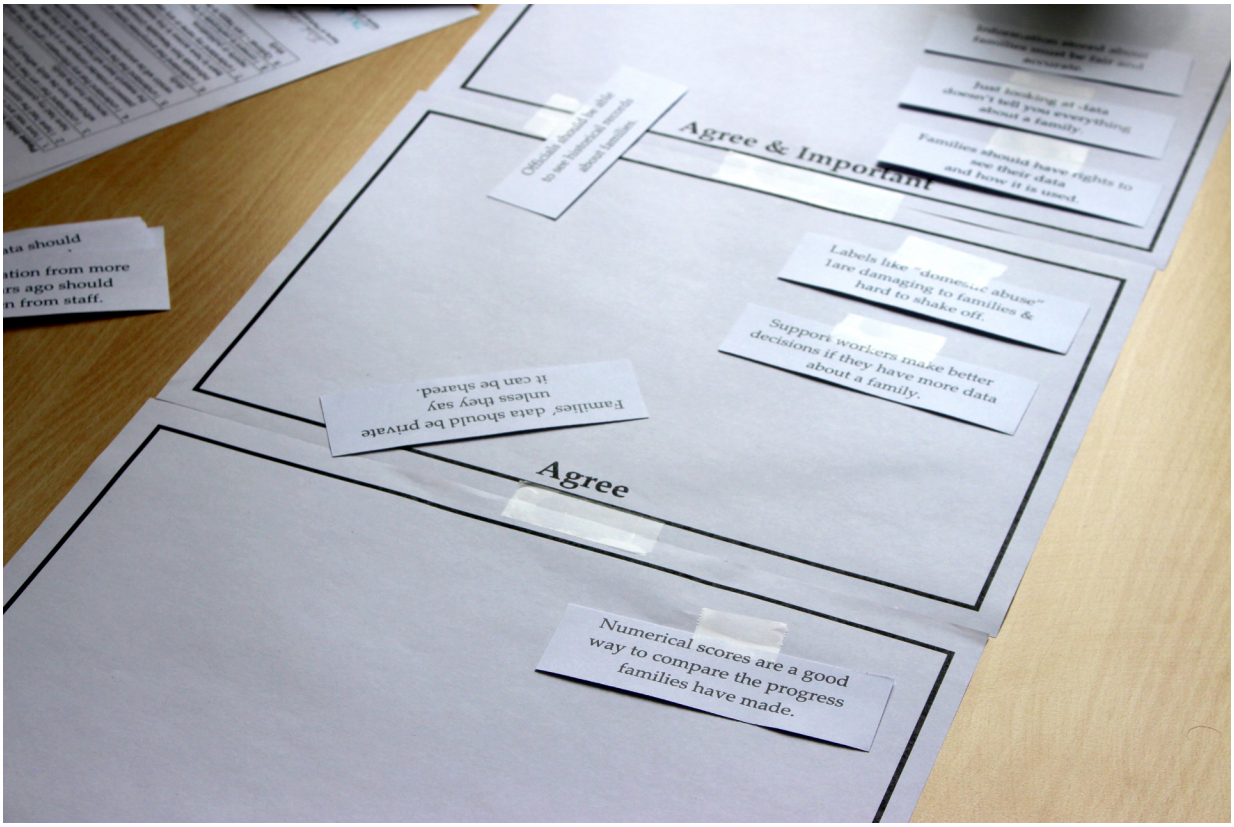


Figure 3.5: Sentence Ranking - Bringing Support Workers and Families to a Shared Problem Space

As described in 3.2.1, an important first step before each research activity is to *sensitise* oneself as researcher to the context, which means to become familiar with relevant issues, systems and practices and increase empathy for the participants. In the pilot study [4.2.2], this entailed a grey literature review to identify the different types of civic data that councils store, and conversations with colleagues and partner organisations within the SILVER project to deepen my understanding of Early Help. The pilot study [1.3.1] combined with this to serve as researcher sensitisation for Case Study One, as participating families had had some contact with the care system. This increased empathy for, and understanding of, supported family perspectives. Working with support workers through the SILVER project also provided empathy for the data needs of staff within the care service. In Case Study Two, self-experiments with GDPR as well as researching privacy policies and GDPR rights provided similar sensitisation before engaging participants.

Participants need to be sensitised too; when planning participatory research activities such as interviews or workshops, it is important to begin the session with an activity that will acclimatise participants both to the specific area of discussion, but also to the problem-

solving mindset required for constructive conversation. This goes beyond ice-breaking, and into thinking about what the participants bring and lack at the start of the engagement. For example, in the pilot study, it was considered that data could be a hard topic for families to engage with, so the *Family Facts* activity [Figure 3.3] was designed. This required family members to consider simple facts about their lives (some provided, some created by them) and discuss whether or not such a fact would be considered data, and additionally whether such a fact should be in the family's control or that of the authorities. This served a double purpose of teaching families that data is simply 'stored information about you', while also getting them used to thinking critically about data ownership. The technique is discussed further in (Bowyer *et al.*, 2018).

For Case Study Two, there was a goal to get participants to think about the data involved in their everyday lives, especially that stored by commercial service providers. So, using a series of posters in the researcher's lab common room were put up. These showed logos of companies that might store data, types of data that might be stored, information about GDPR rights, and possible uses that an individual might have for data they obtain from a GDPR request. Some of these posters are shown in Figure 3.4. These posters served both as a recruitment tool for the project and a discussion prompt for participant onboarding.

Sensitisation activities can also serve an additional purpose of bringing disparate participants to be 'on the same page'. This is known in participatory research as *co-experience* (Battarbee and Koskinen, 2005). An example of this is the *sentence ranking* exercise used at the start of all workshops in Case Study Two [Figure 3.5]. Here, a series of sentences were prepared containing opinions about civic data that had been observed from staff and families in earlier research. Participants were asked to rank these according to *agreement* and *importance*. This allowed validation of whether previous findings still held with these participants, but also acclimatised the participants to considering and discussing the civic data context and the problems experienced by families and staff. Since the sentences included both staff and family viewpoints, and the activity was carried out in all workshops regardless of whether staff, families or both were present, it served to establish a common set of 'requirements' that would be in participants' minds as they began the subsequent co-design activity within each workshop.

3.4.2 Discussion and Exploration with Stimuli



Figure 3.6: Family Civic Data Cards - Things to Think With

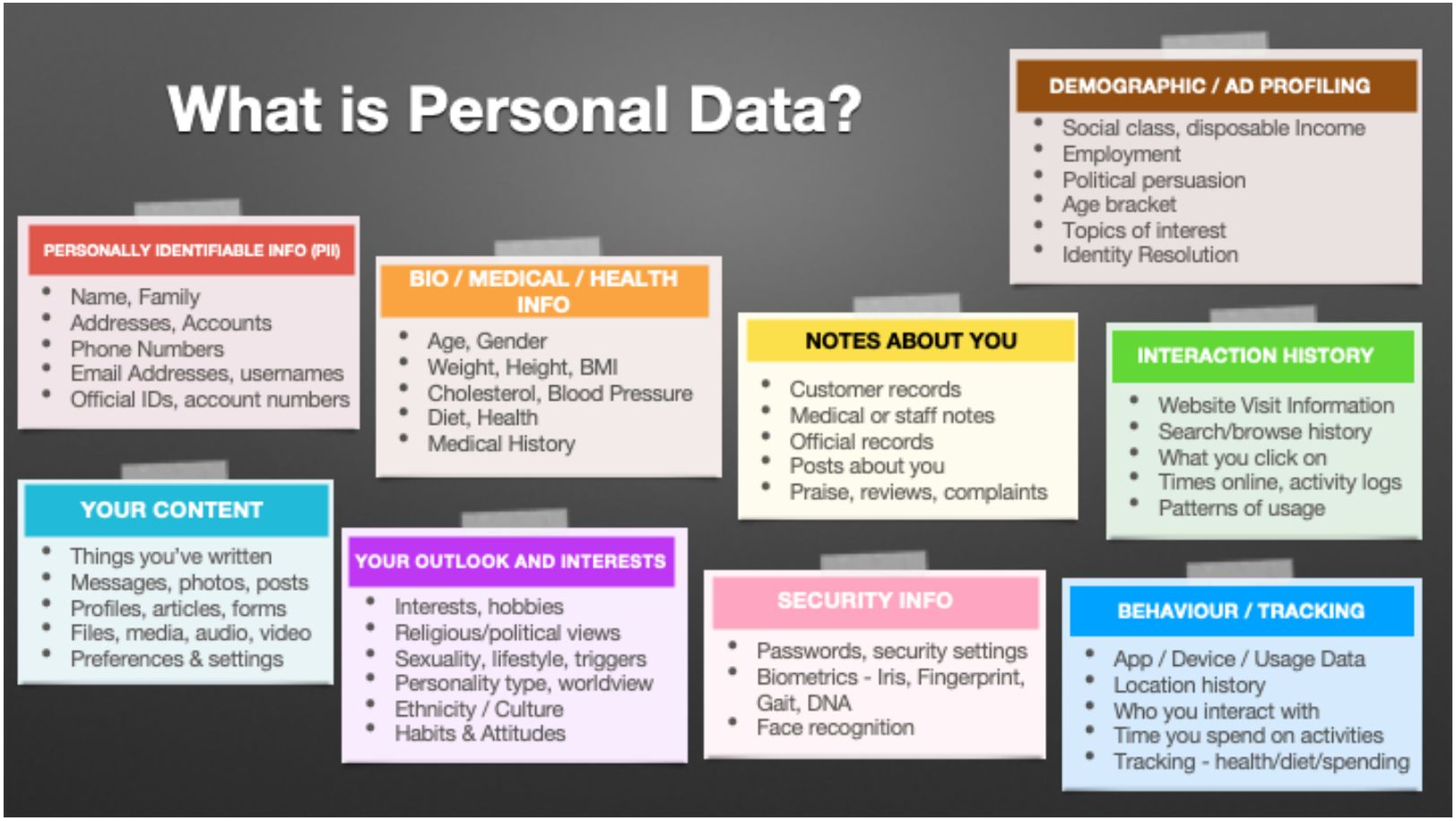


Figure 3.7: Personal Data Examples - Making Data Relatable



Figure 3.8: Home Interviewing: Card Sorting with a Family in Their Living Room

As discussed in 3.2.1, this research seeks to uncover individual perspectives and worldviews. The primary method used in both Case Studies was traditional qualitative interviewing - talking to people about the topic being explored. In Case Study Two, this was largely done on 1-on-1 basis (largely because of the sensitivity of dealing with one's own personal data, and because it allowed the researcher to get closer to the participant's individual experience). In Case Study One, group discussions and activities were mainly used, which brought the advantage of being able to *prime* a discussion between participants and then sit back into more of an observational role. This proved particularly insightful

when observing intergenerational conversations between family members in the pilot study [4.2.2]. In Case Study One, it allowed the negotiation of a 'middle ground' between support workers and supported families to be observed. In some cases, such as the home visits in the pilot study and local authority visits for SILVER, the researcher was able to conduct interviews-in-place (Pink *et al.*, 2013) in participants' own environments, which allowed for additional ethnographic observations to be made as 'life happens around' (Mannay and Morgan, 2015) the participants, as discussed in (Bowyer *et al.*, 2018).

Motivated by a desire to go beyond 'just talking' to achieve a deeper and more detail-oriented conversation, the decision was made to produce suitable stimuli to seed and progress each discussion. Given the abstract nature of the topic of data, it does not always carry a clear meaning in people's everyday lives, so the stimuli needed to make the topic vivid and real. Building on the aforementioned researcher sensitisation, a taxonomy and lexicon for Family Civic Data was created, and *Family Civic Data Cards* [Figure 3.6] were constructed for use in activities and discussions. These serve as *boundary objects* (Star, 1989, 2010; Bowker *et al.*, 2015)–representational artefacts that are understandable by people who come from different perspectives, providing a common vocabulary for discussion (as well as serving to enable co-experience, detailed above). Each card represents a different category of data, including a *summary* and *meaningful examples* to make them be easy to digest, yet still containing sufficient detail to stimulate thinking. The cards were designed to be bright, child-friendly and appealing to engage with. The tangibility of these artefacts was important. An approach of using tangible artefacts is valuable for situating legal and ethical discussions in context, by providing some concrete *things to think with* (Papert, 1980; Brandt and Messeter, 2004), facilitating reflection and discussion. Similar card-based approaches have been used by other researchers with success, such as Mora's Tiles IoT cards (Mora, Gianni and Divitini, 2017), Lane's Unbias Awareness Cards (Koene *et al.*, 2018) and Urquhart's Moral IT Deck (D. Urquhart and J. Craigon, 2021).

More generally, researchers have had success with the use of tangible objects to embody discussion concepts in order to stimulate and structure discussion, for example Coughlan's use of a dolls' house to explore attitudes to home energy use (Coughlan, Leder Mackley, *et al.*, 2013) or more recently Xie's Data City which used AR-enhanced cardboard models to

represent data-processing functions (Xie, Ho and Wang, 2021). These approaches have their roots in Dourish's concept of *embodied interaction* (Dourish, 2001). The cards were used throughout the Civic Data research in both sensitisation and *card sorting* (Spencer and Warfel, 2004) tasks, for example asking participants to position the cards on a pinboard according to perceptions about risk and ownership [Figure 3.8], or sorting them into trays according to relative personal importance. The cards proved very effective at enabling a personal and detail-oriented discussion. Participants voluntarily opened up about sensitive topics (e.g. domestic violence or criminal records) raised by the cards because of their detached-but-relatable nature (Bowyer *et al.*, 2018).

In Case Study Two, discussions around data did not use data cards, but the importance of meaningful examples to make the topic relatable persisted, in this case being demonstrated through posters [3.4.1] and in particular a categorisation of example data by category [Figure 3.7], similar to the examples on Case Study One's data cards. Within the BBC R&D placement, data cards pertaining to held personal data types and actions were subsequently developed for use in participatory research [Figure 9.15; Figure 9.16].

The sketching dialogue technique [Hwang (2021); Figure 5.2] used in the digital life context can also been as another stimulus technique; by putting both participant and researcher's focus upon the page, rather than on each other, it can feel less invasive, more collaborative and makes it easier to focus on details. The ideal stimulus for discussion about data is to view the actual data itself. Due to the sensitivity of personal data, this is more easily done 1-on-1 than in a group. Exploring data together with participants to elicit opinions and insights is a well-established technique (Coughlan, Brown, *et al.*, 2013; Chung *et al.*, 2016; Puussaar, Clear and Wright, 2017). This is the technique used within Case Study Two, asking participants about the data they retrieved from GDPR requests, using a spreadsheet-based approach. This allowed the Zoom-based interviews to retain a 'gathered around the table looking at things together' ambiance despite the remoteness necessitated by COVID-19 restrictions.

3.4.3 Participatory Co-Design of Possible Solutions



Figure 3.9: Ideation Decks - Combining Random Design Ingredients to Generate New Ideas

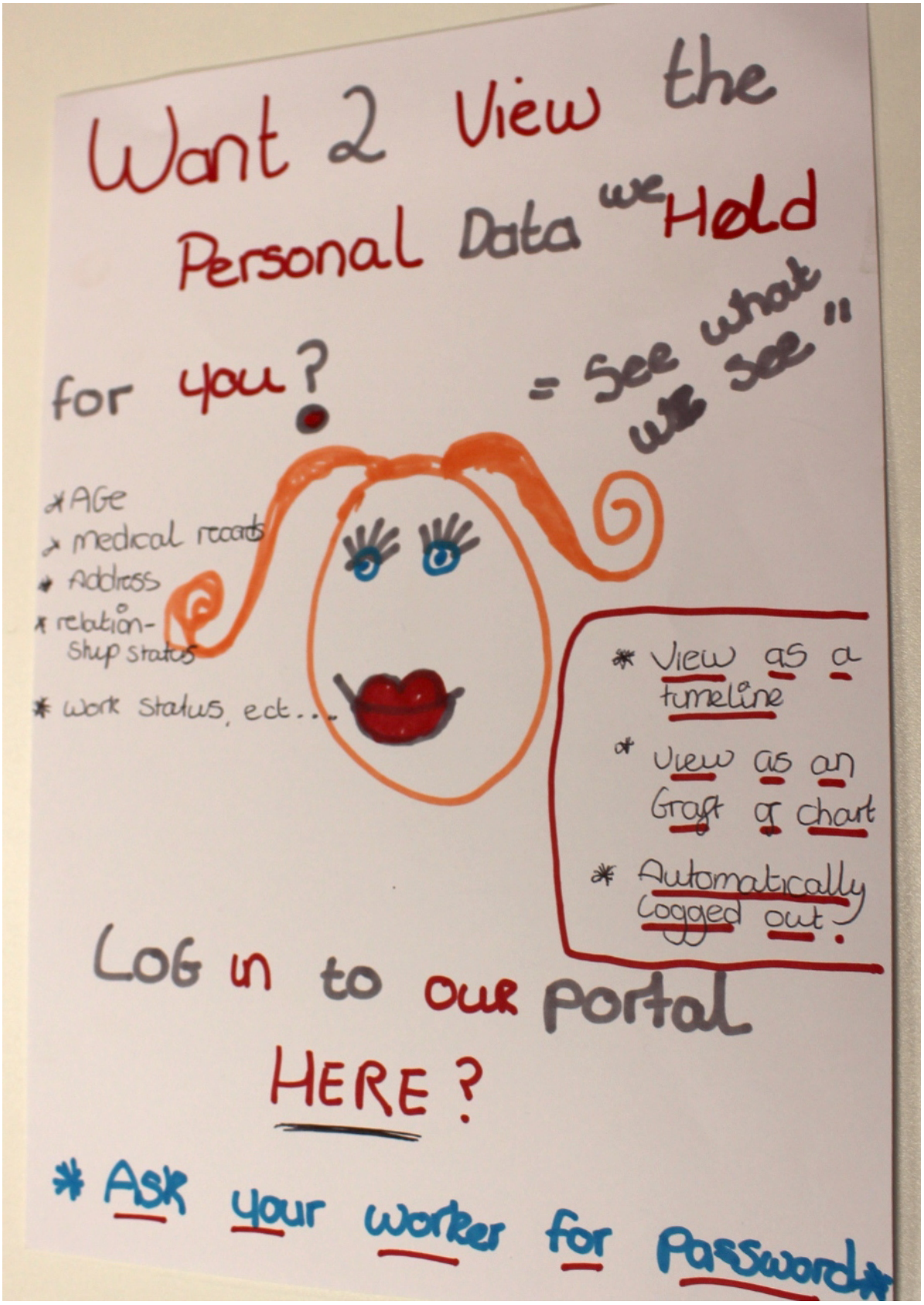


Figure 3.10: Group Poster Design - A Participant-designed Poster to Advertise Features of Imagined Data Interface Products

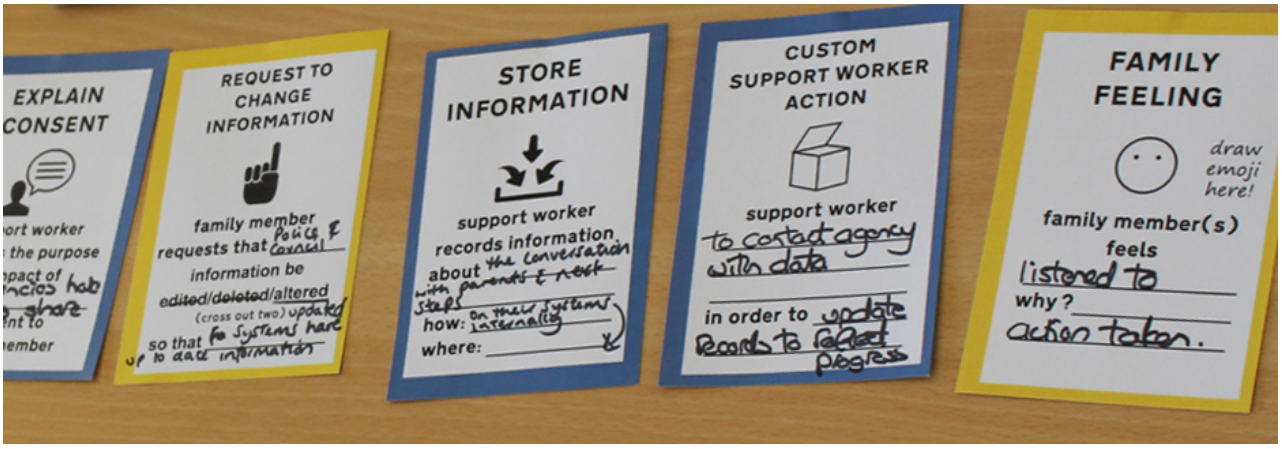


Figure 3.11: Storyboarding Action Cards - A Collaboratively-constructed Narrative Created through Discussion from a Palette of Possible Parent and Staff Actions

Participatory Design (PD) [3.2.1] becomes particularly important when exploring solutions and ideals to identified problems. It involves bringing participants into a *new mental space* where they can imagine the realm of the *possible*, rather than just their current lived experience. PD was an important part of Case Study One research with all participants. In the early stages of a PD activity, it is important that participants are able to generate a wide range of ideas, even fantastical ones, without constraints, self-censoring or judgements. This is known as the *discovery* phase in the UK Design Council's *double diamond* framework. (Design Council UK, 2004). Golembewski's *ideation decks* technique (Golembewski and Selby, 2010) was chosen for this purpose [Figure 3.9], as it allows participants to both select 'ingredients' of a design based on their own experience but also to combine them in a variety of different ways to generate novel ideas, guiding them into a previously unconsidered solution space.

After generating a wide range of ideas using the ideation decks, participants were then invited to pick just one or two ideas to develop into posters, each with three 'features' highlighted. An example is shown in [Figure 3.10]. This activity corresponds to the *define* phase of the double diamond, where participants narrow down options.

For the final workshop of Case Study One, where both parents and staff were brought together to explore possibilities of shared data interaction within the support relationship, a *storyboarding* activity was designed. Drawing from the world of film production, storyboarding is a well-established technique in participatory design (Spinuzzi, 2005; Moraveji *et al.*, 2007). Usually it involves the participants drawing out a series of sketches in

the form of a comic strip ‘telling the story’ of an interaction, encounter or activity. However, given the need [1.1.2; 3.2.1] to focus on interpersonal relations and process rather than the visual aspects of storytelling or interface design, and drawing upon earlier successes with data cards, a card-based approach to storyboarding was used. Participants selected actions from a palette of action cards representing different possible human or data interaction possibilities and annotated these with specific details. These cards are shown in Figure 3.11 and described in more detail in ARI4.3.

3.4.4 Data Analysis

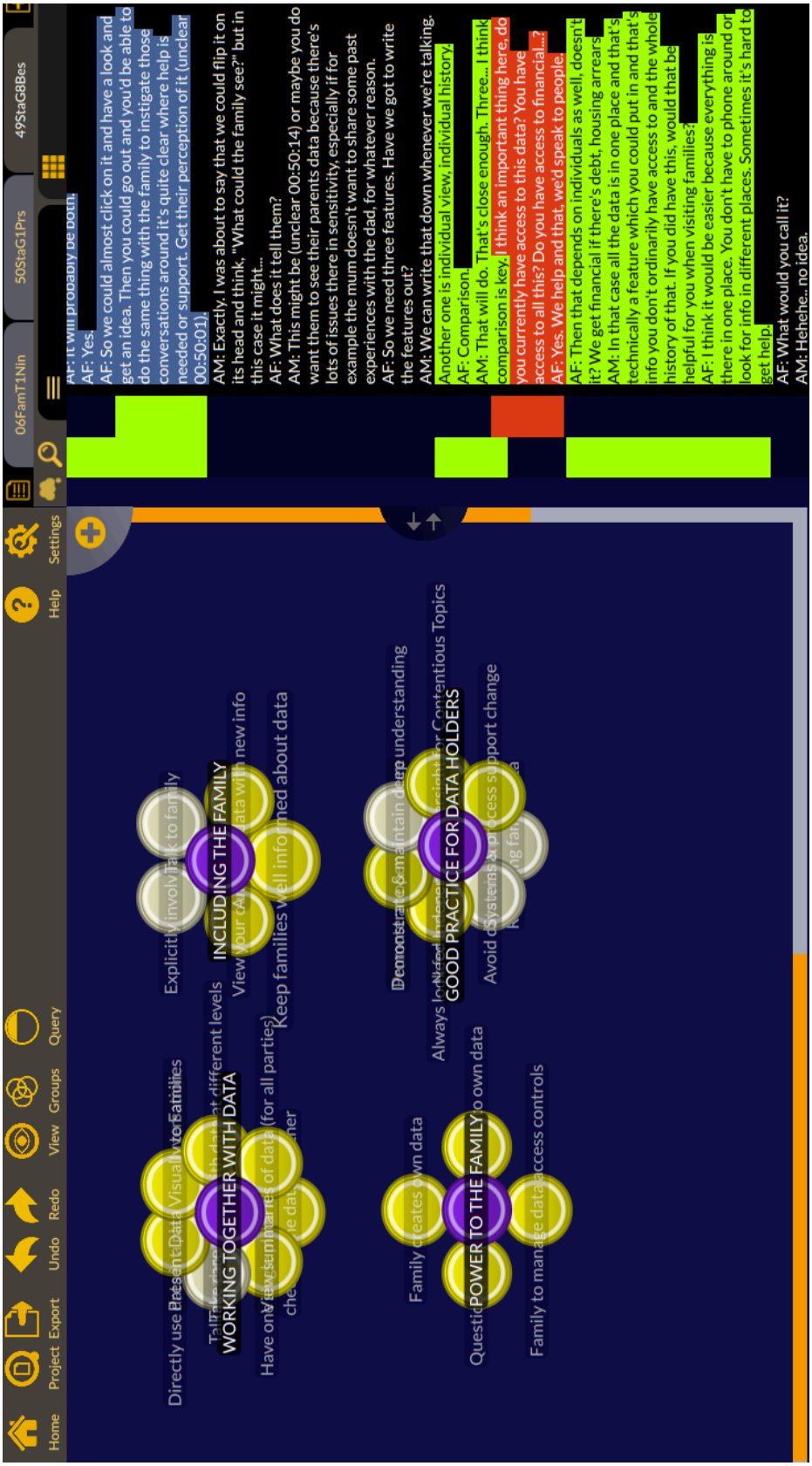


Figure 3.12: Thematic Analysis of Qualitative Data using Quirkos for Case Study One

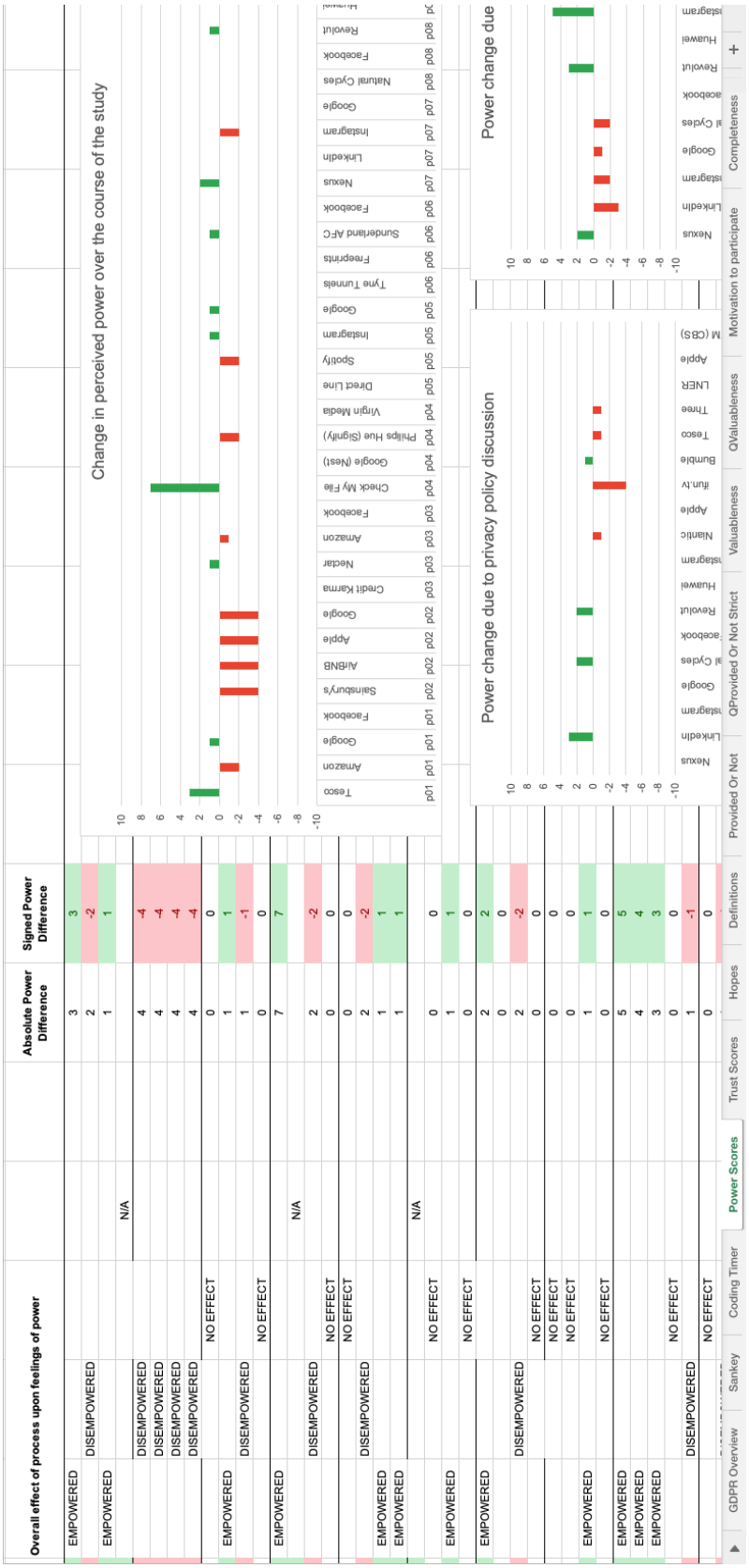


Figure 3.13: Spreadsheet-based Quantitative Analysis of Interview Data for Case Study Two

In order to find common viewpoints and extract insights from the many participatory activities conducted in Case Study One and Two, the qualitative data needed analysis. The general approach taken was to audio record (and, in some cases, video record) all interviews and workshops, and to produce a written transcript of the words spoken. Digital photos were taken to capture card arrangements, rankings and other transitory choices, as well as designs, life sketches and other participant creations. While it is possible to analyse participant designs in more detail, here their sole purpose was to provide additional context for understanding conversation transcripts. Field notes were captured during or soon after each engagement. Then a process of thematic analysis was undertaken. This involved examining the text of the transcripts (with reference to all relevant digital artefacts to add context), and identifying the underlying ideas, themes and opinions of the participants. Thematic coding is a well-established technique in qualitative research (Braun and Clarke, 2006). *Quirkos* software was selected for this purpose, as shown in Figure 3.12, due to it having a more visual organisation and simpler approach than the more commonly used *nVivo*. After initial coding of transcripts, a process of reductive data display cycles (Huberman and Miles, 2002) was used to group codes into themes which became the key findings of Chapters 4 and 5.

While the participant data in Case Study One and Two was largely free-flowing and very loosely-structured conversation, the structure of some activities allowed some data to be captured numerically, notably the sentence rankings and data card placements in Case Study One and the trust/power ratings and GDPR spreadsheets produced in Case Study Two. These datapoints were captured into Excel spreadsheets, and where appropriate analysed using formulae to produce weighted mean averages and standard deviations to help contextualise the findings. An example is shown (as evidence rather than explanation) in Figure 3.13. Due to the qualitative focus of the research, participant numbers were too low to seek statistically significant findings, so all quantitative findings are not intended to be representative of any population at large.

3.4.5 Recruitment

Research Activity	Engagement	Stage or Phase	Duration	Number of Participants	Recruitment Method
Pilot Study	4 x Home-based Interview	preliminary	4 x 2 hours	7 adults and 6 children from 4 families	Posters and Visits to Local Community Centre
Case Study One	1 x Group Design Workshop for Families	1A	1 x 2 hours	8 adults and 9 children from 5 supported families	Selected by Local Authority Care Services
Case Study One	2 x Group Design Workshop for Staff	1B	2 x 2 hours	36 support workers & related staff	Selected by Local Authority Care Services
Case Study One	1 x Combined Staff and Parents Group Design Workshop	2	1 x 2 hours	3 support workers and 4 parents from supported families	Selected by Local Authority Care Services

Table 3.1 - Context One (Civic Data & Early Help):
Participants involved in Research Activities leading into Case Study One.

Tables 3.1 and 3.2 summarise the participants involved in this research. In Case Study One, recruitment was initially attempted using posters placed in local libraries as shown in Figure 3.14 below. When this approach was unsuccessful, participants were successfully recruited with the assistance of a local community centre, which allowed me to visit a community social meeting and talk to residents about my study. This community was located in a low-income area that was known to include a number of supported families; this was chosen with a view to reaching a similar population as SILVER.

For the main engagement of Case Study One, the researcher worked with two local authorities, Newcastle City Council and North Tyneside Council, who were partners on the SILVER project. These authorities provided suitable participants who were actively involved in their Early Help programmes. In the pilot study and in the first families workshop of the main study (stage 1A), activities were designed to include children as active participants in the research. It was felt they would bring valuable contributions to the somewhat abstract creative co-design work and this also allowed observation of intra-family conversations. The final combined workshop with staff (stage 2) was designed to only include adult participants. This is because the focus on processes and on the care relationship itself was thought to be too boring and potentially sensitive for the children to participate.

Have you ever wondered what happens to

YOUR FAMILY'S INFORMATION

when you speak to councils, doctors and officials?

Are you a parent? Can you spare 3-4 hours of your time? Help us with our research and get a £50 family day out voucher!



Every week, doctors, council staff and other officials collect and store about you and your family as you go about your daily life. We call this **Family Civic Data**.

Authorities often **share** this data with each other (for example, your social worker calling your doctor) to try and give you better help and support and know you better.

We'd like to meet you to chat about this and do some creative activities so you can **have your say** on how your family's information should - or shouldn't - be shared.

We need your ideas to help make sure that computer systems that councils might build are **family-friendly**, and don't make your life harder.

To find out more, visit yourfamilydata.com where you can find answers to your questions and sign your family up!

Examples of Family Civic Data			
Names & birth dates	Relationships	Childcare	Living arrangements
Criminal Records	School attendance	Council housing	Medical records
Job interviews	Benefits claimed	Library usage	Adoptions
Child support	Social worker visits	Youth offences	Complaints

Newcastle University

Open Lab at Newcastle University

Volunteer and earn £50! Sign up at www.yourfamilydata.com

Volunteer and earn £50! Sign up at www.yourfamilydata.com
Volunteer and earn £50! Sign up at www.yourfamilydata.com
Volunteer and earn £50! Sign up at www.yourfamilydata.com
Volunteer and earn £50! Sign up at www.yourfamilydata.com
Volunteer and earn £50! Sign up at www.yourfamilydata.com
Volunteer and earn £50! Sign up at www.yourfamilydata.com
Volunteer and earn £50! Sign up at www.yourfamilydata.com
Volunteer and earn £50! Sign up at www.yourfamilydata.com
Volunteer and earn £50! Sign up at www.yourfamilydata.com

Figure 3.14: Recruitment Poster - Poster Used to Recruit Participants for Pilot Study

Research Activity	Engagement	Stage or Phase	Duration	Number of Participants	Recruitment Method
Pilot: Digital Life Mapping Study	5 x 1-on-1 interview	preliminary	5 x 2 hours	5 adults	Convenience sample
Case Study Two	11 x 1-on-1 interview (Life Sketching)	1	11 x 1 hour	11 adults	Convenience sample
Case Study Two	10 x 1-on-1 interview (Privacy Policy Reviewing)	2	10 x 1 hour	10 adults	Continuation[4]
Case Study Two	10 x 1-on-1 interview (Viewing GDPR returns)	3	10 x 2 hours	10 adults	Continuation

Table 3.2 - Context Two (Digital Life):

Participants Involved in Digital Life Research Activities Leading into Case Study Two.

In Case Study Two, no special population was needed, as the issues of living in a data-centric world affect everyone. Therefore, a convenience sample (largely 20-to-40-year-old postgraduate students from Newcastle University) was used. Care was taken to find an even split of male and female participants, but other than that no selection criteria were applied. Given their proximity to the Digital Civics programme, the participants for this study were thought likely to have a larger awareness of societal issues around personal data use and

[4] One participant withdrew from the study after the first interview of the Guided GDPR study due to COVID-19. The other 10 participants took part in all three stages.

greater familiarity with participatory co-design than the layperson. This was considered an advantage as they would require less sensitisation.

In all cases[5] for both case studies, participants were compensated for their time with vouchers—either online/offline shopping vouchers or vouchers for a family day out.

3.4.6 Ethics

All research activities referenced in this thesis were planned in advance, with interview schedules, information sheets, debriefing sheets, participant consent forms and ethics forms being completed and submitted to Newcastle University's SAgE faculty ethics board, which approved all the studies before they commenced. Ethics approvals are included in Appendix B. Most of the engagements were routine interviews and therefore did not require any special measures for safety or ethical reasons. It was made clear to all participants that they were free to withdraw from my research at any time without giving a reason. The following special measures were included in plans in order to satisfy ethical considerations:

- *Visiting private homes*: In order to protect myself and other researchers from any physical risks or any accusations of impropriety, all home visits took place with two researchers present, and contact was made with a colleague before and immediately after the interviews to confirm everything was ok.

- *Working with children*: Activities were designed to be child-friendly (not just safe, but engaging). The families workshop took place at in a function room within a park with a nearby cafe and playgrounds for children, and catering was provided. Within the room, an activity area was provided for smaller children who were not directly participating to play while their parents and older siblings engaged. There was always more than one researcher present and the research team was never alone with children.

[5] (with one exception - the staff workshops within Case Study Two. Because the participants were attending the workshops through their employers (the local authorities), we were not allowed to provide vouchers for participation.)

- *Protecting personal data privacy*: In Case Study Two, particular care was taken to design ways for researchers to talk to people about their personal data without violating participants' right to privacy. The research was positioned such that the data retrieved from companies was participants' own data, that would never be directly collected or handled by the research team. It was made clear that the researchers were only interested in what was said, not personal data itself. Initially, a privacy monitor was developed which could only be seen with viewing glasses that were in the participant's control [See ARI3.1]. This would allow a researcher to sit next to a participant who was viewing his/her personal data, without the researcher being able to see it. Additional measures to protect users' data included clear instructions on how to keep data safe before, during and after the study. A complaints procedure was also written at the request of the Ethics board.

- *Adapting to COVID-19*: As COVID-19 changed working and living conditions in early 2020, Case Study Two was adapted to no longer rely on face-to-face engagement. The in-person privacy monitor approach mentioned above was abandoned and replaced with an online Zoom-based approach. In this model, participants would share parts of their data using screen sharing instead, and could move windows off screen to protect their privacy. The full study plan for Case Study Two was rewritten for online-based participation and was re-approved by the Ethics Board.

3.5 Summation: Towards an Understanding of Better Data Relations

As Section I concludes, we have now established the background and motivation to this PhD research, and explained the approach to be used–a two track research endeavour examining both participant attitudes and pragmatic sociotechnical realities. Section II will describe the participatory research track that seeks to answer RQ1 [3.3.1] and RQ2 [3.3.2]. Section III will then synthesise the findings across the two studies to formulate a research agenda for better data relations, the operationalisation of which will be documented, along with the embedded project work track, in Section IV.

SECTION II:

UNDERSTANDING ATTITUDES AND DESIRES THROUGH PARTICIPATORY CASE STUDIES

Introduction to Section II

Section II will describe the participatory co-design work undertaken to explore RQ1 (what people want from data) and RQ2 (what people from relationships involving data). These RQs are both explored across two contexts. Following a short introduction to these two contexts and their rationales, Chapter 4 will describe Case Study One, which examines current and imagined data use approaches in Early Help in the UK, and Chapter 5 will describe Case Study Two, a study involving the exercise of GDPR data access rights by participants. Each chapter describes findings from each piece of participatory research and includes a discussion to draw insights from each Case Study. Findings and insights will then be examined for commonalities across the two Case Studies in Section III, in order to answer RQ1 and RQ2.

The Two Research Contexts: *Early Help* and GDPR

As described in 3.2.2, a two-track research approach was chosen to both explore people's viewpoints and explore reconfiguration of data relations in light of sociotechnical realities. For the participatory track, I was keen to examine the 'everyday digital life' where I had first observed problems of agency and lost human potential [1.1]. However, at the start of this PhD in 2016, it was hard to see how to frame a study around this. An opportunity arose to take an embedded role as a researcher and developer within Connected Health Cities' SILVER project (Connected Health Cities, 2017). This presented an ideal context for something more focused on a particular demographic and type of personal data: namely the relationship that families have with their civic data while receiving *Early Help* social support from their local authority. The SILVER project ran from 2017 to 2019 and was a UK-government-funded "care pathways" project involving University departments, local authorities, NHS trusts and private technology firms in the North of England. It sought to conduct participatory research and technical prototyping to explore the viability of creating an intervention that would allow professionals across many agencies to access better data about supported families while also exploring and addressing the need to ensure such families' consent to data sharing.

The SILVER project, and its Early Help context in general, became the setting for Case Study One. This choice to focus on Early Help to explore RQs around data is contextualised in Chapter 4 below, but let us first explain the term: Early Help is a particular type of social support offered by UK local authorities as voluntary help to families who are considered to be at risk of falling into poverty, crime, truancy, addiction or other issues which are both problematic for the individuals and costly to the state. Families enrolled in the scheme meet a social worker (called a *support worker* in this context) regularly who can provide advice and connect the family with appropriate health, lifestyle and social services to their needs. As part of this, the support worker has access to a variety of data from civic sources: school records, employment and benefits data, social housing data, criminal records, and more, so that they might be better informed about the family's situation. However, the families do not have any access to this data, and thus despite this being a scheme that is on the face of it intended to empower families to help themselves, it runs the risk of *disempowering* the families through a data-centric power imbalance, exactly the phenomenon identified in the background research [2.1.2]. Therefore, this context would be ideal for the exploration of both RQ1 [3.3.1] (finding out how these supported families feel about their data) and RQ2 [3.3.2] (examining the impacts of data use within a service relationship) as well as to explore how the families could imagine their data relations being improved. An important aspect of this context was that through the SILVER project I as a researcher gained access to support workers and other local authority contacts, allowing both sides of the Early Help data relationship to be explored, a goal I had already identified in 3.3.2.

The coming into effect into EU law of the *GDPR (General Data Protection Regulation)* in May 2018 presented a pragmatic research vector for both participatory research and self-experimentation into the wider aforementioned 'everyday digital life' problem space explored in Chapter 2. GDPR rights would give researcher and participants a means to examine what data was held by service providers, which would provide a suitable basis for investigation and discussion. This data would not be limited to the civic data context but potentially could include any service that an individual might use in their everyday life. This seemed an ideal and complementary context to the Early Help context—focused a little closer to individuals' own lives and personal goals, and this time focused on the average citizen or layperson rather than a specific demographic. Thus, access to digital life data using GDPR became the context for Case Study Two, which is introduced in more detail in

Chapter 5. Initially, I designed a study in this space which would allow exploration of both sides of the GDPR relationship, by working with my University's Data Protection Office and having students or staff (including myself) request their own data while also conducting interviews with those staff who would handle those requests. This study was approved by the University's Ethics Board but had to be cancelled as a result of changing working practices and increased staff workload at the start of the COVID-19 pandemic. Unfortunately therefore, this context was not able to be explored from both sides of the GDPR relationship; nonetheless it was still considered a very useful research context, and Case Study Two was designed—which, after carrying out GDPR requests of my own as sensitisation, would take 10 participants on a journey of exploration of their own personal data and service provider relationships. Post-SILVER project, new opportunities arose for me to take on an embedded role first in BBC R&D's *Cornmarket* personal data locker project, and later in Hestia.ai's *digipower* project. These two external projects were also embedded in the same research context, enabling something closer to a multi-perspective exploration of this research context than Case Study Two would have allowed by itself. These project activities are detailed in Section IV.

In both Early Help and GDPR research contexts, a variety of preliminary activities and preparatory activities were carried out to ensure full sensitisation and to validate existing understandings and research study designs prior to the start of the primary Case Study workshops and interviews. These are described in 4.3, 5.2 and 3.4.5.

This section will now continue with Chapter 4, which describes the Early Help context and Case Study One, before concluding with Chapter 5, which describes the GDPR context and Case Study Two.

Chapter 4. Case Study One: Access to Civic Data in Early Help

"If you can see the invisible, then you can see the possible and provide the opportunities for trust, commitment and ways of empowering others to manage their past, present and future."

—Ken Wyatt
(Australian politician,
teacher and equal rights campaigner)

This chapter describes Case Study One, which involved four 2-hour participatory co-design workshops with local authority support workers, parents and children from supported families involved with *Early Help (EH)* programmes [see Section II Introduction]. Its objectives were to understand family and staff attitudes to civic data holding (RQ1 [3.3.1]) and to explore the role of data within the support relationship (per RQ2 [3.3.2]). A model of *shared data interaction*, was explored, where supported families would interact with data together with their support workers.

Section 4.1 explains the EH context in England. 4.2 explains the various activities undertaken in this context. 4.3 brings together prior findings from the pilot study [4.2.2] and SILVER [see Section II Introduction], and explains how these understandings were used in sensitisation to establish a common ground for participants. 4.4 reports qualitative thematic findings: that families want to be given a voice [4.4.1], that trust can be earned through data and process transparency [4.4.2], and that families need *meaningful* data interaction [4.4.3]. [4.5] discusses these findings in the context of prior literature, focusing on the value of data involvement [4.5.1], the need for human interaction [4.5.2], and the pros and cons of the shifting of the *Locus of Decision Making (LDM)* towards the family [4.5.3].

4.1 Context: Data Use in Early Help

4.1.1 Data-centric Family Intervention in UK Social Care

The UK's social care system has been shaped by a series of efforts to improve the lives of children through data-driven interventions: *Every Child Matters*, *ContactPoint*, and the *Common Assessment Framework (CAF)* (Department of Health, Home Office and Department for Education and Skills, 2006; Gheera, 2011; Holmes *et al.*, 2012; Cornford, Baines and Wilson, 2013). These interventions later took a broader family focus (Wilson *et al.*, 2011; Malomo and Sena, 2017) through programmes such as *Think Family* which focused on *family intervention* (Cornford, Baines and Wilson, 2013; Parr *et al.*, 2016) as a primary approach. Social workers learn about and get directly involved with the lives of targeted young people and their families in order so that they might overcome specific difficulties.

2012 saw the introduction of the *Troubled Families Programme (TFP)* in England, which argued that £9 billion of civic spending was due to just 120,000 families and thus these families should be *'turned around'*. Local authorities had to identify *troubled families*[6]—those *'at risk'* families experiencing multiple issues from a list including unemployment, overcrowded housing, poor education, mental health issues, disability, low income, poverty, truancy, crime and domestic violence. For each family where impact could be proven, funding could be reclaimed from central government (Bate and Bellis, 2018). This encouraged extensive collection and use of data about each supported family to track and demonstrate progress and impact.

This shift towards using data mirrors the societal rise of data-centrism [2.1], but also a wider public sector trend: Pressed to to demonstrate performance and deliver measurable, consistent results, human services including social care, health care and education have all prioritised the collection and use of data about their clients or service users. This use of

---

[6] The term 'Troubled Families', popularised by the TFP, has fallen from use, as it was considered to be negative and judgemental. A latter term 'vulnerable families' has also been criticised for being disempowering. Most councils now refer simply to 'families' or sometimes 'supported families', and the rest of this thesis adopts this convention.

data by the state to represent and think about families is considered problematic (Cornford, Baines and Wilson, 2013; Barbosa Neves and Casimiro, 2018), as records can include both objective facts as well as subjective information such practitioners' observations or numerically-quantified measurements of risk. This increases risks of inaccurate data or unfair judgement. Individuals typically have limited access to verify this data. While families can ask about their data, few do, and the service organisations become de-facto *gatekeepers* to family data (Corra and Willer, 2002). The situation has exacerbated as data-driven approaches to family care have been encouraged through policy and reports about improving quality of the sector (Field, 2010; OFSTED, 2015; Bate and Bellis, 2018; Department for Education, 2018).

4.1.2 Current Practice: Early Help Case Records as a Source of Truth

Through the 2010s, *Early Help (EH)* programmes became a key social care offering from most UK local authorities, aiming to intervene in families' lives before costly additional help is needed. To meet the proof requirements of TFP, new processes were established. Support workers would now carry out an *Early Help Assessment* (a guided enrolment questionnaire) to create an *Early Help Record* for each supported individual and their family, stored in case management or *eCAF* systems supplied by companies such as CareFirst and LiquidLogic. To gain a holistic perspective of a family's situation, a process of information gathering and family-centric inter-agency collaboration is adopted. The record is supplemented on a periodic basis by data from other agencies. Through contact with local authorities as part of its work, the SILVER project [see Section II Introduction] learned that such data is collected ad hoc, via emailed spreadsheets, phone conversations, and in-person meetings. One such meeting is the *Team Around the Family (TAF)* – a bespoke grouping with representatives from other agencies such as police, schools or housing agencies. EH data is used to measure family progress relative to the CAF (Holmes *et al.*, 2012).

Support workers are encouraged to use data as evidence at all stages. A 2015 Ofsted report recommended further evidence-centric standardisation to address inconsistencies and deliver better care (OFSTED, 2015), triggering the transformation of EH into a data-driven service, where professionals seek more and more data about 'at risk' individuals. This belief that better data can drive better care is baked into national policies:

> *"IT systems are most valuable when practitioners use the shared* [between agencies] *data to make more informed decisions about how to support and safeguard a child."*—Department for Education (2018)

The technical reality of pursuing data sharing in cared is problematic. The information ecosystem is vastly complex (Copeland, 2015) with each part of the system having its own ICT arrangements. Teams work in isolation, using different systems and apps. Limited arrangements are in place to facilitate information sharing between data-holding authorities (which can include local charities to which care services are outsourced). Administrative boundaries differ for different authorities and agencies, complicating matters further. With each local authority procuring their own IT systems in the absence (despite recommendations (Harbird, 2006)) of any centralised systems or information sharing standards, the ecosystem has become severely fragmented.

Support workers face barriers to accessing needed data. Care workers can rarely access health data from GPs and have to rely on school nurses, health visitors, or the individual's own word. Information is often shared informally through telephone conversations, meetings or emails, not supported by technical integration. No one team, agency or authority can have a full picture of an individual's data (Malomo and Sena, 2017). Different operating policies, consent agreements, privacy regulations, technical access levels, system functions and staff competences result in different interpretations that further limit what data is shared (Malomo and Sena, 2017). Data should flow freely through the system in the service of individual care, but it does not. The public sector has a closed and fragmented ecosystem (Pollock, 2011).

Processes such as TAF meetings and the attempt to unify all information onto a single EH record can be seen as a recognition of this failure in the system to produce a single source of truth or understanding of individuals from a 'whole life' perspective. Attempts to create and expand the EH record as a central representation of truth about the family can be seen as data-centric solutionism (Morozov, 2013), being applied to a problem that was created by a data-centric approach in the first place.

4.1.3 Rethinking the Role of Data in Early Help Support Relationships

Families' only awareness of EH data occurs when support workers or TAF professionals choose to share elements with them. Such sharing is usually verbal, unrecorded and rarely complete. Critiques suggest the acquisition of additional data consolidates power in practitioners' hands and undermines the families they are meant to be supporting (White *et al.*, 2010; Neff, 2013; Crossley and Lambert, 2017). The scattering of data across so many different systems and organisations, combined with informal processes for sharing, presents serious risk of privacy breaches or mishandling of people's personal data.

Consent may be violated if data, collected for a specific purpose, passes to another authority for some new purpose without explicit consent. The idea that the EH record is a source of truth risks disempowering families further, countering the empowerment goals of the programme itself. Erroneous data is likely, and this may cause prejudice or unfair decisions being made against families. Individual privacy could be violated, or individuals put at risk, if a domestic abuser or criminal gained access to the data. The failure of such systems to properly represent families (Cornford, Baines and Wilson, 2013) produces further risk. Information shared by one individual in confidence could easily be seen by another family member, with potential severe psychological consequences, such as an adopted child finding out in the wrong way that they are adopted.

Data is not neutral (Gitelman, 2013; Neff, 2013). Collecting data for a specific purpose (rather than objectively) undermines local professionals' ability to deliver care (Cornford, Baines and Wilson, 2013; Lowe and Wilson, 2015). The collection and use of data may, instead of helping a family, reinforce the existing asymmetry of power that exists between them and data-holding organisations (Cornford, Baines and Wilson, 2013).

This presents an ideal context to explore RQ2 [3.3.2], specifically the impact of data use upon service relationships. Following preliminary sensitisation research [4.3], a study was designed to work with both families and support staff involved in EH. This study examined individual perspectives per RQ1 [3.3.1] but also looked at the power imbalance and the effectiveness of the support relationship as a whole, in terms of its goal to empower families to build better lives for themselves and get them to a point where they no longer need support. As part of RQ2 [3.3.2], the study also explored possible alternative models for the

use of data within EH relationships. The design and approach of the study is described in more detail in the next section.

4.2 Activities and Methodology for this Context

Within this context I carried out three research activities between 2017 and 2019:

4.2.1 Embedded Research and Development Placement in CHC's SILVER Project

From March 2017 to March 2019, I joined Connected Health Cities' *SILVER* project (Connected Health Cities, 2017) as a part-time research engineer alongside my PhD. This research project was funded by the UK's Department for Health (now the Department of Health and Social Care) and brought together local authorities, health authorities, University researchers and technology partners in the North East of England, in the Early Help context. Its goal was to explore how to unify civic data about a supported family, with their consent, to allow support workers to provide better care to those families. This made it an ideal place to explore my research objectives. It used direct research with families and support workers to inform the system requirements. This also provided an opportunity to deepen understanding of the use of data within Early Help support relationship (RQ2), and both parties' attitudes to this highly personal and real civic data (RQ1). My role was two-fold: as a software engineer, to design and develop user interfaces that would be used to view this unified data, and as a participatory researcher, to assist with the design and execution of focus groups and workshops with staff and supported families that could inform the proof-of-concept data system being built. It is that work as a researcher that forms part of the Case Study in this chapter. The development work is not considered part of this Case Study, but rather as an first embedded research setting [See Section IV Introduction] and will be referenced in the subsequent chapters, especially in Chapter 8, Chapter 9 and in 10.1. The final report from the SILVER project was not published but a public summary is available (Northern Health Science Alliance, 2020).

4.2.2 Pilot: Understanding Family Civic Data Study

In the summer of 2017, in the MRes year of this doctoral training programme (alongside my involvement with the SILVER project), I carried out an initial participatory field study in

order to deepen my understanding of data use and attitudes within this context and to develop appropriate research methods. This study consisted of home visits to four different families in the North East who had interacted in the past with social care & support services [Table 3.1]. During the course of these two-hour visits, I carried out participatory co-design activities and interviewed the families (both adults and children) about their civic data, and in particular their views on how risky different types of data were and how that data should be handled. Fieldwork took place prior to the start of this PhD; however, the data analysis and publication of the findings took place within the scope of this PhD, and directly inform the main Case Study, both through its methodology and as prior work to build upon. This pilot study is given special status in this thesis (as described in 1.3.1) and was published as a first author paper (Bowyer *et al.*, 2018), included in full in Appendix A.

Within the four pilot study home visits [see photograph in Figure 3.8] the following activities were carried out:

- 'Family Facts' Sensitisation [3.4.1]
- Data Card Sorting [3.4.2]
- Discussion stimulated by Example Data, News Articles and Interface Mockups

4.2.3 Case Study One Workshops: Data Interaction in Early Help

In the summer of 2018, informed by the SILVER project and the pilot study, I designed and conducted my first major case study of this thesis: a series of four participatory co-design workshops with people directly involved in *Early Help* relationships in North East England. The workshops were funded by CHC, and were led by myself. They were designed with a dual purpose: to inform the design of the SILVER system but also to serve RQ1 [3.3.1] and RQ2 [3.3.2] of this thesis. These workshops sought initially to validate findings from SILVER's earlier research and my pilot study.

The study comprised two phases:

- First in phase 1, to work with families and support workers as two distinct groups, to identify their different needs [Table 3.1; 4.3.5],
- then, in phase 2, to work with both groups together in an attempt to address their unified needs through novel approaches to data interaction.

The objective for Phase 1 was to develop a deeper understanding of what supported families (workshop A) and support workers (workshops B1 and B2) perceive as problems with data use in the Early Help context, and to explore perceived solutions to these problems. Phase 1 served to revalidate earlier findings of early work, working with each group separately. A further objective was to understand existing data practices and whether they work, or need improving, and to identify known issues.

A second phase workshop (workshop C) was designed to specifically focus on the use of data *within the support relationship* [4.3.4]. This was a joint workshop involving staff and parents working together to explore the *Shared Data Interaction* concept that emerged in Phase 1 [4.3.4]. In Phase 2, the two groups would collaborate to design imagined data practices and interactions. The objective was to understand how *in practice* staff and families could imagine themselves using data together.

Across both phases, various participatory methods were used, as detailed below. Workshops were audio recorded and transcribed. Transcripts were analysed thematically, and in some cases quantitatively [3.4.5]. Thematic findings are documented in 4.4. This Case Study's findings are documented in this chapter, but also contribute to the general findings about RQ1 [3.3.1] and RQ2 [3.3.2] presented in Chapter 6. The workshops are summarised in Table 4.2:

Workshop	**Engagement**	**Phase**	**Number of Participants**
Workshop A	Design Workshop for Families	1	8 adults and 9 children from 5 supported families
Workshop B (2 instances)	Design Workshop for Staff	1	36 support workers & related staff (in total)
Workshop C	Combined Staff and Parents' Design Workshop	2	3 support workers and 4 parents from supported families

Table 4.1 - Case Study One Group Design Workshops.

Within the workshops, the following activities were carried out:

Phase 1: Co-Designing Civic Data Access for Early Help

Workshop A (Co-designing with Families):

- Data Card Sorting [3.4.2]
- Sentence Ranking [3.4.1, 4.3.5]
- Ideation Decks [3.4.3]
- Poster Design [3.4.3]
- Scenario Discussions [3.4.2]

Workshops B1 and B2 (Co-designing with Staff):

- Sentence Ranking [3.4.1, 4.3.5]
- Ideation Decks [3.4.3]
- Poster Design [3.4.3]
- Scenario Discussions [3.4.2]
- Discussion of Interface Mockup [3.4.2]

Phase 2: Exploring Shared Data Interaction

Workshop C (Combined Parent/Staff Workshop):

- Sentence Ranking [3.4.1, 4.3.5]
- Storyboarding Training [ARI4.3]
- Scenario-based Storyboarding [3.4.3]

The next section goes into more detail into how the preliminary work with the pilot study and SILVER established a baseline for families' and support workers' attitudes to civic data use and access. Section 4.3.5 details the commonalities and discrepancies in attitudes between the different participant groups, including details of how these baseline attitudes were validated and captured using the Sentence Ranking method.

4.3 Preliminary Explorations of Family Civic Data: Families' and Support Workers' Perspectives

The first step in study design was to achieve sensitisation [3.4.1] as a researcher, specifically three ways:

- to understand what data is stored about families [4.3.1];

- to understand the family perspective [4.3.2]; and
- to understand the support worker perspective [4.3.3]

4.3.1 What is Family Civic Data?

In order to communicate effectively with participants, an understanding of what data is held in an EH context was needed. Through the pilot study [4.2.2; Bowyer *et al.* (2018); Appendix A] and a grey literature review within CHC's SILVER project, a taxonomy of *Family Civic Data* was established (Bowyer *et al.*, 2018), as shown in Table ARI4.1.

Early recruitment attempts revealed that data is seen as an abstract concept in people's daily lives—a dry, technical topic that many families felt unqualified to talk about. The concepts needed to be made relatable. Drawing on the work of Brandt and Messeter (Brandt and Messeter, 2004) in creating *design games* (which observes that game pieces can be used to create common ground and as *things to think with* (Papert, 1980; Brandt and Messeter, 2004), a set of *data cards* [Figure 3.7] were created. These cards made family civic data categories visual and tangible. Serving as *boundary objects* (Star, 2010; Bowker *et al.*, 2015) the cards aimed to bring researcher and participants' worlds closer together, and to make data relatable to life experience. The design approach for the data cards is documented in [Appendix A; Bowyer *et al.* (2018)].

The cards were used as research stimuli [3.4.2] within the pilot study, which involved participatory design activities and design games in family homes. Once families understood data as *stored information about their lives* they were able to very effectively engage and talk about it. The use of the games and the cards was very successful, keeping a light and playful environment and making the topic relatable. The cards enabled families to talk freely about their own lives and views without feeling personally interrogated, as they were dissociated from the participants' lives.

4.3.2 What is the Family Perspective on Their Civic Data and its Use?

Families participating in the pilot study and in SILVER's early research cared very much about what happened to their civic data, contrary to peers' expectations. They perceived a variety of risks due to data mishandling including identity fraud, criminal targeting and psychological harm. Families felt that data could easily misrepresent them through errors,

prolonged storage of data beyond its need, or the recording of unfair judgements and opinions. Families wanted to view the data stored about them. They wanted a set of basic rights—to be informed, involved and accurately represented, with the ability to see, explain and correct their data to ensure it is fair and accurate. They wanted to know that their data will be handled sensitively and only by those that need to know. They believe that having such capabilities would help them to be able to work together with representatives of the state in a more positive relationship.

The SILVER project [see Section II Introduction] conducted qualitative interviews with EH-supported families, and unpublished findings from these interviews reinforced the need for a more consultative role for families in data handling. Participating families identified that while there was a willingness to consent to information being shared in order to improve their care, they had very little understanding of how it was used and could not be deemed to have given informed consent to the way their data is currently used.

4.3.3 What is the Staff Perspective on Family Civic Data and its Use?

SILVER worked closely with local authority staff, providing an understanding of support worker and staff perspectives. SILVER conducted a series of 'Amy's Page' (Wilson, Wilson and Martin, 2020) focus groups/workshops with support workers and other local authority representatives, where staff mapped out information needs. They desired greater access to health information, particularly mental health indicators. The participants revealed a desire to gather as much data as possible about the families they were working with. It was viewed as a useful raw material that enabled them to do their job better.

Collectively, findings from the pilot study and from SILVER showed a conflict between the desires from families and support workers. Families wanted more involvement and less reduction to data. Support workers wanted to amass more and better data. In part due to its funding and what I would consider *solutionist* (Morozov, 2013) framing, the SILVER project gave priority to the support worker perspective as key requirements. It continued to pursue the building of a richer data interface for support workers. At this point, my research objectives and those of the SILVER project diverged. I was not ready to 'take sides' nor to pursue a purely data-centric solution. I wanted to explore whether it might be possible to

satisfy the needs of both parties and to maintain focus on human-centricity and the need for a balanced relationship.

4.3.4 Shared Data Interaction: An Equitable and Mutually Beneficial Data Use Model

In attempting to reconcile the conflicting desires of families and support workers, **Shared Data Interaction** was conceived. What if, instead of workers gatekeeping data access and using it by themselves, data could be looked at, examined, and updated together, during the face-to-face encounters between families and their support workers? This could potentially bring the benefits of HDI (increased agency, negotiability and legibility) (Mortier *et al.*, 2014) to families (and also to workers), while also serving as a boundary object that might improve the relationship itself (Bowker *et al.*, 2015). I theorised that this would allow families to access unseen data while also enabling support workers to 'fill the gaps' in the data by simply asking questions.

This concept emerged in part from phase 1 participants in the research engagement, and became a main focus for the second phase. In this regard, the workshops would not only explore current practice, but also motivate participants to imagine new practices that could serve their needs better. This would allow a preliminary assessment of shared data interaction—whether it might address both groups' needs and whether it would be perceived to benefit the support relationship. Even if the model was not well-regarded, such an exploration would still help to put participants in a speculative, co-design mindset that could elicit deeper insights about how civic data *should* be used.

4.3.5 Sensitising Participants and Discovering Shared Values

In all workshops, it was important to ensure participants arrived at a common understanding of their 'design brief'. To achieve this, and to validate if these participants' perspectives matched those observed previously, an activity called *Sentence Ranking* was conducted. Participants considered a number of opinionated statements and ranked them according to:

(a) whether they agreed, disagreed or were neutral on that statement, and

(b) whether or not they felt that statement was important.

An example statement is:

'Families should always be able to talk to someone about their data.'

The complete list of sentences is in ARI4.2. These sentences were collated from family and staff perspectives observed during the pilot study, in SILVER's research engagements, and from researcher observations from the SILVER project [see Section II Introduction].

Through discussing and reaching consensus on these opinions, families and staff would be in effect *agreeing requirements* that could inform their thinking during design activities. By conducting this same activity across all participant groups and across both phases, this would also allow comparison between the different groups, to identify differences and find shared values. Within each workshop, groups of participants sat at tables of 4 to 6 people, and each table provided its own sentence rankings.

The resultant rankings were analysed as described in ARI4.2. The visualisation of these findings on shared values is shown in Figure 4.1.

As the figure shows, there was universal consensus that:

- families should be able discuss their data with someone from the authorities,
- public sector officials cannot make good judgements solely by looking at families' data,
- data cannot adequately represent a family,
- families should be treated as more than just what their database record says,
- information stored about them must be fair and accurate,
- families must have rights to see it and how it is used, and that
- support workers really need to know mental health details of family members.

Participants felt it important to address that current consent practices were inadequate. There was also strong consensus that families did not want to be responsible for looking after their own data, though this was felt to be an unimportant matter.

Participants showed considerable contention over whether or not support workers should be able to access historical family records [4.4.3], about how families would feel about the

collection of data about them and about likely attitudes to setting privacy preferences. There was moderate consensus over most other sentences.

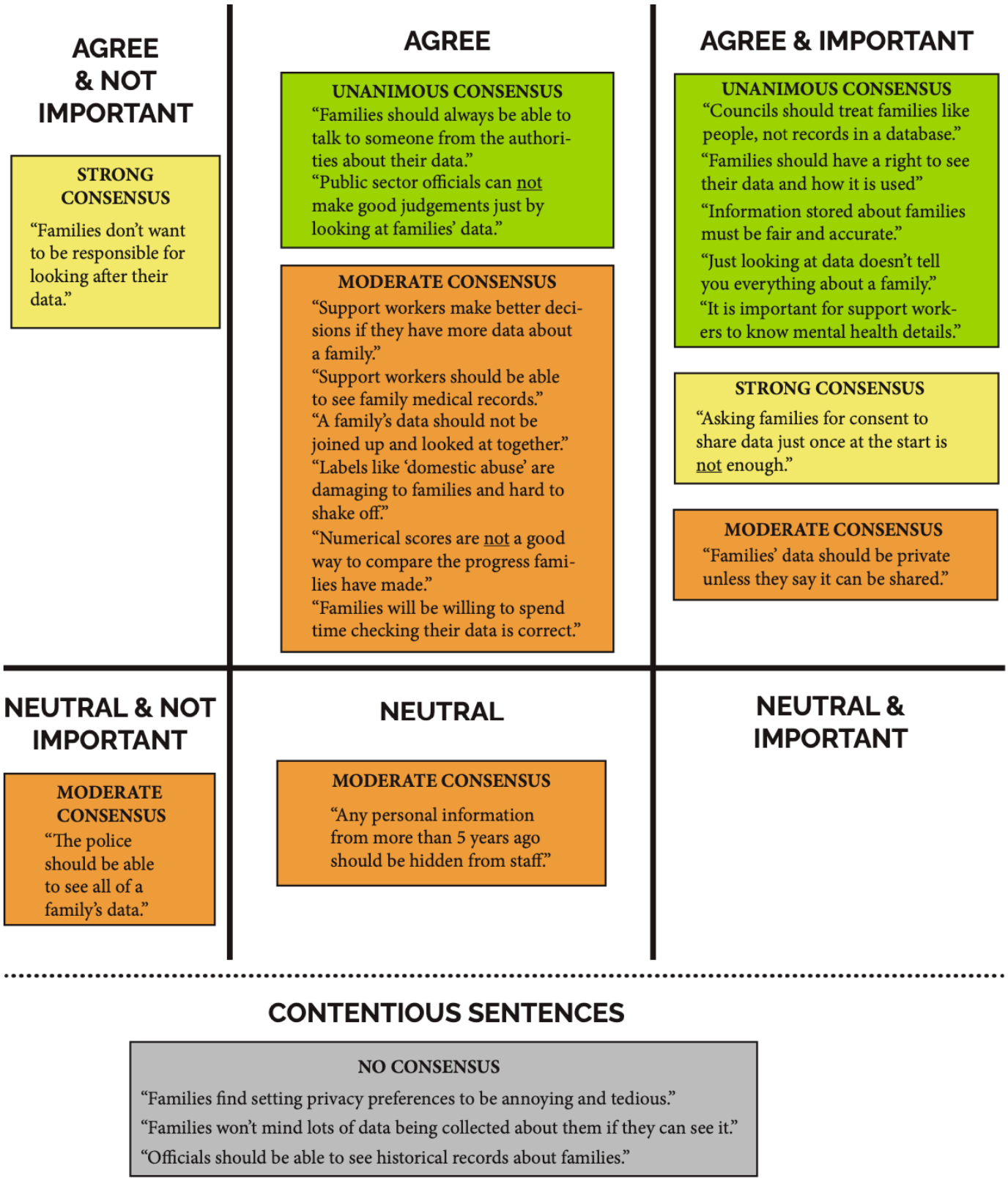


Figure 4.1 - Participants' Shared Values Deduced from Sentence Rankings Data

After this exercise, participants were considered sensitised and went on to carry out the other co-design activities [Table 4.1]. Transcripts of these activities were analysed to produce thematic findings, which are detailed in the next section.

4.4 Thematic Findings

The 120,000-word corpus from audio recordings of workshops A, B and C was divided by activity, group, and family/staff focus into 85 different source texts. Each text was thematically coded. The coded texts were analysed through four cycles of analysis (Huberman and Miles, 2002). During this reductive process, participant creations, activity outputs and ranking data were referenced for context. Results of the thematic analysis are presented below. In 4.4.1 the three main themes and subthemes are introduced. Each theme is detailed in 4.4.2 to 4.4.4, including participant quotes. Notation for quotes is explained in ARI4.4.

4.4.1 Themes & Subthemes

Since workshop discussions were framed as explorations of data use within the EH relationship, thematic findings are expressed as desirable best practices. These best practices are divided into three themes:

- Meaningful Data Interaction [4.4.2],
- Giving a Voice to the Family [4.4.3], and
- Earning Trust through Transparency [4.4.4].

Explicit and implicit statements from participants, contextual clues, and accumulated knowledge allowed a judgement to be made as to whether each discussed best practices was:

- commonly in use ('current'),
- happening occasionally/partially ('emergent') , or
- not yet occurring at all ('imagined')[7].

[7] As judged at the time of the workshops—Summer 2018.

Tables 4.4, 4.5 and 4.5 show subthemes, illustrated with participant quotes, as well as the current, emergent or imagined status of each subtheme. Structuring the themes makes the findings actionable for social care organisations.

Subtheme	**Description & Quote**	**Status**
Understandable Information Summaries	To maximise understanding, simple summaries of the information within families' data should be available to both families and support workers. Visualisations should be used to ease comprehension, and information should be contextualised at different levels (individual, family, community). *'There's so much data that's stored. For me, for a parent [I want] to understand that through a text or email but just in point form. […] The less written, the better for the parent. [What we need is] a small synopsis […] like a summary view.'* [Parent, SQ44] *'Some families will go, "Well you know that information because it's all there somewhere." We're like, "Yes, but we don't want to trawl back to eight years ago." There's reams and reams and reams of it [data].'* [Worker, SQ40]	Emergent
Interact with Data Together	Support workers should work to actively counter the knowledge imbalance by informing families what their data says. They should make use of specific datapoints as talking points to aid planning conversations. *'You could have a table, you'd look at where they are and where they could be. [You could say] "This is where you are now but if you [take these specific steps], even though you've got a criminal record, you could progress to this level."'* [Worker, SQ29]	Emergent / Imagined
Direct and unified data access	Individuals should be able to directly access their civic data through a personal interface; this should be a	Imagined

Subtheme	Description & Quote	Status
	single, common place where all of an individual or family's data is brought together to give a complete and consistent overview to all parties with a need to know. *'[I'm imagining an] online database of personal family info accessible [only] by people, practitioners that have permission […] I would say that it's only who you want [to give access to, that can see it]. You would have your private code which you could hand out, like the doctors give you appointments.'* [Parent, FQ8]	
Ongoing Data Access and Support	It is not sufficient simply to give access to data. Families should be able to access information in their own time and should be supported in understanding it. Most importantly they should be able to ask questions, challenge data records or start a conversation to discuss their data at will. *'[The families would have] a little app which they can log into and read all their information - what's recorded about themselves, […] who we share the information with […]. If they're not happy […] they can fire off an email to us and let us know what they disagree with or if they want their information taken down or their consent.'* [Worker, SQ51]	Imagined

Table 4.2 - ***Theme 1 - Meaningful Data Interaction for Families.***
Subthemes & Participant Quotes.

Subtheme	Description & Quote	Status
People not Records	Support workers must always treat people like individuals, that are more than a data record. They should review family data before contact, but must always engage at a human level too, avoiding making any judgements based solely on data. Worker A: *'You should never make a judgement on data… that data could be wrong.'* Worker B: *'It takes individuality, working with that person as well, doesn't it?'* [SQ11]	Current / Emergent
Checking Data Together	Families should be explicitly invited to review, discuss, check, correct and approve data records. Data recording should be visible, and workers and families should check data together. *'[The parent could] countersign. [The worker would] say, "I feel that we've talked about this today so I'm going to write that down. I'm going to show you. Can you sign and me sign if you're happy and I'm going to share this." That's a bit [better].'* [Parent, FQ12]	Emergent / Imagined
Changing Lives Means Changing Data and Changing Consent	Recognising that families' lives are in constant flux, routine reviews of data should occur, and they should be invited to regularly review their choices regarding data collection, keeping and sharing. All systems and processes should treat data as fluid and flexible, not static unchanging facts. Feeds of recent changes should be available to both parties. *'[There's] this perception of something sticking with you even after you've potentially reformed. […] That's something that happened a long time ago and that judgement is still there but [you'd be wondering] "Okay, is it [true]?"'* [Worker, SQ61]	Imagined

Subtheme	Description & Quote	Status
Individual Agency & Family-sourced Data	Individuals should be able to create or contribute their own data to tell their own story and annotate particular datapoints with their own explanations. Worker A: *'If you read information […] about me, you wouldn't expect to meet the person you meet.'* Worker B: *'That's it. It's the same for everybody.'* Worker A: *'[…] It just [has] basic things in most of the time, doesn't it […]? You're not a person [in the data record] are you really?'* Worker B: *'[I'd like it if you could] give your bit of personal data, your own story.'* Worker A: *'Yes, because everybody makes mistakes and there's probably thousands of people out there who have got a criminal record and have never done anything since. [They're] getting judged by having one thing [but they should be able to write] "Yes, I did this because of this situation but this is what I've done to make myself [better]…"'* [FQ10]	Imagined
Granular Access Controls	Families should be given controls to manage access to their data and configure and change preferences at a fine-grained level. *'[Families need to] feel they're being involved. […] [We need to be able to] sit together and say, "Right, that's the information I'll allow you to share. I don't want that bit shared. But this bit, because it will help me and the family […]".* *Say in this [scenario] family, she might have been married before and had domestic violence so she doesn't want that bit shared, that's in the past. So, it's [only] certain up-to-date information about the family [that would be shared]*	Imagined

Subtheme	Description & Quote	Status
	because this [the family suggested by the data] isn't her family.' [Parent, FQ16]	

Table 4.3 - ***Theme 2 - Giving a Voice to the Family****.*

Subthemes & Participant Quotes.

Subtheme	Description & Quote	Status
Transparent, Respectful Data Handling	Support workers should treat families' data with the utmost respect, keeping it safe, ensuring it is not used beyond its intended purposes, shared without consent or put at risk. When talking to families about data, it is helpful to focus on positives and strengths and not use it as a means to criticise. *'There was a time where I was at the doctors' and they asked how many units of alcohol I drank, and I said, probably about three bottles a week, at the time, not any more but later on [the support worker] pulled me up on it and they had it down as three bottles a day. That could have caused an issue was anyone ever to ask.'* [Parent, CQ7]	Current
Always Seek and Demonstrate Greater Understanding	Support workers should always assume that their understanding from data is incomplete and should seek to learn about individuals and build a more complete picture of their lives. By showing this effort and their growing understanding, they will engender trust. *'You don't want to reduce them to this number in a database. You want to understand their actual experiences and support them in getting better.'* [Worker, SQ74]	Emergent
Pro-actively Challenge Data-	Support workers and agencies can recognise that current systems and processes are data-centric and imbalanced,	Imagined

Subtheme	Description & Quote	Status
centric Norms	and can strive to change this through their actions: being as open as possible about how families' data will be handled, ensuring that proper oversight mechanisms exist for data handling especially in the sake of contentious issues, and that data is shared openly but consensually between authorities. *'It hasn't been explained property to this [scenario] family that their information will be shared with other professionals. So, they've been left feeling really let down and probably quite angry about it. So, although that information does need to be shared, they [the support workers involved] ought to make the family properly aware that information will be shared.'* [Worker, SQ18]	

Table 4.4 - ***Theme 3 - Earning Families' Trust Through Transparency****.*
Subthemes & Participant Quotes.

4.4.2 Theme 1: Meaningful Data Interaction for Families

Through discussions with families and support workers a deep understanding of what sort of families' ideal data interactions was obtained. Setting aside interface considerations, which were not the main focus of the inquiry, and focusing on the wider sociotechnical context around the data and its access, it seems that in order to maximise understanding for all parties, data interaction needs to be *meaningful*—the first theme of these findings. Encompassed within this concept are:

- the need for understandable and effective summaries and visualisations,
- the need for direct and ongoing data access with human support, and
- the recommendation for families and support workers to interact with data together within the support interaction.

Understandable Information Summaries

Written summaries of information were independently considered to be critical for both parents [SQ44] and support workers [SQ40]. These could also be used as a mechanism to protect privacy, by keeping sensitive details hidden:

'In that example, depression, ten year ago, that shouldn't be on there for the support worker. All they should get is if Social Services have been involved and it should just be, "Please contact for more information." […] [The system should stop workers from] getting a list of all the kids who have ever missed dental appointments or when you were depressed ten years ago. […] There needs to be a thing where it's, sort of […] key trigger words, where if the word comes up a lot of times, it spots the patterns. Whereas, if [a problem] is mentioned once, it should only be [shown] at the highest level.' [Parent, CQ10]

Because the amassing of large volumes of historical data is expected, families expect (though are not happy about it [FQ6]) that any aspect of their past life may be *findable*: *'We go to them and say, "We're aware that you've got these issues going on" […] and not one family I've ever met has said, "How on earth have you got that information?"'* [Worker, SQ42]. Managing expectations can be problematic [SQ40] and some workers felt they should not be given greater data access, fearing greater liability to *'trawl through data'* so that they know everything.

This need for summaries can also be seen an echo of Gurstein's call for *effective data use for everyone* (Gurstein (2011)). It is not sufficient to simply open up public sector databases to allow individual record access; families need not just the opportunity, but the technology, skills, formatting, interpretation and sensemaking to make the access effective. Some individuals may lack *'proper access to a computer'* [Parent, CQ9]. Data tables are insufficient and may need to be supported by visualisations: *'Some families might not understand [a data viewing interface]. They might not be technical… I think sometimes it's easier to do it in pictures.'* [Worker, SQ43]. Participants suggested pie charts, graphs, spider diagrams and timelines [SQ30, SQ31] or even an audio interface for the visually impaired [SQ45] to aid understanding. Visualisations also need verbal explanations [CQ11].

It is not clear who could or should do the skilled knowledge work of creating these representative and accurate tailored summaries and visualisations.

Interact with Data Together

Directly using data together within a support conversation is seen as a key element of making data interaction meaningful for families. For support workers, the use of data can form *'a way in'* or conversation starter:

> *'[Showing the data could be] an ice breaker [with] a new case. So, "We've got this information; can you tell me more about it?" That opens it up, like a can of worms and it all just comes out; you know what I mean? Then you're able to have that open and honest conversation with them to see what level of support that they need.'*— [Worker, SQ28]

The showing of data performs an additional important purpose, combatting the lack of *awareness* of what data exists and who holds it [SQ39]. Currently, much of the data stored about families is invisible to them: *'Families really only see the data that we [support workers] want to present.'* [Worker, SQ37] Regardless of families' legal rights to request copies of their data, it seems this right is rarely used [SQ38], and typically only when filing complaints. Lack of awareness can not only cause suspicion [SQ17], but also incorrect assumptions that support workers *'already know everything'*.

Participants particularly recognise the value of referencing data points over time (such as a record of welfare scores that support workers have previously given them), for example to track progress [SQ29, shown above in Table 4.2]. This could motivate and reinforce progress [SQ6] by relating behaviours to consequences [SQ32]–essentially facilitating data-based decision making. Reviewing historical data is preferable to verbal description: *'Whenever you go through stuff like that [verbally], especially historic stuff, they can be quite remote so [having the data in front of you] would be good for that.'* [Worker, SQ33].

Direct and Unified Data Access

Despite the reality that families currently have no direct access to their civic data, family participants all eagerly described designs including apps, intranet terminals, online chat facilities, and self-service webpages, all offering individuals the ability to view their own data; there is a clear demand for *personal data interfaces*, which could empower families to use their own data:

'They could quickly tap onto the app […] and show somebody else where they're at.'—[SQ54]

'Our first [idea] is the lovely [child's name] has made an app. [It's] free to download, you can make your own password and there's going to be a button on it so you can press it and then query the information that's held on you straight away.'—[Parent, FQ7]

Workers and families shared a desire for one single point of access for data, useable by all parties [SQ25, SQ26], though families *'don't want to be responsible for looking after all our data'*[FQ17, S5]. Bringing together data from multiple sources would allow patterns to be spotted by correlating data from different sources, which workers perceived would help their preparation:

'[This imagined interface] would provide individual histories but you could also pull them all together so you can prepare, so for instance if mum was having some significant issues with mental health, you might be able to correlate the [child's] school attendance alongside that and find out why that's happening.'—[SQ8]

Ongoing Data Access and Support

Families, being accustomed to accessing information in other parts of their lives through smartphones and web interfaces, expect that any civic data interface would allow them to access data *'in their own time, at their own pace'* [Parent, CQ12]. Currently access only possible via the support worker, functioning as a *gatekeeper* within the support interaction, so opportunities to reflect upon the data are limited in time and coverage:

'[If conflict occurs,] I would need to go away and seek some advice on what can happen next, but it could be useful for the family, to spend that period of time, perhaps looking at all the information and identifying what it is that they feel they're being judged on.'—[Worker, CQ13]

Timely access to data could be empowering, as families could track their own progress, enabling them to make plans outside of the support relationship, reducing dependency upon support, in line with the ultimate goals of the programme:

'If we were working with a family about school attendance, could we then link that in to [the families'] app so parents [would be] aware of what their attendance looks like at this point in time and they […][could] monitor it themselves and take accountability.' [Worker, SQ49]

As well as having ongoing access to data, families need human support to understand that data [SQ49, CQ11]. All participants agreed that *'Families should be able to talk to someone about their data'* [S7]. Explanations are needed [CQ11] with language and vocabulary adjusted to individual literacy [SQ46] or age [SQ47]:

> *'No matter which [presentation of data is offered], you'd have verbal context for it as well, wouldn't you? You wouldn't just go, "There's your app" or "There's your piece of paper" and leave them. You'd just talk it through with them anyway.'*-[Worker, SQ49]

Key to meaningful involvement is the ability to start a conversation. Groups imagined families being able to send a message [SQ51] or record audio to raise an issue for discussion, letting their disagreement be known and empowering them to be part of a dialogue about what is recorded [SQ60].

4.4.3 Theme 2: Giving a Voice to the Family

The second theme systems and processes currently rely excessively upon the 'facts' within the data record, and they need to be updated to give the family an empowered role within their civic information ecosystem. The purposes of an EH intervention are to obtain more information for a better understanding of the family's situation and to make evidence-based plans and decisions to improve the situation. Seeking objective truth is clearly central. Impressions of that truth can be formed either by reading the data or by talking to the family. There are benefits and dangers of relying solely on either source. Families should become agents in the data ecosystem, and this involvement should lead to both greater empowerment and better evidence-based decisions.

People not Records

It was evident, consistent with literature (Gitelman (2013)) and the pilot study (Bowyer *et al.*, 2018), that data can never represent absolute truth - it is often biased or incomplete, and this can mislead [SQ59, FQ11A]. For example, a lack of mental health information could make an individual look like a poor parent [SQ12]. Families may be less willing to 'open up' if they feel they may be judged unfairly [SQ14]. Therefore, developing a strong relationship between worker and all family members is key to understanding the full picture [FQ1]; to ensure fairness [SQ77], data must be current and complete [SQ13], but this state can only

be achieved with the family's cooperation. Looking at data will never provide support workers with a complete understanding. Yet, workers often *'tend to just trust that everything that has been put down is right'* [CQ1], allowing the data perspective to dominate. Such assumptions should be avoided [SQ10]; processes must recognise maintaining human face-to-face dialogue as a priority. Data should only provide supplementary insight: *'You should never make a judgement on data… that data could be wrong. It takes individuality, working with that person as well, doesn't it?'* [SQ11]. All participants presented with the sentence *'Public sector officials can make good decisions just by looking at a family's data'* [S18] disagreed with it.

In spite of the warnings above, the data record is undeniably useful; over 80 comments from workers contend the current practice of reviewing a family's data prior to meeting in person is beneficial, because it provides useful background that will help them identify support needs. For example:

> *'I had a family where trying to unpick what had happened, over ten years, to the child, was really difficult. So, I went away, got the information and came back and if you have* […] *that picture of how the family works* [when you meet them], [that helps]*.'*—[SQ1]

Additional benefits identified included safeguarding workers [SQ3] or giving them an ability to *'check the family's claims'* so that they might constructively challenge individuals [SQ4]. Supported families echoed the value of workers reviewing data [FQ1A], and saw benefits included *'not having to repeat your story'* [SQ5].

The compromise that participants identified over the use of data is that workers should avoid making judgements based solely on data. While sometimes providing essential background to a worker [FQ11B, SQ62], historical data in particular often leads to inadvertent prejudice, especially where labels are used [SQ9]. No participant disagreed with the sentence *'Labels like "domestic abuse" are damaging to families and hard to shake off'* [S15], and workers recounted experiences of being uncertain how to judge historical issues:

> *'[There's] this perception of something sticking with you even after you've potentially reformed. […] That's something that happened a long time ago and that*

judgement is still there but [you'd be wondering] "Okay, is it [still true]?"'—[Worker, SQ61]

Many participants concluded that only *'relevant'* information should be available, to those who *'need to know'*, but the wide range of opinions expressed suggest that this is a highly subjective judgement that would be difficult to determine. A cut-off period before which workers should have no right to look was suggested [Parent, CQ15], but the sentence *'Officials should be able to see historical records about families'* [S17] was contentious. Some workers feared any restriction in access might mean they miss important background on an individual's past, such as sexual abuse or mental health issues [Worker, SQ76]. The solution to this dilemma is unclear, but transparency about what is in the data would seem to be a critical ingredient [4.4.4].

Checking Data Together

The idea of families and support workers reviewing data *together* arose from many participants in workshops A and B, and this motivated an exploration of the concept of shared data interaction in more depth through the storyboarding exercise in workshop C [4.3.4]. Families perceived value in having not just data representations as above, but also a data interface present within their care meeting, so they that they could see actual data and have it explained to them. One practice embodying the concept of transparency that is emerging in some care services is the use of '2-in-1' devices (laptop/tablet hybrids) within the care engagement. Workers can then visibly record data in front of families and then ask them to *'approve'* the accuracy on screen [FQ12, SQ67]. Participants believed this would help to build trust between the support workers and families; if a family begins to feel powerless, they may disengage [SQ35], but even minor involvement, such as this emergent practice of signing off approval of data records [FQ12] or an imagined process of checking & correcting data records together (see next section) could make families feel more empowered which could make the support relationship more productive.

Family participants imagined going beyond just seeing and getting verbal explanations of their data to being able to review their data and be asked for their approval of accuracy [FQ3]. Maintaining accurate data is important because that data is used to decide care plans and support strategies. Families are thought to be better placed than anyone else to identify inaccuracies or gaps in their civic data. Participants believe family corrections would

increase data accuracy. This does not mean free editing of records (as, for example, fears and/or self-interest could lead to families misrepresenting themselves in data (Bowyer *et al.*, 2018)). Instead, taking a role in reviewing, annotating, explaining, or requesting changes, through direct data-centred collaboration between involving workers and family members is desirable:

> *"*[There would be an] *individual view where each person within the family would have their own section* […] *you could sit with them* […] *and go through the data that we have got which would enable them to change anything that they want taken out."*—[Worker, SQ66]

Shared data interaction carries the potential to bring benefits in accountability, accuracy, simplicity [SQ25, SQ26] and consent.

Changing Lives Means Changing Data and Changing Consent

One reason for reviewing historical data and for requiring dialogue with the family to gain an up-to-date picture, is that the truth changes over time. People are not static, and families' lives are always changing; given marriages, divorce, birth, death, house moves and other changes, data can become out-of-date simply through inaction. Given this, asking consent once at the start was considered insufficient [S3]. Data is inherently static – it does not change, but people do [SQ61, SQ63]. This was the basis for participants' desired practice that not only the content of the data, but the family's consent over what happens to that data that both need to be reviewed regularly [CQ16]. A process of regular reviews around data use could prevent unwelcome surprises about how family data is handled [CQ2, CQ17] which could damage trust and hinder co-operation. Participants imagined data systems issuing notifications or update feeds for families and support workers showing significant events or data updates [SQ64]. Support workers currently get notified of police incidents, safeguarding concerns and hospital admissions, but alerts of data changes across the care ecosystem could provide useful triggers for reviews or discussions:

> Worker A: *'We would get a report through to say…'"*
> Worker B: *'They've recorded something.'*
> Worker A: *'Yes. Then I suppose we would follow it up […] face to face.'*—[SQ65]

Regardless of the particular mechanism, it was ultimately felt that both data systems and support processes need to do a better job of supporting change.

Individual Agency & Family-sourced Data

The idea of families reviewing data has significance not just for how it can help within the support interaction, but because it can give families an independent role in their data ecosystem. Both families and support workers imagined the family being able to interact with their civic data on their own, something that is currently not possible. This is a vital step for empowerment: if something goes wrong, families must be able to discover this and must be able to do something about it. Without a cycle of feedback involving individuals as stakeholders having the ability to review and correct data, data will quickly become inaccurate (Pollock, 2011). Thinking about data interaction at home unlocked additional thinking, such as families helping to fill gaps in data [SQ57] or contribute new data that may not otherwise be recorded [SQ58]. Giving families the ability to contribute new data would empower them to *'tell their own story'* [FQ10]. Many participants saw this as-yet-unavailable capability as expected common sense:

> *'I just generally want to see [what is stored about me] just to know what people are saying and then obviously if it's wrong, I can correct them on it.'* [Parent, CQ14]

Rather than solely relying on dialogue, families could provide new data more directly, e.g. through a 'family network app', which could also increase their sense of data ownership:

> *'It would [ask them] who [professionals the family is involved with] they could name outside of their family to create a network. […] But it would collect more than that, […] it would allow the family to be accountable for their data collection and making sure that it's accurate […] because we often go away and record it all on [our existing database] and it's our story rather than their story of how the events occurred.'* [Worker, SQ36]

With new ways for self-expression, families could add context for support workers [FQ9, SQ55], unlocking new support topics [SQ56]. The overriding sense from both groups was that families having the ability to annotate or explain their data would allow them to hold authorities to account, and empower them to tell their story and *'show the real me'*, as illustrated in [FQ10, shown in Table 4.3 above].

Granular Access Controls

Participants identified that it is important to consider that different individuals within the family would have different roles, access and summaries, in order to respect individual privacy [SQ52, SQ48]. Psychological harm could be caused through information leakage, for example an adopted child finding out their true parentage (Bowyer *et al.* (2018)). To avoid this, data should be managed carefully with consent being less binary and more fine-grained access controls being offered:

> Worker A: *'When a child turns 16, when they go to the doctors, is that confidential between me and my GP or can my parents see that?'*
> Worker B: *'I think it's confidential.'*
> Worker A: *'Exactly. So in this interface, I [would be] able to see that – [as the] 16 year old - you as my support worker could also, but not my mother.'*—[SQ53]

Once such capabilities are established, this could enable much more careful and deliberate forms of data-sharing which could support the creation of a personal data ecosystem [2.3.4] beyond, but centred upon, the individual family member, all the while remaining under that individual family member's control:

> *'[I'm imagining an] online database of personal family info accessible [only] by people, practitioners that have permission [...] I would say that it's only who you want [to give access to, that can see it]. You would have your private code which you could hand out, like the doctors give you appointments.'*—[Parent, FQ8]

Theme 2 shows see that giving families a role in the creation and stewardship of their data selves may unlock new capabilities and a sense of empowerment for families.

4.4.4 Theme 3: Earning Trust through Transparency

The third theme looks at these imagined new data access capabilities and empowered role for data subjects in the wider sociotechnical context of how they could affect the support relationship. The topic of trust arose directly or indirectly in almost all participant conversations, and the findings show that transparent and open data handling and decision-making processes are key to support workers to earn the trust of supported families. Currently, families are mostly unaware of what data is held about them and what discussions about them are being had and have no choice but to trust both the support workers, and all the parties and technologies involved in the surrounding care ecosystem, which is very hard to do when they have little to no visibility of it. Without visibility, any

error or surprise can be very damaging to this fragile trust and can harm the relationship, and conversely, increase transparency and explanation can avoid surprises and increase trust, improving the relationship.

Transparent, Respectful Data Handling

The findings in Themes 1 and 2 above clearly suggest that in seeking the best possible understanding, families must be engaged in a fact-centric way, which requires trust in the support worker (to interpret and record data fairly and accurately) and in the system (to keep data safe and prevent misuse). A good relationship with the support worker is critical [FQ1] to the family's care. Workers recognise the importance of being transparent with families:

> *'I think that [families] have got a right to know what is held about them and what is said about them.'*—[Worker, SQ50]

Even for data that would itself would be considered uncontroversial, a lack of awareness to that data or a lack of transparency on how data is informing judgements can cause great worry to families:

> *'Some people that I've worked with, I think as soon as they know you're holding information about them they get really tight and [say], "What are you holding about me?" […] They don't like people knowing what's going on in their lives.'*—[Worker, SQ70]

The current approach, which relies on the support workers mentioning data that they consider relevant, can reassure families when they are kept thoroughly and regularly informed, but can result in expectations being broken by accidental sharing of information if its sensitivity is overlooked:

> *'That tends to be the biggest problem with this, these little bits of information that nobody ever thinks are relevant to bring up in everyday conversation and they're coming out.'*—[Parent, CQ3]

Data must be handled respectfully, with attention to family and individual privacy. A lack of transparency and trust can lead to an atmosphere of suspicion [SQ17] where families have *'a totally overwhelming feeling of people checking up on them'* [SQ71] and apply extreme scrutiny to what they are told: *'You can get families who [no longer] believe what's being said*

about them.' [Worker, SQ73]. Fearful of consequences [SQ72], families may withhold information:

> *'Well my thing would be who is [my data] going to be shared with? Which authorities? What is going to be shared? […] If I ask for help because my son has got massive behavioural issues and I've been trying for years to get help with him and […] if I go to social services, are they going to come in and think I can't cope because I'm on my own with five kids? Are they going to take all the kids away? That's my thing. So I'm terrified of Social Services, I really am.'* [Parent, FQ14]

Respectful data handling also includes using tact and discretion when referencing data, and a common current practice is the use of a *strength-based approach* [multiple workers in workshop B] when presenting or referencing data that could be perceived as particularly negative or judgemental; looking for the opportunities for growth rather than seeking to criticise.

An open and respectful approach is rooted not just in decency but in practicality as a co-operative family is easier to support: *'Because if someone is feeling judged or stressed or angry or whatever, then they can stop the conversation'* [Parent, CQ5]. Being transparent with data can also help with accountability and accuracy, which can detect and prevent mistakes earlier:

> *'There was a time where I was at the doctors' and they asked how many units of alcohol I drank, and I said, probably about three bottles a week, at the time, not any more but later on [the support worker] pulled me up on it and they had it down [in the data record] as three bottles a day. That could have caused an issue was anyone ever to ask.'*—[Parent, CQ7]

In current practice, data handling *is* generally respectful - data mishandling and unexpected uses of data are currently mostly avoided; but transparency is low, making the perception of respectful handling quite fragile and entirely based upon trust rather than direct experience.

Always Seek and Demonstrate Greater Understanding

In order to earn, build and maintain trust, support workers must always be seeking to form a more complete and up-to-date picture of the family, in line with the finding in 4.4.3 that individuals are more than what is stored in their records. This requires human interaction

to uncover. Demonstrating a deep understanding of the family, and that a family's lived reality has greater priority than what a database says can be a critical to trust-building:

> *'You don't want to reduce them to this number in a database. You want to understand their actual experiences and support them in getting better.'*–[Worker, SQ74]

It is important that families understand workers' good intentions when accessing data about them [FQ15]. However, if workers had to show all available data to families this could make it challenging to maintain good relations, *'because literally [the data we have] is like everything, isn't it? So I don't know how I would feel…'* [Worker, SQ21]. In addition to avoiding breaches of expectations (see Theme 2 above), a transparent approach ensures that the privacy of families is respected, because data is not used in decisions without the chance for explanation:

> Parent: *'I don't want everybody knowing how rubbish I am with money.'*
> Child: *'That's my life.'*—[FQ2]

Participants also indicated that families' desire for transparency (as mentioned in the previous section) does not just imply reporting data usage, they need dialogue and human engagement to give them reassurance; trust and reassurance can is best achieved through a conversation [FQ1], not a data interface. Support processes need to change to better recognise the role of dialogue, rather than just consultation of a database, as the best way to achieve a rich and nuanced understanding.

Pro-actively Challenge Data-centric Norms

Exploring this need for reassuring dialogue in more depth, it is clear that to avoid damaging negative spirals of emotion, deliberate openness is needed from support workers (and the entire care system) [SQ18] as to what information is held, and how it will be used and shared, in order to alleviate fears of data being used *'against'* families that can arise without that transparency—giving them instead confidence that their interests are being protected, thus putting them at ease [SQ20]. Data handling processes appear to be only explained once in very loose terms during initial engagement, for the purposes of collecting informed consent, and are rarely revisited. Workers could easily imagine explaining data practices in greater detail than they currently do [SQ41] and clearly there is a need for proactive action

by workers to counter the inherent knowledge imbalance of data being collected into systems that they are gatekeepers for.

Workers however lack control over the quality, coverage and timeliness of the family data and see this as a systemic issue they could not adequately address. From my experience with EH teams through the SILVER project [see Section II Introduction] it became clear that while support workers can see more data than most, they have far from the complete picture; in fact, there is no one organisation or individual with visibility of the entire family-information ecosystem, suggesting that greater openness with data would benefit not just the family, but other civic actors involved in the family's lives and in their care. Some participants suggested that openness about data handling needs to accompany data access, so for example if browsing information together [4.4.2; 4.4.3], it would be important to explain where the information has come from and why the support worker has it, rather than just reporting its content:

> Parent: *'[If the worker knew sensitive medical information] the family would be really annoyed, they would just want you [the worker] to go.'*
> Worker: *'I'm the same, me. I'd be like* "I don't know how you got all this?". *That would be my first reaction but then if we [were to] discuss it and browse the information with the family [that would work better].'*—[CQ6]

As mentioned in 4.4.3, there is a need to replace the current practice of treating consent as a one-off formality at the start of the support process with something better. This has previously been conceptualised as a need for *dynamic consent* (Bowyer *et al.* (2018); Kaye *et al.* (2015); Williams *et al.* (2015)]. A common heuristic expressed by families here and in the earlier study is that data should only be seen by those that *'need to know'*, but this is very hard to achieve: first, because without transparency of data handling, a family cannot verify whether this is happening, so has to rely only on feelings and supposition to inform their trust. Second, the need for fair judgement over who should access families data is objectively important given that some support workers expressed a belief that their right to access families' data should overrule families' consent:

> Worker A: *'I think to enable us to work with families, we need to have as much information to give them the best possible service. So, I think we should be able to [access their information] regardless of what families say.'*
> Researcher: *'Regardless of what they say?'*

Worker A: *'I do, yes.'*
Researcher: *'Does everyone feel the same way then, that they don't get a say?'*
Worker B: *'Yes, because you need as much information as what you can.'*—[SQ22]

This suggests that to ensure the 'need to know' is determined fairly and accountably, independent oversight might be needed; other situations that would benefit from this include deciding what parts of a medical history are *'relevant'* [SQ23], arbitrating situations where legal duties may require the breaking of consent [SQ24], and being able to identify and address situations where recorded information may not tell the full story [CQ8].

These findings suggest that not just transparency but a progressive attitude to data practice, actively challenging current data centric norms, would enhance trust around data handling access and decision making as requirements and lead to a healthier support relationship. This could even include thinking about new ways of using data, for example at a collective community level [SQ78], to promote an open data-sharing culture.

4.5 Discussion

The findings have provided a deeper understanding of families' experience of data RQ1 [3.3.1] and the role of data within EH relationships RQ2 [3.3.2]. This section discusses three implications:

- the value of involving people with their data [4.5.1],
- the need for human interaction to make data access effective [4.5.2], and
- the implications of a shared data interaction approach in terms of shifting the LDM closer to the supported family [4.5.3].

4.5.1 The Value of Involving People with Their Data

Data about supported individuals is integral to current care practice, improving decision making by providing a more complete picture of a family's life [4.1.2]. This prevalent mindset reduces families' autonomy. Just as in the commercial sector [2.1.2], families' civic data is considered a resource to be utilised. The implicit assumption is that data is a complete and objective source of truth [4.1.2], yet participants agree it can never be [S14, S18'].

Families' lack of awareness of held data and its use leads to false expectations and surprises, and in the worst cases, fear and distrust which can harm the care relationship. Family civic data currently serves a *proxy for family involvement in decision making*. Families are cut out of the loop. They should be able to take a role in relation to their data (Bowyer *et al.*, 2018). A lack of such involvement removes any possibility for data checking, increasing the chance that data can contain inaccuracies or errors of judgement due to out-of-date or missing data, which can cause harm (Bowyer *et al.*, 2018; Crossley, 2022).

Trust

Participants' responses confirm prior findings (Dijck, 2014) that trust in the independence and integrity of data holders is critical to an effective support relationship. Yet trust in EH is currently fragile, resting upon feelings and impressions. The findings suggest that a support worker can build trust by continually striving to develop and show a deep understanding of the family as individuals, beyond 'what the computer says'. Trust—in the system and the worker—can be earned by treating them as people, not *'objects to be administered'* [Cornford, Baines and Wilson (2013); S4], leading to more effective support.

Shared data interaction practices (such as checking data together, visible data recording, family sign-off, or contribution of individual perspectives as data) could give a family direct evidence that they are being listened to and that their viewpoint is important even when it contradicts the digital record. This would be powerful for trust-building. Transparency of processing—something that is currently near impossible—could empower families by giving them confidence through their ability to hold providers accountable by verifying that (as per the need established in Bowyer *et al.*, 2018) data is fair and accurate. It is evident from the findings that an EH system built upon strong trust would require direct involvement of the individual(s) being cared for. Shared data interaction as conceived by the participants suggests mechanisms for such involvement.

Shared data interaction would also align with the desire for evidence-based decisions [4.1.2]. An earlier WHO study found that a similar approach lead to more effective collaboration and better decisions (Johnson, Kim and Church, 2010). Inefficiencies that could be avoided include spending time correcting misunderstandings, or 'damage control' following misjudgements. Participants described emergent practices of using data with

families to track progress; these are apparently already effective, allowing families to see their own progress. If such data were available outside of the support engagement, this could empower families still further to be self sufficient.

A health innovation project in South Africa echoes these findings on the importance of trust, agency and involvement of the individual:

> *'The user must feel or experience trust, […] feel that they can control and increase their own access to a system. Their uptake and use are essential for such a [digital ecosystem] to work or be regarded as a sustainable solution.'*—(Herselman *et al.*, 2016)

Consent

Viewing data as a shared resource to be curated together could provide an effective alternative to EH's currently ineffective consent model [4.3.2]. Currently consent collected as part of the EHA serves as a a *point of severance* (Luger and Rodden, 2013) that hands authorities *carte blanche* powers to collect and use families' data. Ongoing access and direct use of data by families would offer practical *dynamic consent* (Kaye *et al.*, 2015; Williams *et al.*, 2015). If families were regularly 'talked through' their data, consent would become more reliable and less bureaucratic. Consent could shift from being seen as a formal permission to be certified and tracked, relying instead on the natural human instinct to speak up when you are shown something that seems amiss. Families will be happier with the use of their data if they can see it, notice issues and speak up when they feel something is amiss.

If implemented in a robust manner, this could simplify consent processes and, by sharing responsibility for data stewardship, limit support workers' liability. Conversations centred upon data would allow mistakes to be spotted sooner, easing workers' fears of *'missing something important'*. Shifting the focus of the support relationship towards *discussions around data* might help alleviate the inadequacies (Cornford, Baines and Wilson, 2013) of the EH record to accurately represent families.

Giving the user a role in understanding and influencing the life of their own data is identified as a key ingredient of moving towards a more progressive model of digital citizenship:

> *'If, instead of disempowering users in the name of simplicity and ease of use, we acted to empower them and ourselves through increased literacy in the technologies employed, and constructed systems where data about behaviour can be more easily quantified and controlled by the user, then we would have the tools at our disposal for a more equitable negotiation with commercial and governmental forms of power.'*—(Bridle, 2016)

Families as Agents of Self-Care

Shared data interaction could make supported families stakeholders in their own case. The care worker need take less of a position of authority, instead becoming an ally to the family member(s), now empowered as an agent in their own self-care. Individuals would be more able to act and improve their situation than previously [4.4.3], approaching the HDI concept of *agency* (Mortier *et al.*, 2014). Such a shift could serve as an antidote to data-centrism in the system and society at large [2.1.2]. With control over and input into their *'data self'*—the representation of them that is seen by the state—families would have greater trust that their interests are being served. Through data, they could take part in informed decisions to could improve their own lives.

4.5.2 Effective Data Access Requires Human Interaction

The Need for Support

The findings shows that current data inequalities will not be solved simply by opening up databases to families and giving them access. They must be able to meaningfully comprehend the data and effect change based on what they learn. This involves the translation of raw data into meaningful information [2.1.1]—through summaries, visualisations and explanations.

It is not clear who would have the access, skills and mandate to create such information representations. Participants' designs and desires echo prior claims that information available to must be *legible* [2.3.2; Mortier *et al.* (2014)] and that access must be *effective* [2.1.4; Gurstein (2011)]. This includes providing suitable opportunities for access—for example via personal data interfaces not just within meetings—as well as addressing technology, literacy, mental or physical handicaps. Participant ideas around audio interfaces illustrate the extra steps that would be needed to provide effective access for all. To support varied needs, information access would need to be supported by a human

relationship—one where someone can both explain the data as well as answer questions about it [4.4.2].

It is the human-to-human interaction that can make data access meaningful. Data use necessitates an ongoing conversation. Systems need a human face or point of contact that individuals may put their trust in and to whom they can address questions. Access only to raw data would be inadequate and limiting (Cornford, Baines and Wilson, 2013).

Working Together

Explorations of human interactions around data within the support relationship suggest that shared data interaction could lead to more effective discussions. The use of printouts, tablets or 2-in-1 devices to show data rather than verbally report it, can provide a focal point, optimising the discussion. The data representations would function as a *boundary object* (Star, 1989, 2010; Bowker *et al.*, 2015), just as the data cards did (Bowyer *et al.*, 2018). The families understand the data because it is *about* them, and the support workers are familiar with the systems it came from. This could encourage families to 'open up'. Many participants talked about how looking at data could serve as a conversation starter—echoing this study's use of stimuli [3.4.2]. The ongoing use of such data representations as a metric to observe change from from meeting to meeting could bring a feeling of reward and accomplishment to the family and contribute to their future success.

Shared data interaction would also enable support workers to be less adversarial. They could position themselves as equals (*'Let's make sure this data is right.'*) and avoid appearing to side with the data (*'Our records say that you have…'*). As we see in this study, data representations serve as *'things to think with'* [Bowyer *et al.* (2018); 3.4.2]. Shared data interaction—played out in abstract through the use of storyboarding cards in Workshop C—helped participants to navigate scenarios and quickly imagine possible actions together.

Agency & Negotiability

Yellow-bordered cards (for families) and blue-bordered cards (for staff) triggered both parties to take ownership of their piece of the puzzle, placing their 'own' actions without any direction having been given as to who should place which cards. Parents and staff had each taken a role in the scenario and felt ownership over the choice of options available to

them. Similarly, green-bordered cards (actions involving both parties) usually resulted in both parties discussing and agreeing a view before the card was placed, showing that the way data is presented is key to how people respond to it. This gives some insight into how the dynamics of shared data interaction might work if implemented in practice. Feeling able to perform actions such as commenting or correcting data would provide some *agency* (Mortier *et al.*, 2014) for family members; The availability of capabilities over data, including the ability to raise a question or start a conversation would satisfy the second HDI requirement of *negotiability*. Conversely, an inability to act upon or influence shown data and its use would indicate a lack of negotiability, reinforcing the idea that simply viewing data is insufficient.

Efforts to deliver effective HDI capabilities in future should therefore focus upon the role of the human in the information system, as a data interface is limited by its operational context as to its ability to truly offer meaningful actions to users.

As our participants all strongly agreed, supported families *'should be treated like people, not database records.'* [S4; 4.4.3]. Excessive focus upon the record can inadvertently become problematic in EH practice focusing upon child welfare:

> *'Children* [can be seen as] *the objects of a variety of concerns which need to be acted upon rather than agents of their own lives'*—(European Commission, 2014).

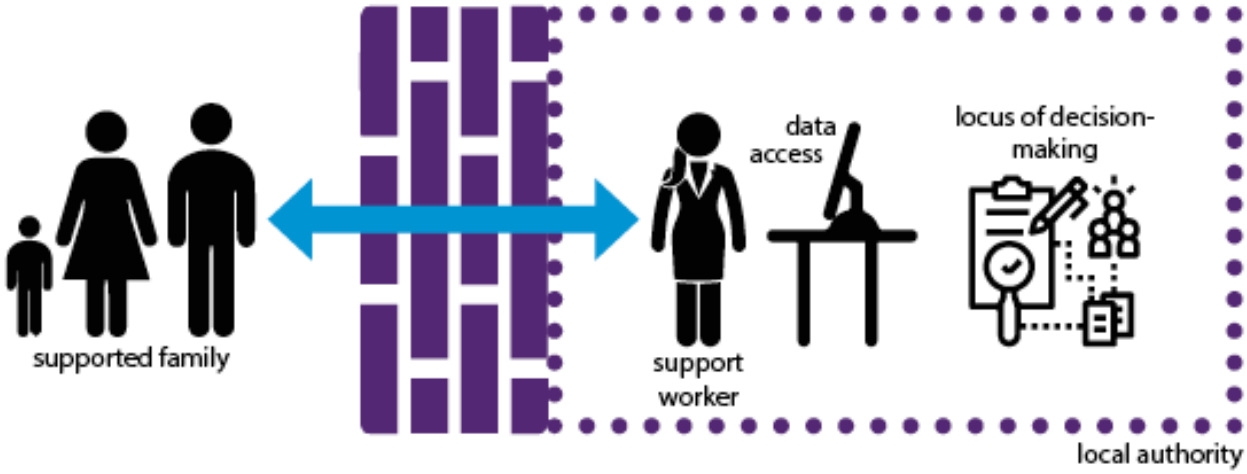
Analysis of the Child Index, an early warning child welfare system in the Netherlands, drew a similar conclusion on the importance of maintaining a compassionate human aspect in family-state relations:

> *"Taking into account that* [care] *professionals' first love is the best interest of and care for a child,* [policymakers should] *provide room for the 'love' between future technologies and their social actors to flourish."*—Lecluijze *et al.* (2015)

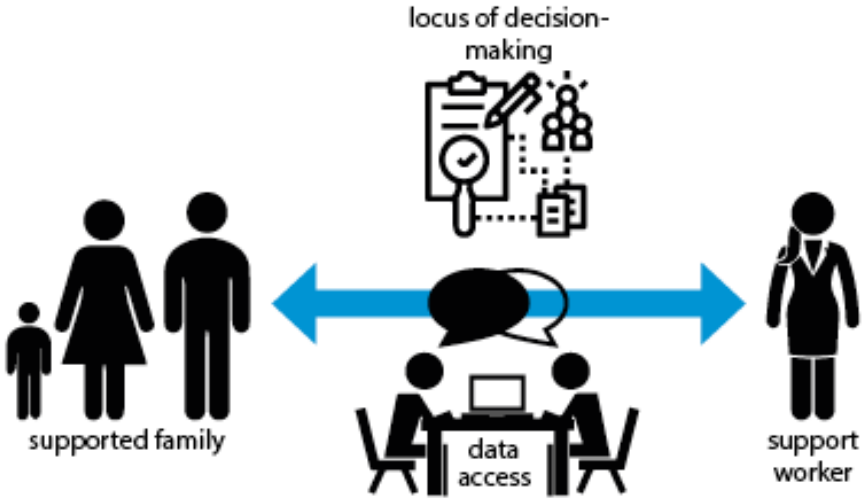
4.5.3 Shifting the Locus of Decision Making Through Shared Data Interaction

In pursuit of RQ2 [3.3.2], the workshops explored the role of data within the EH support relationship. Workshop C brought parents and workers together to explore the mechanics of shared data interaction at an interpersonal, sociotechnical level [2.2.5; 2.3.3] to explore the mutual benefits that might be offered by *shared data interaction*. The participants worked together to construct possible narratives expressed through human-human and

human-data interactions. In doing so, they modelled how shared data interaction could rebalancing power over data. I conceptualise this as **shifting the Locus of Decision Making (LDM)**. LDM is a concept distinct from *locus of control* (Spector, 1982) which normally refers to personal willpower, and *locus of power*, which refers to the concentration of power within an organisational hierarchy. The LDM is the the place where decisions are made. It may or may not coincide with existing authority structures. Accumulated understandings of EH practice, and more generally of data-centric service provision, suggest that decisions are typically made, germinated or championed close to where data is accessed. In an effect that has been expected since as early as 1970 (Klatzky, 1970) the increasing use of data in service provision [2.1.2] has concentrated the LDM with data holders, who collect service users' data to serve their own purposes.



Current model: Support workers are gatekeepers of family data



Ideal model: Support workers and family members use data equitably to plan together

Figure 4.2: Current Model of Data Interaction, and Proposed Model of Shared Data Interaction

Removing the Gatekeeper

The current and imagined approaches are represented in Figure 4.2.

In the current model (top), the only access to data by families is through the support worker as gatekeeper, who decides the scope, content and nature of their access. The LDM is 'back at the office', locked away from the family's participation.

In a more equitable model (bottom), support worker and family member are positioned as allies looking at the data together. This changes the nature of the support relationship. Some of the work that was previously done solely by the data holder (such as data maintenance and data-driven decision making) would now take place in the two-party context of the support meeting itself.

The removal of the gatekeeper role redistributes the power to interpret, select and judge data. Families would no longer be prevented from participating in data-based decision making. I theorise that shifting the data access from the domain of the support worker to the shared domain of the support meeting would move the LDM closer to the heart of the support relationship, creating a more balanced relationship and increasing families' agency and power.

Support for shifting the LDM through shared data interaction is seen in the findings: Both families and staff said they would value a shared data interaction approach. Multiple participants independently imagined benefits of reviewing data and consent together [4.4.2; 4.4.3; 4.4.3]. Participants perceive shared data interaction as an improvement, but it has not been tested in practice, so the remainder of this subsection will consider the benefits and implications of such a shift.

Individual/Family Benefits

Shared data interaction could empower families by giving them a role to play as agents in the life of their data, and a new ability to create and curate their own *data self* so that it is as fair, accurate and representative as possible [Bowyer *et al.* (2018); 4.5.1]. They would be further empowered by having access to view metrics by which their progress is judged, as they could take steps to influence any poorer metrics, and then use the improved metrics as

confidence-building evidence of growth—a positive feedback cycle that is hardly possible at present.

Shifting the LDM could enable families to take more responsibility, through an increased ability to reflect and make plans—an important element of harnessing one's personal data for self-improvement [2.2.3; Abiteboul, André and Kaplan (2015)], and through better accountability [SQ75]. Exposure to data is required for accountability (Crabtree and Mortier, 2016). The perceived benefit of individuals directly using data-based interfaces for health and wellbeing are already accepted; 93% of doctors believe apps can improve health outcomes (Kostkova, 2015).

Care Provider Benefits

Benefits to supported individuals can be seen as benefits to the EH practitioner too, given their role to help improve the family's situation. But co-stewardship of family data can also reduce the burden and responsibility upon the authority to look after that data. Responsibility for ensuring completeness, accuracy and fairness is now shared. This could also reduce the likelihood of complaints or litigation, not least because families can shift from an 'us and them' mentality towards a more favourable perspective.

Provided the individual remains engaged, informed and understands the data and processes that exist, the scope for breaches of consent by workers is reduced because at every meeting (and, with personal data interfaces, outside those meetings) supported families are involved in a conversation that directly enables them to voice their approval or concerns for the ways their data is being used.

Practical Challenges

Implementing shared data interaction would be challenging. Costs could be incurred if new equipment such as 2-in-1 devices were needed. New software interfaces would need to be commissioned, developed and purchased. Such interfaces would be technically difficult, given the fragmented care sector infrastructure [4.1.2; Copeland (2015)]. Identity management in this context is already challenging (Wilson *et al.*, 2011), and child involvement requires special care (Tregeagle and Darcy, 2008). Support workers would need additional training both on software and hardware, this is already a current issue in

the UK (Honeyman, Dunn and Mckenna, 2016) and a critical one in Poland (Soja, 2015). Training becomes particularly important in a system where the care workers must take on an educational role as sensemakers of digital records.

Involving individual members of the public as actors within systems previously targeted to staff would likely carry fresh considerations for access control, technical support and public liability insurance. Providing personal data interfaces to the public, and new communication channels for asking questions would carry a large human resource support cost.

The creation of a direct communication channel between supported individuals and support services offers potential savings for the state in terms of reducing the amount of costly (Kriisk and Minas, 2017) 'in the home' contact. However, broad human support for data access might change these savings into a net cost.

Dispute resolution procedures and additional legal and information governance support would be likely to be needed. New challenges might occur, such as individuals with destructive, manipulative or otherwise challenging intent, who might try to mislead workers for person gain.

It would be fair to criticise human-centred state interaction as something that would not be cheap or scalable. With more individual-state interactions, every case could take more worker time in a system that is already overburdened and underfunded (Copeland, 2015; Local Government Association, 2017). The state has adopted a data-centric approach to citizen interaction in part because it cannot manage to provide human relationships with every individual citizen. But now this approach has become ingrained into government approaches to citizen relations:

> '[Data-centric citizen interaction] *is no longer a technological necessity, but it has become a political intention.'*—(Bridle, 2016)

There is a need to reverse this trend, not only in practice but in political ambition, if people's interests are to be best served, and if a welfare state is to be truly *enabling* (Miettinen, 2013). By taking a more innovative approach to digital policy, it is possible that governments could be more effective in helping to involve those citizens that have become disadvantaged by the current system. A more human-centred approach could help to

combat the digital divide (Kalvet, 2005; Steyaert and Gould, 2009). Pragmatic considerations for organisations wishing to explore new more human-centric models for service user data relations are explored further in 10.2.

My proposal for shifting the LDM is theoretical; it does not provide an implementable solution that could be rolled out at scale. Instead, it can serve as a mental model to stimulate discussions about potential change. Its value is in shining a light on the positive and negative impacts upon relationship effectiveness of current data practice in EH. These findings offer imagined practices that could be more efficient, and serve as a challenge to the status quo that should encourage EH providers to question their priorities when it comes to the use of people's civic data. The primary goal of EH is to empower families to help themselves as effectively as possible, and EH's focus on data arguably works against this.

4.6 Summation of Case Study One

Through four participatory co-design workshops with supported families and support workers in North-East England, this study highlighted five problem areas which participants perceive with current personal data practices:

- **A power imbalance** – Families' personal civic data is collected by care organisations who view it as a resource to be utilised, creating a structural power imbalance against families who are barely able to influence data values or practices.

- **A closed and opaque data ecosystem** – Families lack awareness of what data is held about them and how it is used, with support workers (who themselves have limited access) functioning as gatekeepers to what families will be told.

- **Ineffective, meaningless consent** – The current consent model, while legally satisfactorily, is ineffective. It is a one-time initial hurdle after which support workers can do whatever they deem necessary with families' data. Families are never again given any meaningful choices about what happens to their data.

- **No accountability and fragile, limited trust** – Without transparency or an ability to request or demand changes to data or data practices, families have no ability to hold

data holders to account. This makes families' trust in the system hard to earn and fragile to maintain.

- **A lack of agency or true empowerment** – With families having no ability to shape the way they are seen in data, or even just to see how the state sees them, opportunities are missed to empower families for personal growth.

These explorations of shared data interaction and personal data interaction show there is both a need and a desire for a new approach. Deliberate openness with families' data and direct use within face-to-face consultations could address all five of these problems. The removal of the gatekeeper role over families' civic data would give families a role in the stewardship of their own data, shifting power in their favour. Clear visibility of data recording and usage would enable accountability (currently absent), engendering trust. With ongoing family involvement, the consent problem would be largely solved; families could immediately speak up at any point in the light of new developments or new information. Data-informed and support conversations could enable better decisions. With an ability to influence how they are represented and observe changes through data, families would be empowered to make changes in their own lives, acquiring a previously unattainable level of agency.

Data visualisations and summaries could serve as conversation starters and as boundary objects, enabling more effective conversations. A shift from verbal reporting to 'looking at data together' would change the dynamic of the support interaction away from 'us and them' towards an ally-based approach. Data accuracy should improve, given that for a full picture both data and dialogue is needed (Bowyer *et al.*, 2018). Given the opportunity, individuals could contribute explanations, fill gaps or correct mistakes. Decision making would naturally improve through a greater focus on discussions around accurate evidence.

This chapter establishes that designing new service relations that give the family a role could be very powerful, because visible data processing would provide them with direct evidence that they are being listened to and that their perspective is seen to matter more than 'what the computer says'. The ability to ask questions about data treats the family with greater respect. Personal data interfaces, enabling families to act independently in their own time and in contexts outside of the support interaction, would allow individuals to

alleviate concerns quickly and maintain confidence that their data selves remain fair and accurate. At the same time, new opportunities could arise for use of one's own data. The adoption of such measures could facilitate the emergence of a human-centred personal data ecosystem [2.3.4] in a civic context.

Capabilities – or their absence – matter more than the on-screen technicalities of the data interaction. Data interfaces are limited by their operating context as to how much they can offer, but in this chapter data interaction is examined as a sociotechnical process [2.3.3], looking beyond interface interaction to the human relationship between the individual and the state, which allows more holistic and human-centric solutions to be imagined, in line with the objectives of this thesis [2.4].

The sociotechnical perspective allows deeper consideration of the purpose of EH interactions, refocusing attention on the core goals to empower individuals to better themselves. The human perspective on family civic data should be given the highest priority, so that professionals' flexibility is not limited, but also because data cannot adequately represent the complexities of human life (Bowyer *et al.*, 2018). People are more than just data, and you have to talk to them to make sense of their lives and to avoid excluding them. Usage of data should always be supported with dialogue and engagement. It is this need to focus on the human aspect that explains why trust underpinned nearly every single problem imagined by participants. Without an open system that encourages dialogue and discussion, it is very hard not to close doors, create suspicion and harm trust.

Figure 4.1 informs RQ1 [3.3.1], showing perspectives upon family civic data and practices. The thematic analysis [4.4] of workshop transcripts identified positive and negative impacts on the support relationship of current civic data practices within EH, identifying best practices, seen in the subthemes and expressed as 38 specific practices for EH services (Bowyer *et al.*, 2019). Many of these are currently imagined or only just emerging. Participants believe these practices would improve families' engagement and the support they receive. These proposed practices (which inform RQ2 [3.3.2]), as well as the shared data interaction approach, challenge the status quo. They can inform policy, process and system design for all kinds of attempts to reform care services or digital citizenship offerings. Such changes carry significant challenges in cost, training, manpower and emergency planning, as with any systemic practice change in an organisation, but such an

approach may get closer to the heart of the real issue of empowering *'left behind'* (disempowered) families, perhaps moreso than a purely state-centred approach. Human-centred data use may offer a route to a more enabling welfare state. This work serves as a reminder that as we move into the data-driven age it is important that data should stay close to the people it is about, rather than to those that use the data to provide services.

The general principles expressed here could be equally applied to other domains including education, healthcare, democracy and commerce, and this emphasis upon individual capabilities over interface design is a mindset that could be applied to many human-computer interaction and design endeavours.

Chapter 5. Case Study Two: The Human Experience of GDPR

"The Crystal Wind is the storm,
and the storm is data,
and the data is life.
You have been slaves,
denied the storm,
denied the freedom of your data.
That is now ended; the whirlwind is upon you . . . . . .
Whether you like it or not."

—from *'The Long Run: A Tale of the Continuing Time'*
by Daniel Keys Moran
(computer programmer and science fiction writer)

In this chapter, I will describe the second case study of this PhD, in which I took 11 participants[8] through an longitudinal in-depth one-on-one process of three interviews with coaching and support in between, with the total engagement per participant lasting approximately 4 hours over a three-month period. The purpose of the research was gain a deeper understanding of people's attitudes to the kinds of personal data held by companies in people's everyday lives and what they want from that data (in pursuit of RQ1) and specifically to examine the human experience of existing in a data-centric world [2.1]. Each individual has a number of relationships with service providers that involve the use and holding of personal data. In line with RQ2, the goal is to better understand the role of that data in those relationships. In particular, having gained an initial understand of attitudes, hopes and expectations, a further objective was to examine how those expectations might change during the journey of digital life mapping, data request making, receiving and

[8] 11 participants started the study but one dropped out after the first interview due to COVID-19, so only 10 participants conducted GDPR requests. 31 interviews were conducted in all.

examining of data, and scrutiny of responses, collectively forming a holistic understanding of *the human experience of accessing your data with GDPR*.

In section 5.1, I will expand on chapter 2 to explain the context of using GDPR in research as a means to retrieve personal data. After explaining the overall strategy and preparations made to research the GDPR context in 5.2, I will in 5.3 explain the stages of the interview process (including details of how participants were sensitised). In section 5.4, I will explain the model of personal data types developed for this study, and will present quantitative and summary data from the interviews, explaining how participants' GDPR access requests progressed, highlighting participants' shared hopes and goals, and examining in particular how their perceptions of power and trust were affected by the experience. In section 5.5, I will describe the three themes uncovered through thematic analysis:

- that organisations provided participants with insufficient transparency to meet participants' hopes and their legal obligations [5.5.2],
- that people struggle to find meaning and value in their data when they do manage to access it [5.5.3], and
- that providers' data practices (in particular their GDPR request handling) can be harmful to their users' trust, but that greater openness can have an opposite, positive impact [5.5.4].

I will discuss the implications of these findings with reference to prior literature, from the perspective of policymakers [5.6.1], data-holding companies [5.6.2], and individuals [5.6.3]. Finally, in 5.7, I will summarise these insights in terms of how they can advance our understanding of the research questions and their wider significance.

5.1 Context: Accessing Your Personal Data Using Your GDPR Rights

5.1.1 The Current Need for Data Access

As established in 2.1.2 and 2.2.4, people live digital lives, inevitably involving the use of myriad digital services that collect personal data, which is subsequently mined for value and exploited at scale, creating an imbalance of power between data holders and data subjects, and a exclusionary landscape around data use which is difficult for individuals to

navigate. Having acquired data about individuals, this becomes a focus for service providers' decision making and customer relations become less important.

Section 2.1.4 established how unaware many people are of this imbalance around data, that there is a want[9] for effective access to data to restore individual agency. As detailed in section 2.1.3, policymakers have been attempting since the 1970s to introduce legislation to tilt the balance of power back towards individuals, most recently and most notably the European Union's General Data Protection Regulation, which legally endows at least 513 million individuals[10] with new rights to timely data access, explanation, erasure and correction (Information Commissioner's Office, 2018).

Data protection and misuse issues have grown in the public awareness since the Snowden revelations in 2013 (Gellman, 2013), and have become even more important following the Cambridge Analytica scandal in 2018 ('Facebook–Cambridge Analytica Data Scandal', 2014; Chang, 2018), which may have resulted in manipulation of voting outcomes through personal data use, and the COVID-19 pandemic in 2021 (O'Donnell, 2020; Hamon *et al.*, 2021). Since the GDPR's launch in May 2018, it has undoubtedly resulted in new data access offerings; many large consumer companies have developed 'privacy hubs' or improved privacy policies where individuals[10] can learn how their personal data is handled or access data download portals to easily download copies of it ('Privacy - Apple (UK)', no date; 'Privacy & Terms – Google', no date; 'Privacy', no date; 'Facebook - Data Policy', no date). Almost all data controllers and processors have now updated their privacy policies to

[9] In this study and throughout this thesis, my usage of the word 'want' in the context of data capabilities deliberately includes both meanings of the word: the need or desire of the individual, but also that which they **lack** (see opening of Chapter 6).

[10] At the time of writing (summer 2022) the GPDR legally applies in both the European Union and the United Kingdom (which adopted a copy of EU legislation when it left the EU in 2020), which have a total population of 513 million individuals (Eurostat, 2020). GDPR rights are also conferred to any individual who is a customer of businesses with registered offices in EU or UK countries, meaning that these rights are in effect globally available for non-EU, non-UK users of many multi-national digital service providers.

include clear processes for data subjects to request copies of their personal data per their GDPR access rights.

However, the effectiveness of these offerings and processes for service users has not been studied, nor how individuals feel about them in light of this backdrop of public concern. No service providers make data access statistics publicly available, but anecdotal reports from industry insiders suggest GDPR access rights and data download dashboards are not well-known and hardly used (except by activists [9.3; 9.5]). This presents an opportunity to take individuals who have not previously used these capabilities on a journey of discovery. Through this, we can assess the impact of these processes over time and whether–by compelling data holders to create such offerings and respond to access requests–the GDPR succeeds in its goals to 'enhance the data protection rights of individuals' (Council of the European Union, 2015) and to give people 'control over their personal data' (The European Parliament and the Council of the European Union, 2016a).

5.1.2 Current GDPR Research and its Limitations

Since it came into effect in May 2018, the GDPR has opened up new possibilities for research (Comandè and Schneider, 2021); the ability to obtain one's data records from organisations provides the general public with a potential deeper view inside those organisations, much like the UK's Freedom of Information Act has provided a view into governmental and public sector organisations, enabling research and improving accountability (Savage and Hyde, 2014). Such legally-enforced transparency also provides researchers with a window into organisations and their processes that was previously only available based on goodwill. Ausloos and Veale (Ausloos, 2019; Ausloos and Veale, 2020) outline an approach for using the GDPR in research as well as describing the many ethical and methodological considerations that should be made. GDPR research can however be as simple as inviting participants to exercise their rights of access and talking to them about the experience and any changes in their perspective, which is the approach this study uses.

The GDPR process itself has also been examined from many perspectives by researchers:

- to understand data holder's compliance with legislation (Ausloos and Dewitte, 2018; Arfelt, Basin and Debois, 2019);

- to evaluate data portability (Wong and Henderson, 2018) and 'privacy by design' (Waldman, 2020);
- to compare its effectiveness in
  - public/private sector contexts (Quinn, 2021) or
  - in improving explainability (Hamon *et al.*, 2021),
  - fairness (Kasirzadeh and Clifford, 2021),
  - consent (Human and Cech, 2021),
  - transparency (Spagnuelo, Ferreira and Lenzini, 2019) and
  - the reduction of data breach risks (Gonscherowski and Bieker, 2018).

Potential negative impacts have also been considered; the GDPR could be seen as a threat to privacy (Bufalieri *et al.*, 2020) or as an impediment to health research (Clarke *et al.*, 2019).

The GDPR has spurred a broad variety of research, spanning legal, social and technology domains, yet there is scant research into the individual human experience of the GDPR. Alizadeh et al. conducted a study with 13 users of a German loyalty programme and interviewed them before, during and after they made GDPR data requests (Alizadeh *et al.*, 2019), finding better responses and GDPR education were needed. This is a good example of the sort of work that is needed to explore the human perspective on the GDPR journey, though this particular study was limited in breadth (only one service provider was targeted) and in depth (the data returned from companies was discussed largely at a high level of 'were your expectations met?' and potential to use the data for one's own benefits was not examined). The implications of the experience upon the participants' relationship with the provider were also not explored. Impacts of data handling practice upon relationships is an under-researched area in general. Recent work (Bufalieri *et al.*, 2020; Glavic *et al.*, 2021; Zuckerman, 2021) has established that openness and transparency around data handling are key to services establishing individuals' trust; indeed an echo of this was seen in a public sector context in Case Study One (see Chapter 4). In a commercial context, such changes in trust can impact customer satisfaction and business success.

At a more fundamental level, there is a need to understand how people *experience* the GDPR; companies' GDPR processes have been designed to comply with litigation but often with insufficient design thinking (Cormack, 2021). GDPR-handling processes and data

access systems have been motivated by a need to comply rather than by focusing on individual needs or desires [Abowd and Mynatt (2000); McCarthy and Wright (2004); Wright and McCarthy (2008); 3.2]. It is highly likely that many user needs or desires have been overlooked. Such experiential understanding could inform the design of improvements to companies' GDPR mechanisms (be they interface interactions or response-handling procedures), as well as identifying specific needs that might be best met through improvements to policy, including to the GDPR itself.

5.1.3 Human-Data Interaction: Towards a Human-centric Personal Data Ecosystem

Given that data-centric services now span all aspects of our lives, and the amount of personal data about individuals has grown, it has become critical to think about the way people interact with data as a 'whole life' problem. This is one of the reasons this study focuses on the layperson rather than a particular demographic, and 'everyday services' rather than a particular domain. Data has transcended the machine and now encodes facts about our lives, it exists across devices and across providers (Weiser, 1991; Mydex CIC, 2010; Abowd, 2012). This means that personal information management has become a sociotechnical problem [2.3.3], that can no longer be solved as a filing-and-retrieval problem as per traditional PIM approaches [2.2.2], but only when considered as multi-party negotiation over representation, ownership, access and consent. It is important to evaluate the GDPR in this context. Up to now, individuals have not had the means to participate in or initiate such negotiations. On paper, it would seem that GDPR rights do convey this capability, but it is not known whether in practice, service providers' responses to GDPR can actually deliver data subjects the ability to take part in negotiations around data in a fully-informed way. While some research on relationships around data and data as a shared resource is now emerging [2.2.5], the relationship with data-holding service providers has not been examined in this way.

A roadmap for best practice in this space can be found in the emergence of the PDE concept [2.3.4]. Researchers have identified that a human-centric approach to personal data is needed, placing individuals at the centre, as controllers and overseers of their own personal data (Mydex CIC, 2010; Symons *et al.*, 2017). This is an emergent space of much activity and research ('Human Data Interaction Project at the Data to AI Lab, MIT', 2015; BBC R&D,

2017; MyData, 2017; Symons *et al.*, 2017; MyData.org, 2018; 'HDI Network Plus, University of Glasgow', 2018; 'HDI Lab, Heerlen', 2020) which provides a strong framing for us to evaluate the human experience of—and interaction with—the GDPR. Given people's diminished agency and control over their data (Woolgar, 2014; Crabtree and Mortier, 2016), do the GDPR's access rights, as implemented by service providers, provide the effective access (Gurstein, 2011) people need? Does the GDPR help people to achieve legibility, agency and negotiability, the three tenets of Human-Data Interaction [2.3.2; Mortier *et al.* (2014)]?

This case study aims to explore the research gap of 5.1.2 above, from this perspective of greater human-centric need in a multi-party data use context. It scrutinises the experience of using one's GDPR rights in order to discover how well the process meets individuals' needs and expectations. In the process we can uncover problems and identify possible solutions that could address them.

5.2 Overview of Activities for this Context

As mentioned in the introduction to Section II, a core motivation for this thesis was to conduct research into the power imbalance around personal data from the *everyday life* perspective–to explore our relationship with and through the data that we hold, use or live with as we go about our lives, online and in person. This power imbalance seems to touch everyone, and therefore for my second research context I chose not to focus on a particular community or group but to look at these problems at the level of ordinary day-to-day digital life. In 2018, during this PhD, the European Union's GDPR regulations came into force, enabling people to obtain copies of their own data. This provided a means by which to examine any individual's data-driven service provider relationships (including my own), as data access requests could be executed which would return copies of held data that could be examined (in my own case) or used as stimulus for discussion (in participants' cases); the core everyday digital life context could now be researched in a grounded way.

To serve research in this context, I carried out formative research activities [5.2.1] between 2016 and 2019, leading to a major participatory study [5.2.2] which ran from December 2019 to June 2020. After the participatory study concluded, I continued to make my own GDPR requests throughout the period of this PhD running up to submission in late 2022,

both as researcher sensitisation for the thesis, and also in the context of my embedded project work with BBC R&D and with Hestia.ai [see Section IV Introduction], which also contributed to insights presented in Sections IV and V.

5.2.1 Formative Research in Personal Data and Digital Life

In late 2016, during the MRes programme, I conducted qualitative research in order to validate my own prior perspectives on useability[11] and to explore attitudes to everyday technology use. I interviewed three participants about smartphone use, with particular attention to perceived usefulness or barriers. I sought to understand what motivates people as users of technology, an important consideration when looking at disempowerment. Findings from this pilot (Bowyer, 2016) contribute to some of the interpretations expressed through Chapter 5 and Chapter 7.

Then, from November 2017 to February 2018, In order to dig deeper into data attitudes and trial methodological approaches, I conducted 5 two-hour interviews with individuals about their digital lives, looking at how they mentally segment their lives, and the roles and functions of different technologies, and especially of data, across those different parts of their lives. I began to explore participants' perceptions of their relationships with service providers, identifying potential disempowerment or pain points. The interviews were conducted using the Sketching Dialogue (Hwang, 2021) technique, which uses collaborative sketches as a basis for a semi-structured interview. A light analysis of the recordings and data was carried out. This trial investigation helped to inform the main study, both in study design and in interpretation, but findings were not formally documented or published.

To sensitise myself as researcher within this context, it was vital to become well-versed in the GDPR. Since 2018, I have practiced obtaining my own data from companies and organisations in my life. This has entailed making over 70 GDPR requests to a variety of organisations including retailers, device manufacturers, online service providers, local and health authorities, banks and leisure services. Additionally, I have experimented with self-service download dashboards and third-party 'get my data' tools. In some cases, I have engaged providers in communication to try and get better data or ask questions about my data. These activities have provided multiple benefits:

- they have enabled me develop a detailed understanding of what actual stored personal data looks like [Figure 3.7];
- they have given me an awareness of the evolving response to GDPR from data-holding organisations and an understanding of the limits of what is and is not possible with GDPR.
- These investigations also contribute to the activist research strand of Section IV, and one particular self GDPR experience with Spotify is used to illustrate how an individual can influence a provider using GDPR [Principle 9–10.1.9].

5.2.2 Case Study Two: The Human Experience of GDPR

The major study for this context was to guide participants through the process of GDPR and retrieving their own personal data, to enable a conversation that included not only attitudes to personal data and the use of data within service relationships, but discussion of how those attitudes were changed by the experience as it happened and how well expectations and hopes were met by the process. Similarly to the approach [4.2.3] used in designing Case Study One, the general methodological elements described in 3.4 informed the study design. I planned to work with participants on an individual basis to help them obtain and subsequently review their own personal data, which would become the stimulus for discussion [3.4.2]. This would be done in five stages:

- Sensitisation using wall posters about data holding organisations, types of personal data, GDPR rights and possible uses for your retrieved data [Figure 3.4; 3.4.1];
- A life mapping exercise, using the *Sketching Dialogue* (Hwang, 2021) technique, enabling target companies to be selected for GDPR [Figure 5.2; 5.2.1];
- A discussion and guided walkthrough of the target organisations' privacy policies, in particular their stated data collection and usage practices [3.4.2];
- Guidance and support in making and seeing to conclusion a GDPR request from the individual to each target organisation; and
- An individual interview where the participant would be guided through reviewing their data and be asked about their experiences and reactions to the data and the GDPR process [3.4.2].

Through these stages the objectives were to understand how people view the data that exists about them as they go about their everyday life and what they would ideally want from it (in service of RQ1 [3.3.1]), as well as what role data plays in their relationships with companies and other data-holding organisations in their lives, and what they would ideally want from those relationships with respect to data (in service of RQ2 [3.3.2]).

All interviews would be recorded, transcribed and analysed to produced both qualitative and quantitative data for later analysis. The next section goes into more detail into the actual execution of these interviews.

5.3 Methodology for the Case Study Interviews

5.3.1 Interview Timeline and Stages

Between December 2019 and April 2020, 31 qualitative interviews were conducted, with a convenience sample of 11 individuals from a population of researchers and students at (or connected with) Newcastle University, aged 20-40 years and identifying as 5 females and 6 males. Participants were not data experts (only 1 had previously made a GDPR request), but were computer-literate, educated to degree level, and used to reflecting critically on their own behaviours and opinions. Participants were compensated for their time with Amazon vouchers worth £20.

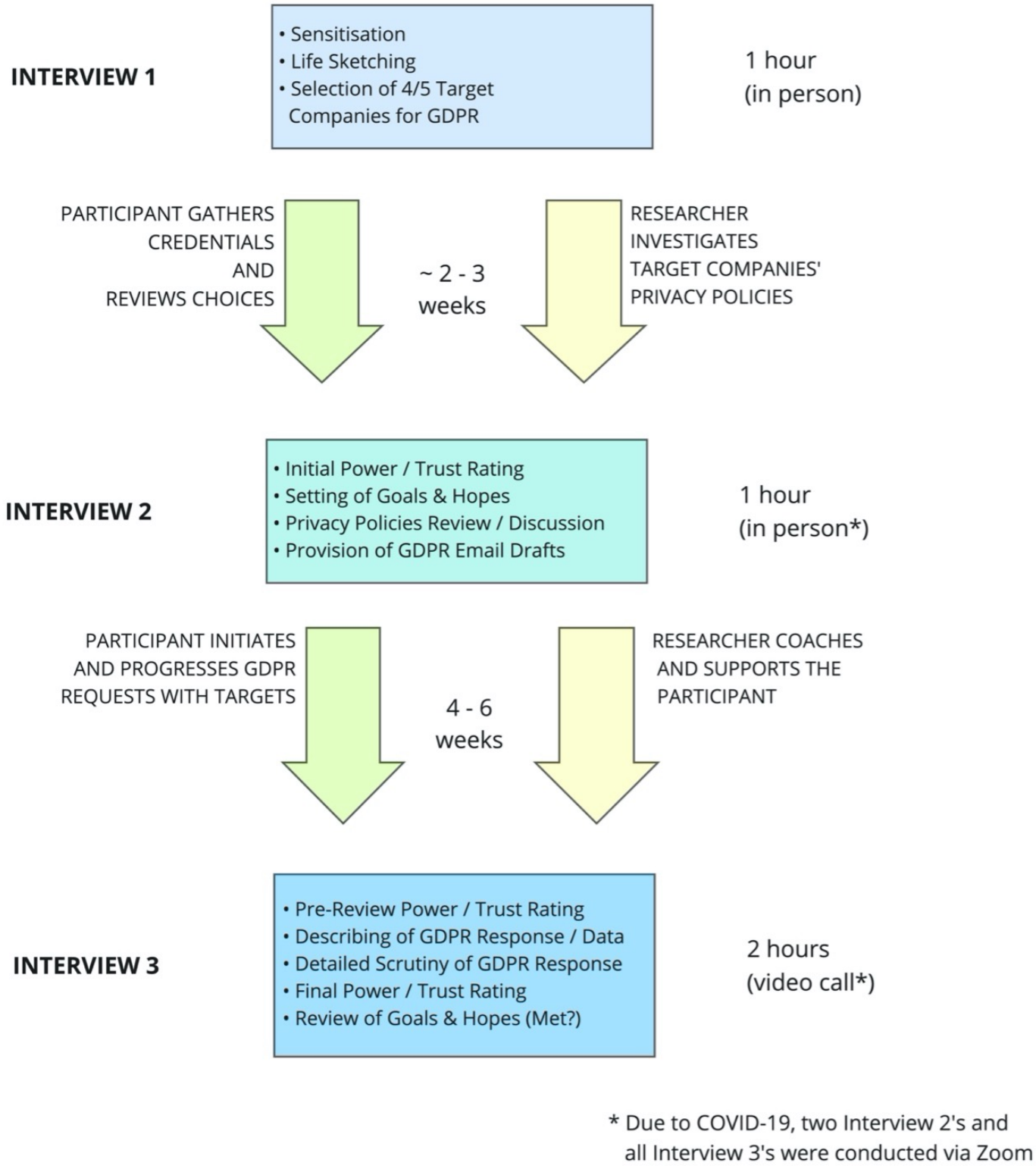


Figure 5.1: A Journey Map of Each Participant's Study Progression

Each participant's journey progressed at its own pace [Figure 5.1] with participants invited to three separate 1-on-1 interviews totalling 4 to 5 hours each, over a period of around three months. The scope and purpose of each interview was as follows:

- **Interview 1: Sensitisation, Life Exploration and Company Selection** [1 hour, in person]. The participant was sensitised to the research context using an interactive tour of a poster display on the topics of GDPR rights, potential data-holding organisations, potential types of data and potential uses for GDPR-obtained data. Baseline data was collected on the participant's hopes and motivations, their current understanding of personal data, data access, data control, and power as it relates to data. Using a sketch interviewing (Hwang, 2021) technique, the participant mapped out their 'data life' (e.g. Figure 5.2), annotating key organisations that they have relationships with, types of data those companies might hold, and feelings about such data use and storage by each holder. The participant then selected 3-5 candidate companies to target with GDPR requests.

- **Interview 2: Privacy Policy Reviewing, Goal Setting and GDPR Request Initiation** [1 hour, in person]. To stimulate reflective thinking and measure impacts, the participant was asked to discuss and score their initial feelings of trust and power with each company. They then viewed key sections of privacy policies on a screen with the researcher, to identify each company's statements on collection and use of personal data. The participant then initiated an email GDPR request for each company, which had been prepared for them using a tried-and-tested template generated by personaldata.io (Wiki.personaldata.io, no date). Interview 2 took place in person, except for P10 & P11 whose interviews took place over Zoom, due to face-to-face meetings not being possible during the COVID-19 pandemic.

- **Interview 3: Detailed GDPR Response Review** [2 hours, online video call] Having allowed sufficient time for GDPR requests to conclude (there is a legal duty to reply within 30 days), a deep dive into the specifics of each GDPR experience took place. The participant's personal data was not collected by the researcher, only described verbally by the participant; screen sharing was used to show excerpts to the researcher where the participant wished to do so. The participant was asked a structured set of questions about the completeness and value of any data returned, as well as to provide new evaluations of trust and power, whether their hopes had been met, and any general feelings about the experience. Answers were recorded in a

screen-shared spreadsheet, which was also used to structure the discussion (for a sample, see the supplemental materials of (Bowyer, Holt, *et al.*, 2022)).

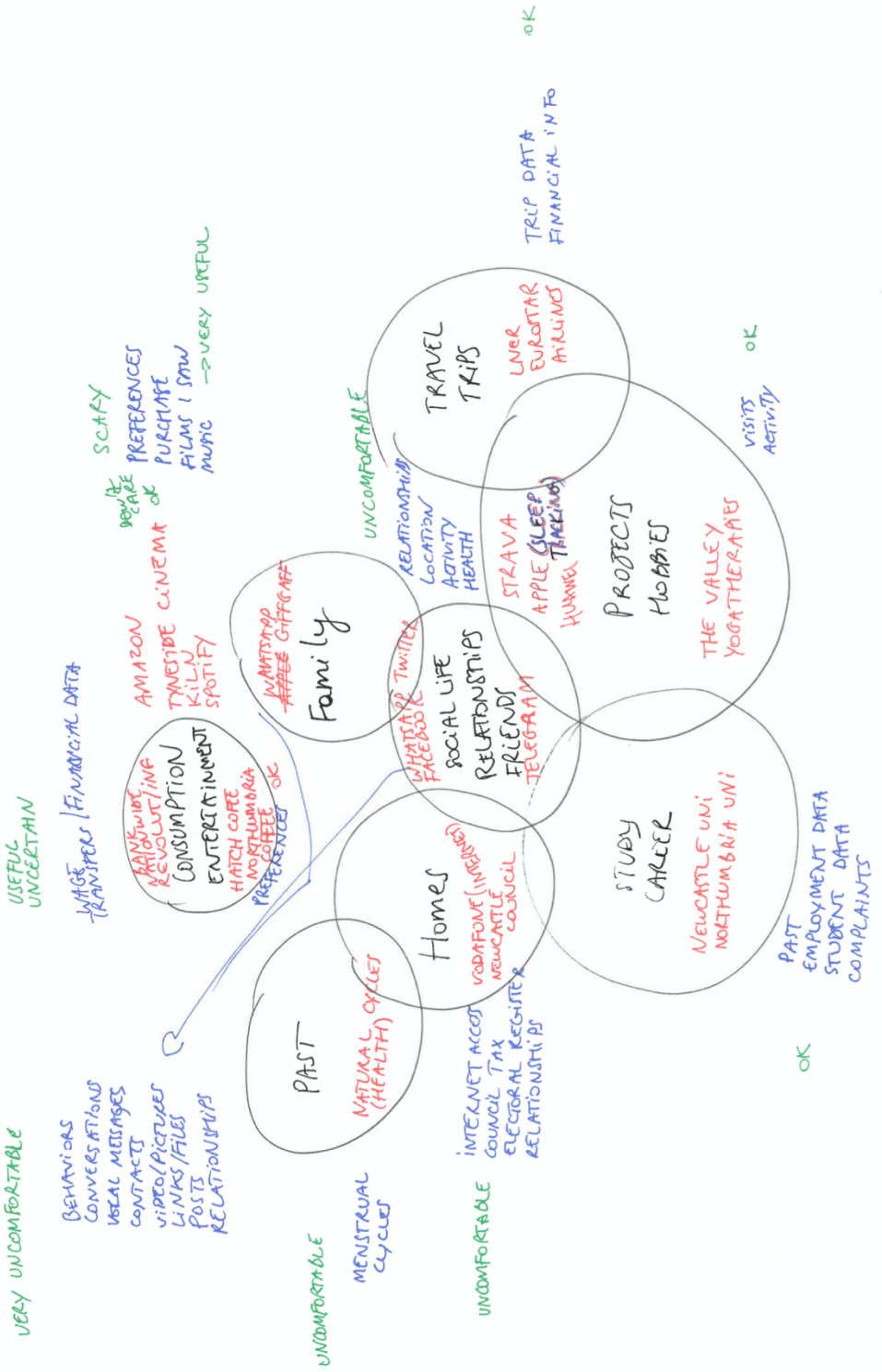


Figure 5.2: An Example Life Sketch from Interview 1, with Data Handling Companies in Red, Data Types in Blue, and Feelings in Green

All 31 interviews were audio and video recorded, then auto-transcribed using Google Recorder/Zoom, producing a 370,000-word corpus. Transcripts were split up and recombined across all participants into six topic areas–digital life, company-specific discussions, general discussions, power, trust, and hopes/goals. These transcript topics were then analysed through reductive coding cycles to produce thematic findings [5.5]. Quantitative data from interview spreadsheets was also summarised and analysed [5.4]. Sketches, recordings, screenshots and field notes were referenced throughout thematic analysis to aid interpretation of the transcripts. The processes described here are further detailed in ARI5.1.

5.3.2 Interview Structure and Questions Asked

Interviews included the following questions, steps and areas of exploration:

Interview 1 (Sensitisation, Life Exploration and Company Selection):

- Researcher explains, using posters, different types of companies, types of data, GDPR rights, and possible data uses, while participant asks questions and shares opinions.
- What is your motivation for taking part?
- What do you understand by, or how would you define, each of the following terms?
    a) personal data;
    b) access to data;
    c) control of data; and
    d) power (as it relates to data).

4. Sketch (in black) and explain the different parts of your life.
5. Label your sketch (in red) to name companies that have data in each area of life.
6. Label your sketch (in blue) with types of data each company might have.
7. Label your sketch (in green) to show your feelings about combinations of company/type of data/area of life.
8. Select 4 or 5 companies you are most interested in making GDPR data access requests to.
9. Review each target company and tell me:

- Who has the power (you or them)? Provide a score for where the power lies with respect to data, where -5 = fully with them, 0 = balanced/equal, and +5 = fully with you.
- How much do you trust the company? Provide a score for how much you trust the company, from 0 ('I wouldn't tell them anything') to 10 ('They can know whatever they like about me')
- What things do you hope to find out?
- What things could you do with the data you get back from this company?

Interview 2 (Privacy Policy Reviewing, Goal Setting and GDPR Request Initiation):

- For each target company:
    a) provide a current score for how much you trust the company.
    b) provide a current score for where the power lies with respect to data.
    c) researcher presents and explains key information from the privacy policy (prepared in between interviews)
    d) discussion about privacy policy: researcher answers questions and collects participant viewpoints.
    e) having reviewed and discussed the privacy policy, revisit and, if desired, update your trust score.
    f) having reviewed and discussed the privacy policy, revisit and, if desired, update your power score.
    g) Researcher provides participant with a generated GDPR request e-mail to send to the target company.

Interview 3 (Detailed GDPR Response Review):

- For each company:
        a) provide a current score for how much you trust the company.
        b) provide a current score for where the power lies with respect to data.
        c) describe (or optionally show via screen-share) returned data to the researcher

d) answer detailed questions (where possible/applicable) around each of the five types of data (volunteered, observed, derived, acquired, and metadata) [Table 5.2]:

i) Is the data complete?

ii) Is it accurate?

iii) Is it understandable?

iv) Is it meaningful?

v) Is it useable?

vi) Is it useful?

vii) How do you feel about this kind of data w.r.t. this company?

viii) For each subtype of data here (e.g. location data, web history data, address data, financial data, etc.):

- has it been provided?
- is it (or would it be) valuable to you?

c) In general, for your GDPR interaction with each company,

i) Was there anything surprising?

ii) Are there any causes for concern?

iii) what benefit, if any, have you got from doing this?

iv) What one thing would you do differently in the future?

v) Any other comments?

d) Having reviewed and discussed the response and any returned data from this company, revisit and, if desired, update your previous trust score for said company.

e) Having reviewed and discussed the response and any returned data from this company, revisit and, if desired, update your power score for said company.

2. Across all companies targeted, for each of your previously stated general hopes:

a) how well has it been met?

b) how do you feel about this?

3. Across all companies targeted, for each of your previously stated specific desired data uses:

a) is this now possible?

b) is this now practical for you to do?

4. Across all companies and GDPR requests made:
   - How do you now feel about your data being stored by organisations?
   - What would you like to do next regarding these companies and your data (if anything)?
   - How has the experience of GDPR been for you?
5. How has the experience of this study been for you?

Before Interview 3, participants were not given any training or instructions in advance about how to examine or assess their data returns, other than to check they could open and read the data (in some cases data was password protected or needed to be downloaded from provided links). Within interview 3, some help was provided in viewing data, if the participant struggled or asked for help. Other than this, the only structure to the examination of the data was the question schedule above.

5.4 GDPR Request Outcomes

In this section, outcomes from each interview stage are summarised, and a basic quantitative analysis of the interview data is presented, before qualitative findings are presented in 5.5.

5.4.1 Interview 1: GDPR Target Selection

Initially eight participants chose 5 target companies and three chose 4 to request data from. One participant (P9) withdrew from the study due to COVID-19 after Interview 1. Five participants withdrew a chosen company upon further consideration. Reasons for withdrawing chosen targets included:

- having one's personal data mixed with other household members (Netflix),
- the account being in someone else's name (Morrisons),
- not wishing to impact active customer support matters (LNER), and
- inability to contact the provider by email (ifun.tv, see below).

One participant selected Newcastle University, which was vetoed by the research team to avoid conflicts of interest. Hence, 41 out of a possible 52 GDPR subject access requests were made (to 28 distinct data holders) as shown in Table 5.1:

Type of Company	Company Names[a]
Major Internet Companies	Apple (3), Amazon, Facebook (4), Google (5)
Hardware Companies	Apple (3), Huawei, Google(5), Philips Hue (smart lightbulb manufacturer)
Online Platforms/Websites	Airbnb, Bumble (dating site), Check My File, Credit Karma, Direct Line, last.fm, LinkedIn
Social Networks & Dating	Facebook (4), Instagram, LinkedIn, Bumble (dating site)
Software/App Manufacturers	Freeprints, Niantic (creators of Pokémon Go), Natural Cycles (a menstrual tracker), Revolut, Spotify
Transport Companies	Tyne Tunnels, Nexus (Tyne & Wear Metro), LNER
Retailers & Loyalty Schemes	Amazon, Tesco, Sainsbury's, Nectar
Telcos	Virgin Media, Three
Sports Clubs	Sunderland AFC

[a] Where a company was chosen by more than one participant, the number of participants choosing that company is shown in brackets.

Table 5.1 - Types of Data Holding Organisation Targeted for GDPR Requests by Study Participants[a]

To ensure fairness and consistency, the aim was that all GDPR requests be sent by e-mail by the participant to the data holder's identified Data Protection Officer, requesting both a subject access request (Information Commissioner's Office, 2021a) and a data portability request (Information Commissioner's Office, 2021b) be initiated, and specifically

enumerating and asking for all those datapoints that the company stated in its privacy policy, as well as others which the GDPR entitles individuals to obtain. To identify these datapoints, company privacy policies were analysed and the necessary information was compiled in personaldata.io's semantic wiki ('List of target companies for GDPR requests', no date). This has a feature to generate bespoke GDPR request emails, which were adapted then provided to participants (Bowyer, Holt, *et al.*, 2022, supplemental materials). Facebook, Apple, Huawei and Philips Hue do not offer a contact e-mail address, so the email text (shortened where length restrictions applied) was pasted into a contact form. In one case, entertainment website ifun.tv, the only available means of contact was via WeChat, resulting in the participant (a Chinese citizen) choosing not to contact ifun.tv due to fear of Chinese government surveillance. Through analysis of companies' privacy policies and with reference to GDPR rights, a taxonomy of the types of personal data that could be returned was constructed, using terms from those privacy policies and GDPR legislation. There are five types of personal data, as shown in Table 5.2:

Type of Personal Data	Description	Examples
Volunteered Data	Data that the data subject has directly provided to the company through upload, contact or form completion.	Personally Identifiable Information (PII), contact details, user-generated content, photos, files, profiles, settings, communication history, financial information, security credentials, surveys/forms.
Observed Data	Data that has been indirectly or automatically collected about the data subject through product/service use or customer/staff interaction.	App usage information, behaviour on website, search/browse history, location tracking/tags, activity/health tracking, technical/device information, network/telco/ISP information, cookies & pixel trackers, staff observations, customer interaction notes.
Derived Data	Inferred data or profiles that have been created through algorithmic or human analysis of volunteered, observed or acquired data.	Interest profiles, advertising demographics, market segmentation, customer categorization, product/service recommendations, internal customer codes.
Acquired Data	Data that has been obtained or purchased from external sources such as civic records, reference agencies, advertisers or third parties.	Public records and information from internet searchs, reports or reviews from individuals, electoral roll data, credit checks, fraud checks, criminal record checks, e-mail/interest lists from advertisers, information shared between affiliates, sister companies or partner organisations.
Metadata	Information about how the other four categories of data have been handled, including storage, processing, uses, decision-making and external sharing.	Names of third parties data has been shared with, details of where data is stored and when/where it has exited the EU, explanations of how data has been used in automated or human decision making, legal bases for storage and processing.

Table 5.2 - Types of Personal Data Potentially Accessible from Data Holders via GDPR Rights

5.4.2 Interview 2: Privacy Policy Review and Goal Setting

In Interview 2, participants had reviewed and discussed privacy policies for their chosen target companies and been asked to define hopes and expectations for each GDPR request (see Table 5.4). 74% of goals expressed related to participants wanting to have greater insight and control into their personal data ecosystems; most commonly a desire to see the breadth and depth of data collection by companies, to understand what was being inferred and how personal data was used, and to use such information to better assess trustworthiness of those companies. Such goals were often motivated by curiosity or suspicion, or a desire to shed light on specific incidents or answer specific questions. In some cases participants wanted not just to learn and acquire knowledge but to take control of or delete held data. In contrast, 26% of goals related to gaining personal benefit from one's obtained data: motivators included the desire to reflect on past data to gain self-insight, as well as goals relating creativity, fun, and nostalgia.

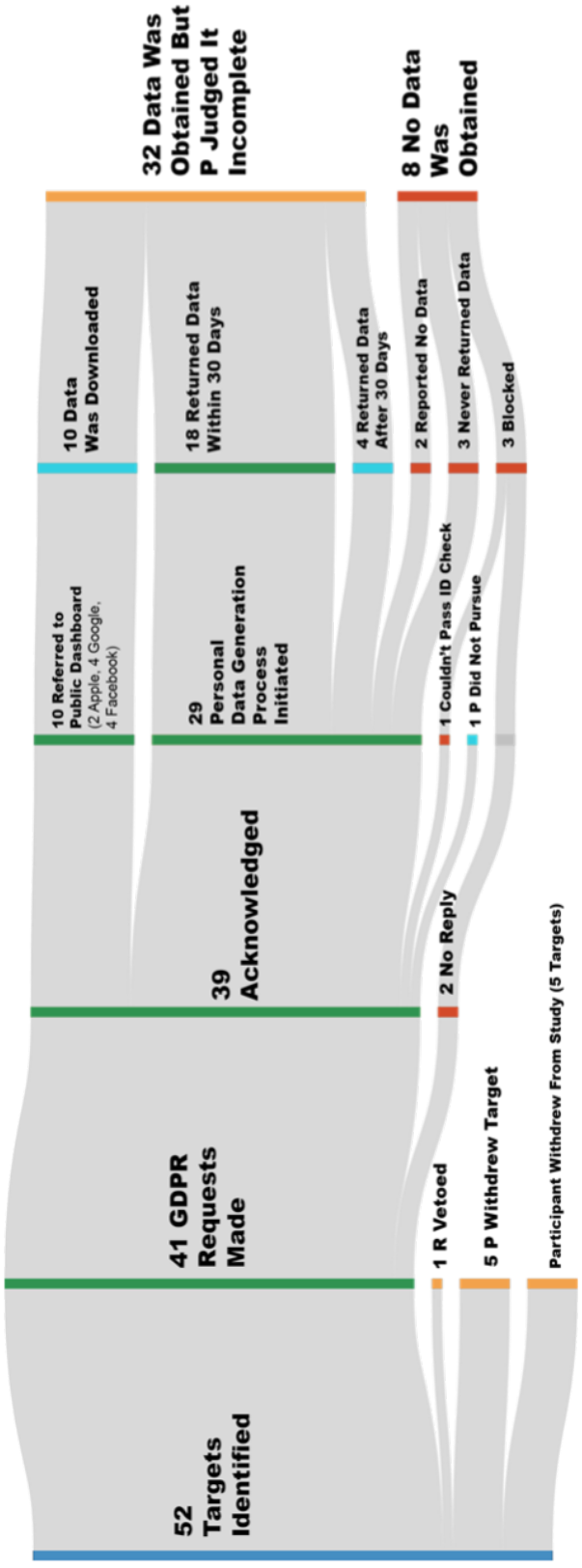


Figure 5.3: Sankey Overview of Participants' GDPR Requests

At the conclusion of interview 2, participants had been provided with the emails and instructions to start their GDPR requests, which progressed as illustrated in Figure 5.3. Eight requests resulted in no data being obtained, due to either data holder non-responsiveness, inability to access the right account or satisfy ID requirements, or confirmation being received that there was no data to supply. 32 requests (80%) resulted in at least some data being returned; 10 of these directed the participant to use a publicly-available download dashboard such as Google Takeout, and the rest resulted in data being made individually available. Of these, one was mailed as printouts, another was mailed on CD-R, and the rest were delivered by e-mail (sometimes involving a secured online website to download). While 22 companies supplied bespoke data packages, 4 did not return it within the 30 days the legislation specifies. Please note that requests took place within the context of a global pandemic so response rates may not be typical. Following discussion in Interview 3 (see below), participants judged that all 32 requests receiving data had failed to return all requested data (across all five of the categories in Table 5.2).

5.4.3 Interview 3: Reviewing the GDPR Response

Once each participant's GPDR requests had concluded (as described above), they were invited to discuss the GDPR responses in detail. Participants were asked to describe (and optionally show) the data they had received, then to evaluate the data holder's response for each data type, according to multiple metrics designed to assess the perceived quality of the GDPR request handling and the subjective value of any returned data. All questions were posed from the perspective of (a) the data that providers said they collect and process in their privacy policies, and (b) the rights that the GDPR specifies, to ensure discovery of missing data or unfulfilled rights would be considered objectively. Participant responses were considered quantitatively (see Table 5.3 below) and qualitatively [5.5].

Type	Valued? [b]	Returned?	Complete?	Accurate?	Understandable?	Meaningful?	Usable?	Useful?
Derived	82%	39%	10% (dk:13%) [c]	20% (dk:20%) [c]	40% (p:40%) [d]	40%	0%	20%
Acquired	81	49	16 (dk:16)	50 (dk:25)	75 (p:0)	50	25	17
Metadata	73	4	0 (dk:7)	0 (dk:0)	0 (p:100)	0	0	0
Volunteered	57	53	55 (dk:0)	92 (dk:0)	72 (p:20)	72	52	58
Observed	48	33	18 (dk:12)	57 (dk:30)	61 (p:20)	57	52	61

[a] Percentages represent the proportion of "Yes" answers to each question, per data subtype, from all those *where a judgement was given.*
[b] Participants were asked whether this category of data from each provider would be valuable *if they were to receive it.*
[b] dk = don't know (percentage of cases where participants felt unable to assess data accuracy or completeness).
[c] p = partially (percentage of cases where data was judged *partially* understandable).

Table 5.3 - Presence and Quality Assessments of GDPR Responses by Data Type (as Percentages[a])

Table 5.3 shows quality assessments for each data type, with rows descending by subjective value. Notably, the kinds of data participants value most (derived, acquired and metadata) were less frequently returned, especially metadata (returned in 4% of cases). Where data was returned in these categories, it suffered from poor data quality, often judged as incomplete, inaccurate, unuseable[11] and not useful (although acquired data was largely

understandable). At 53%, even the most returned category, volunteered data, was lacking. Where it was returned, accuracy (92%), meaningfulness (72%) and understandability (72%) were high. Observed data was least valued and also rarely returned or complete (yet judged to be of moderate quality). From the perspective of individual data categories, data was only judged to be complete in 22% of cases. In 62% of cases, personal data specified in privacy policies to be collected was not returned, despite the legal obligation.

Data collected in these interviews also allowed comparison of different companies. This analysis is included in ARI5.2.

At the conclusion of the final interview, participants were reminded of the specific hopes and anticipated data uses they had expressed at the start of their journey and asked about how well each goal had been met. These answers were recorded and combined to produce percentage values showing in how many cases goals were fully met, partially met, or not met at all, as shown in Table 5.4.

Participants felt their goals were not fully met in 78% of cases, and 54% were not met at all. Specific shared problem areas included (1) the desire to understand what providers infer from held data (7 participants), which was unmet in 73% of cases and only fully met in 7% of cases; and (2) the desire to delete one's data, which was a stated goal in 10 cases but was only met in one of them. Four wholly unmet hopes were to investigate specific incidents (GDPR responses were often delivered as a one-off package without any kind of backchannel or opportunity to ask questions), to secure data, to check accuracy, and to move data to another service.

Hope or Goal	Distinct instances of this goal	Distinct participants	Specific companies in mind for this goal, if any	Was this hope met? Unmet?	Partially met?	Fully met?
GOALS RELATING TO ACCOUNTABILITY AND CONTROL (74%):						
Understand the breadth and depth of what data is collected	24	7	Amazon, Apple, CheckMyFile, Credit Karma, Facebook, Google, LNER, Nectar, Philips Hue, Spotify, Tesco, Three, Virgin Media	42%	17%	42%
Understand what is inferred about you from your data	15	7	Amazon, Apple, Direct Line, Google, Instagram, last.fm, LNER, Spotify, Tesco, Three	73	20	7
Assess provider trustworthiness	12	6	Apple, Credit Karma, Direct Line, Facebook, Freeprints, Nectar, Niantic, Sunderland AFC, Tesco, Three	42	42	17
Remove your data & control/limit its use	10	3	Bumble, ifun.tv, Instagram	90	0	10
See inside 'black box' algorithms & processes	9	4	Amazon, Facebook, Google, Tesco	56	11	33
Understand how and why your data is used	6	5	Direct Line, Google	50	33	17
Investigate specific questions or incidents	4	4	AirBNB, Three, Credit Karma, Instagram	100	0	0
Learn about data use and how to be safer online; educate others	3	2		0	33	67
Secure data about you and identify risks and leaks	2	2	Apple, Facebook	100	0	0
Check accuracy of data about you	1	1	CheckMyFile	100	0	0
Move your data to another service	1	1		100	0	0
Test your data rights	1	1		0	100	0
GOALS RELATING TO USING DATA FOR PERSONAL BENEFIT (26%):						
Reflect on past activities & gain insights	14	5	AirBNB, Apple, Google, last.fm, LNER, Tesco, Virgin Media	57%	36%	7%
Find patterns/habits & track goals	6	5	last.fm, Nectar, Spotify, Tesco	17	50	33
Combine data from many sources for deeper insights	3	2	Philips Hue, Google	33	67	0
Play with, create, hack & remix your data	3	3	Google	67	0	33
Nostalgia, fun & inspiration	3	3	Spotify, Niantic	33	33	33
Keep your own data archive	2	2	last.fm	0	50	50
OVERALL	**18 goal types**	**10 people**	–	**54%**	**24%**	**22%**

Table 5.4 - Participants' Hopes, Imagined Data Uses and Goals for GDPR, as well as Resultant Outcomes

5.4.4 Perceived Power and Trust

Scoring questions were used to examine how participants' feelings towards the data holders changed throughout the process: Participants were asked:

- to assess trust from 0 (total distrust) to 10 (total trust), and
- to assess their perceived power on a scale of -5 (total provider power) through 0 (balanced power) to +5 (total individual power).

Explanations and reasoning for initial ratings and for any changes were uncovered through questioning. By repeating the same question at different times, longitudinal comparisons could be made. Many participants' attitudes did change as a result of the experience [Figure 5.4] for both perceived power (45% of cases) and trust (66% of cases). For those with changed attitudes, the change was often negative: in 63% of cases where participants perceived a change in individual power, that change was a loss in individual power, and in the majority (52%) of cases, participants felt more distrustful of GDPR targeted companies after completing the process (constituting 79% of cases where a change in trust was perceived). However, it is important to note that in some cases GDPR had a positive impact; in 17% of cases participants felt their perceived power had increased, and in 14% of cases participants felt more trusting of providers after GDPR.

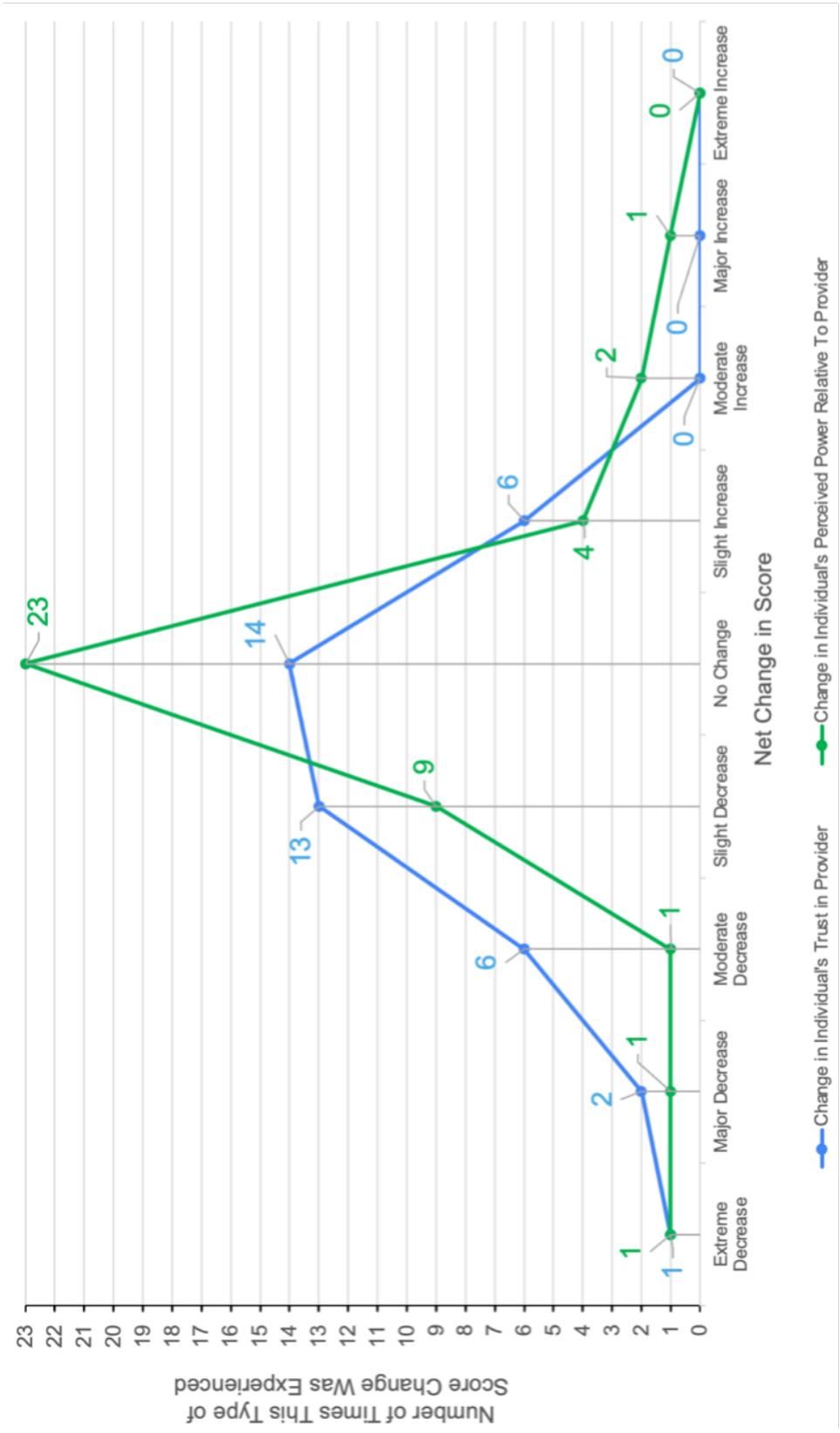


Figure 5.4: Longitudinal Distribution of Net Changes in Participants' Perceived Power and Trust Scores

Looking deeper into these datapoints, changes in attitude could be attributed both to the impact of reviewing the privacy policy as well as to the experience of the GDPR process and the discursive review of GDPR responses. Figures 5.5 and 5.6 show snapshots of power and trust ratings at different points in the process which illustrate these impacts. Looking to explain these changes qualitatively, it was found that privacy policies often contradicted participants' expectations, resulting in discomfort. In two cases (Philips Hue and last.fm) privacy policy review revealed that the service relationship was with a completely different company than the participant thought, which was disturbing to them. LinkedIn's privacy policy was noteworthy as being exceptionally clear, reassuring and trust-enhancing to the participant, largely due to its 'easy read' text sidebars but also good use of examples. However it does appear that simplifying privacy policies can go too far: Google's privacy hub (which includes video explainers) was considered easy to understand but necessarily broad (given their breadth of services) and thus over-simplified, raising uncertainty about generalisations made, and in some cases increasing distrust.

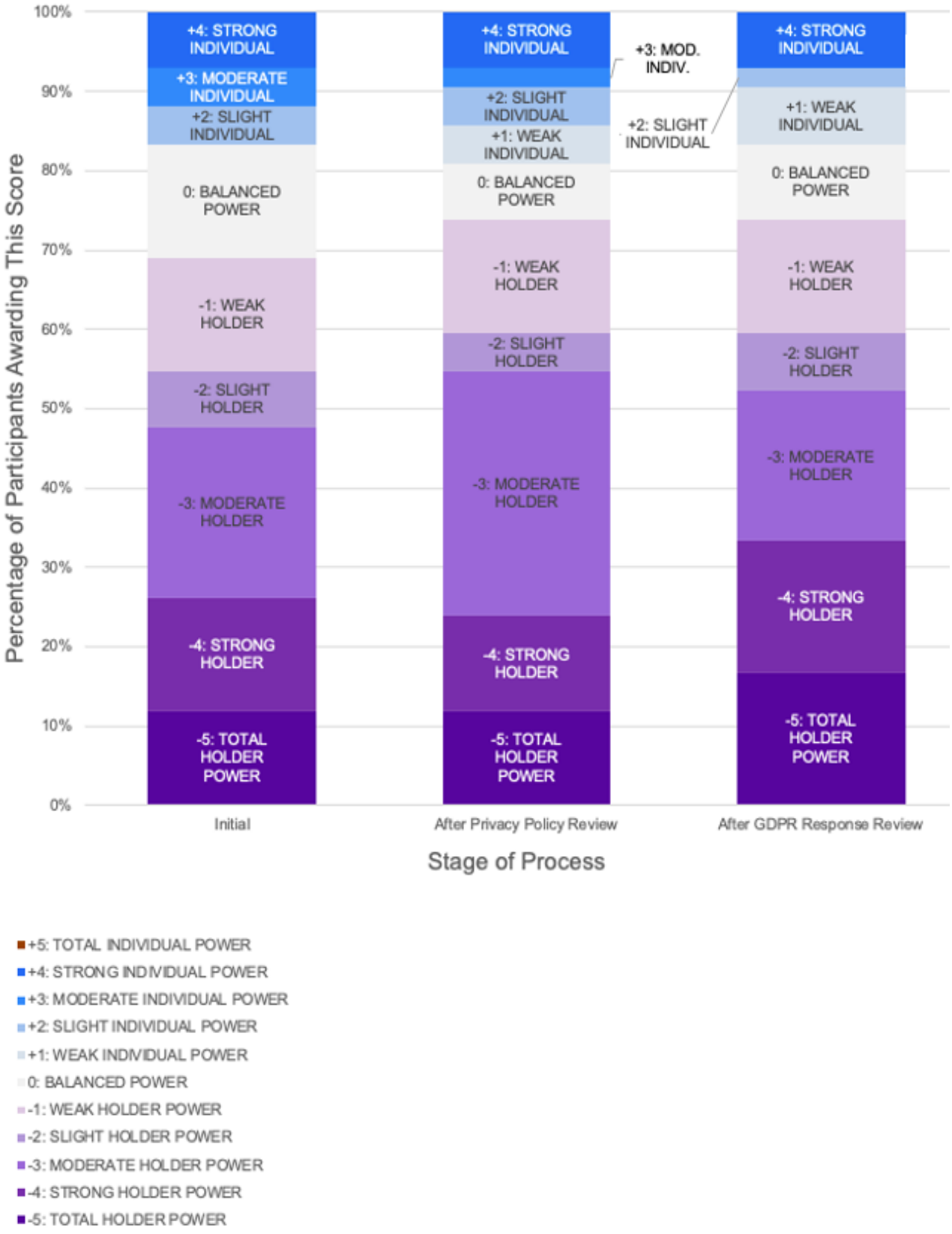


Figure 5.5: Perceived Power Balance Between Individual and Data Holder at Different Stages of the GDPR/Study Process

Considering the process as a whole, participants' attitudes were impacted particularly by the *"hassle"* (P11) they experienced in getting through the data access process, and from the realization that what seemed at first glance to be a thorough response, when scrutinised more closely in Interview 3 and viewed through the lens of the privacy policy promises and one's GDPR rights to the five categories of data, was in fact quite poor.

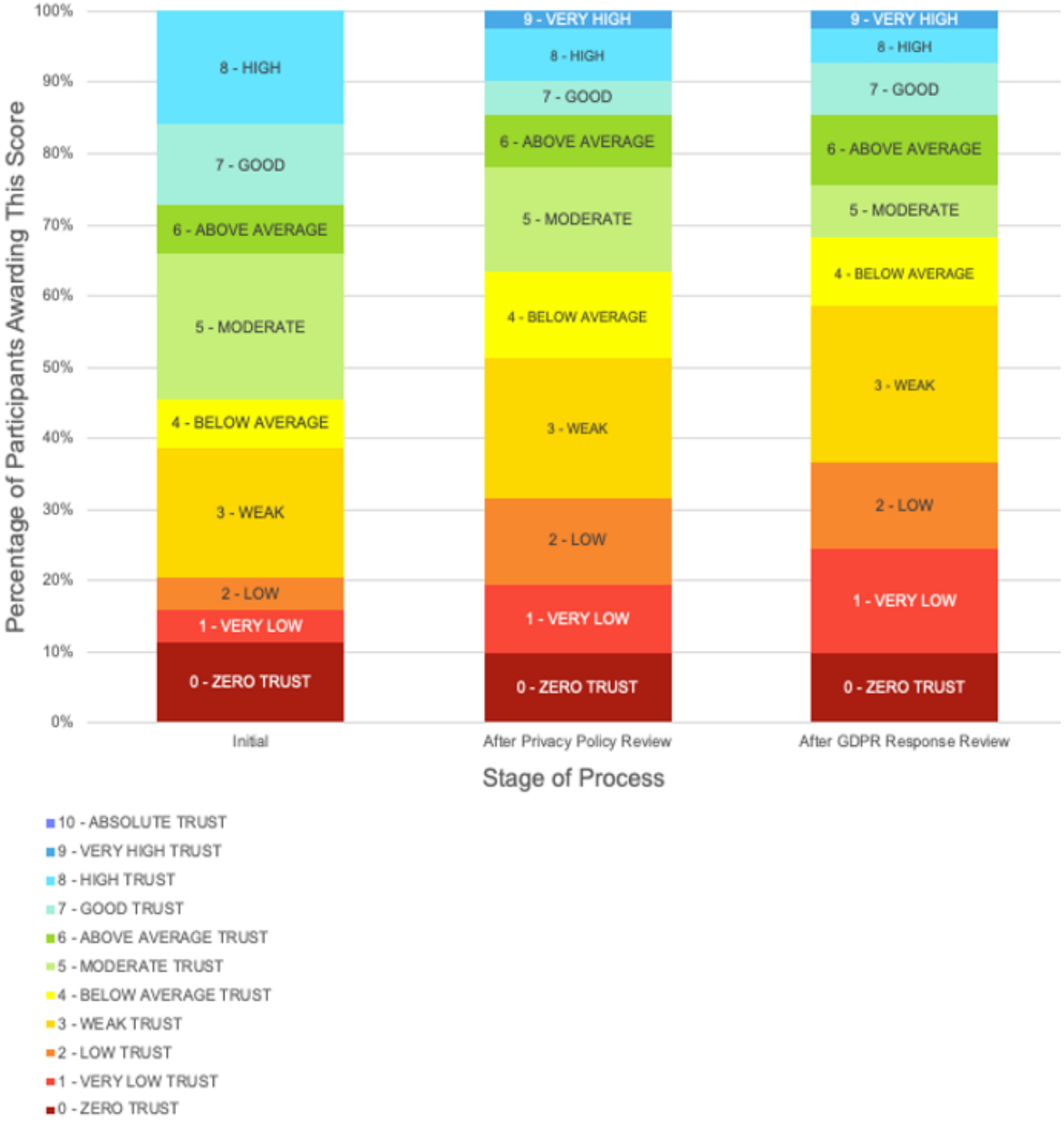


Figure 5.6: Participants' Perceived Trust in Provider at Different Stages of the GDPR/Study Process

5.5 Thematic Findings

As described in 5.3, the topic-focused transcripts were carefully examined to identify themes and findings, a process involving over 200 person-hours of iterative data analysis (Huberman and Miles, 2002) of the interview transcripts see ARI5.1]. In this section the findings from that analysis are presented and summarises, with the three main themes being introduced in 5.5.1, then each theme is further detailed in the the following subsections, including participant quotes.

5.5.1 Themes & Subthemes

The findings are presented through three main topics:

- Insufficient Transparency [5.5.2],
- Confusing and Unuseable Data [5.5.3], and
- Fragile Relationships [5.5.4].

Theme 1 (see 5.5.2 and summary in Table 5.5 below) describes the **Insufficient Transparency** that participants experienced in privacy policies and through the GDPR process; Organisations appear evasive over data when responding to GDPR, leaving people "in the dark" even after making GDPR requests.

Subtheme	**Description**	**Quote**
A Desire for Awareness and Understanding	Participants want to see, know and understand the data held about them. There was particular interest to see data collected or inferred about them without their involvement, and to understand how data is used and shared and how that might affect them.	*'[Companies have more power] because they're making decisions about things and you don't know how they're making those decisions.'* (P5)
Non-Compliance Without Consequence	Many providers failed to provide data on time or at all. In 100% of cases, returned data was incomplete, and many viewed this as non-compliance.	*'I am surprised at Google's unwillingness to provide me with all of the data … they haven't provided me with all of*

Subtheme	Description	Quote
	Data holders' freedom to disobey legislation was attributed to a lack of enforcement and seen as an exertion of power.	*my data. And that's not legal.'* (P7)
Inadequate Data Responses	Participants judged data holders to be unhelpful, GDPR procedures to be painful and ineffective, and returned data to be lacking in coverage and in quality. Their questions remained unanswered; after GDPR they were still *'in the dark'* (P4). There was widespread disappointment and a view that GDPR did not confer any power to the individual.	*'It's kind of disappointing because I would have hoped that this process would have levelled the user power versus the organisation power in a way that holds them accountable and [it doesn't] seem to be doing that.'* (P1)

Table 5.5 - ***Theme 1 - Insufficient Transparency****. Subthemes & Participant Quotes.*

Theme 2 (see 5.5.3 and summary in Table 5.6 below) explains how participants received **Confusing & Unuseable Data** from data holders; When presented with their data, people struggle to understand it and relate it to their lives and are not able to make use of it.

Subtheme	Description	Quote
The Search for Personal Value in Data	Participants found the large volumes of data that were sometimes returned overwhelming, and wanted summaries and breakdowns to understand it, as well as tools to help them make sense of and explore the (often technically formatted) data. Data that spanned a period of time was judged particularly meaningful as it could serve as a window into past memories and would allow for trends	*'[It's] almost too much […] for a normal person to be able to process and understand […] It could do with a document detailing, like, "this is what is in here".'* (P1)

Subtheme	Description	Quote
	and changes over time to be observed.	
Unuseable Data Formats	Participants anticipated receiving data in formats they could explore, visualise, mashup and play with, but in fact often received data that lacked explanations. Data was often arranged in ways that were more reflective of internal systems than being optimised for use or understanding. Both useable data files and explanations of how to use it are needed.	*'They did give me the data, but not how it fitted together. It's like being given the bricks to a house, and then they're like "Here's your house". It doesn't really mean anything when it's just bricks, if you don't know how to put it together.'* (P5)
The Liability of Data You Can't Delete or Control	Having understood that the amassed data about them could be exploited by businesses or third parties, participants wanted to see their data so that they could understand the extent of that capability, and wanted to be able to restrict its use and, in some cases, delete their data. No clear pathways to take such control were offered, nor was this control practical to achieve; the retaining of data against their wishes was seen as a liability and a lack of control.	*'[Companies did not] tell me what they are doing with [my data]… And sometimes I think my willingness to give a company data might be quite intrinsically linked with what they're gonna do with it.'* (P7)

Table 5.6 - ***Theme 2 - Confusing & Unuseable Data****. Subthemes & Participant Quotes.*

Finally Theme 3 (see 5.5.4 and summary in Table 5.7 below) examines the **Fragile Relationships** that individuals have with data holders; Companies' data practices, and in particular their privacy policies and GDPR response handling, can be impactful to customer relationships, carrying a risk of damaging trust but also the potential to improve relations.

Subtheme	Description	Quote

Subtheme	Description	Quote
Power and Enforced Trust Through Data Holding	Participants feel that the sacrifice of (or the giving of permission to collect) personal data is a necessary cost in order to get the valued benefits of the services they want to use, something they are pressured to do and have no choice about. Such sacrifice is seen as the giving up of power, as participants lack access and control to that data. In the face of providers making decisions based on data and processes that they could not observe, participants felt powerless. This amassing of data was sometimes seen as surveillance, and some saw great potential for misuse and abuse of it.	*'For me to have power over my data, I think is a fair and normal thing. But for a company to have power over [my] data means that it's basically a proxy to have power over me.'* (P8)
Perceptions of Data Holders	Participants entered the study with varying perceptions of providers' integrity, influenced by reputation, business model and size. Participants' various observations reveal a strong link between their perceptions of providers' data handling practices and the trust they hold in those same providers.	*'When I like the company already, I'm more willing to give them my data.'* (P2)
Changed Perspectives Through Scrutiny	In general, the more that participants found out about data-centric practices through the process of scrutinizing privacy policies and making data access requests, the more they distrusted providers. Failure to explain or provide complete data was harmful to trust. Conversely, where	*'If someone's not completely open with you, then you're like, "What are you hiding?"", which means you trust them less.'* (P4)

Subtheme	Description	Quote
	providers were more transparent or participants did obtain interesting data insights, trust was increased.	

Table 5.7 - ***Theme 3 - Fragile Relationships****. Subthemes & Participant Quotes.*

5.5.2 Theme 1: Many Companies are Evasive and People are "Still in the Dark"

A Desire for Awareness and Understanding

As Table 5.4 shows, in the vast majority (62%) of cases, participants wanted to see, know and understand what data was held about them and how it was used. For example, P11 wanted to know what data was collected by train company LNER when he bought tickets, so that he might judge whether it was appropriate:

> *'I'd be interested to understand what data they have [...] Is it just the patterns of my spending on trains, or is it a bunch of other stuff that they're using for advertising to me?*'—P11

Beyond the data that participants had directly volunteered [Table 5.2], most data was currently unknown to participants. In particular they wanted to gain awareness of what data might have been collected without their knowledge.

> *'The bit that concerns me is where I don't know what data is being taken by companies. If I'm registering for a library or something, I know [what] data I'm giving to them, but what I don't know is all the other stuff that they're recording'*—P9

Participants were equally unaware of what holders might infer from the data they had collected. P4 wondered if Philips could use data from his smart home lighting to deduce his sleep and TV-watching routines. P7 had received targeted advertisements relating to pregnancy that she felt weird about because she did not understand why she had been targeted in this way. P5 raised concern about how data inferences could affect decision making, surmising that the data holder had greater power than him because *'they're making decisions about things and you don't know how they're making those decisions'*. Sharing of personal data is also insufficiently visible to participants; two participants (P3,P4) targeted

GDPR requests to credit-check websites (Credit Karma, CheckMyFile) - P4 wanted to get *'a picture of what other companies can currently expose'*.

Non-Compliance without Consequence

As detailed in 5.4.2, few requests resulted in a timely provision of requested data (44% or 68% depending whether referral to a download portal is excluded or included in the count). Many data holders responded late or not at all; such actions are objectively a breach of legislation. However, participants were broadly unsatisfied even when they did receive a GDPR response. In 100% of cases where data was obtained, it was considered incomplete, and this was usually seen as further failure to comply. Participants had reviewed their GDPR rights in Interview 1 (though, as expected (Rughiniș *et al.*, 2021), most were already aware), and so several participants saw this apparent non-compliance relative to their understanding of their rights as a poor quality of response, for example:

> *'I feel more concerned now, […] what they've given me seemed reasonable. But then comparing against what we asked them for, what I'm legally [entitled to], it's a fraction.'*—P5

For some participants, sceptical from the start, such poor responses were consistent with their expectations; P6 found the incompleteness of Facebook's response *'alarmingly unsurprising'*. Others had expected compliance:

> *'I am surprised at Google's unwillingness to provide me with all of the data… they haven't provided me with all of my data. And that's not legal.'*–P7

Many participants, reflecting on a feeling of having less power than they had initially thought, felt that the prevalence of non-compliance showed that too much power relative to the authorities, that a lack of pressure is being applied by regulators and that *'there needs to be more enforcement'* (P11). P6 revised his view of Facebook's power versus his own because he felt that after review he now could clearly see *'which [data] they are prepared to share and which they aren't'*. P11 also framed the selectivity of responses as an exertion of power:

> *'It seems like there's a lot of derived data about things like purchases and stuff [that I would expect] that just isn't there. So they're free to not give me the data. That, to me, suggests [that despite GDPR] they retain an awful lot of power.'*–P11

Inadequate Data Responses

While in some 22% of cases participants did meet their goals through GDPR (see Table 5.4), when it came to the desire for greater awareness and understanding [5.5.2], this want was largely unmet. Only volunteered data such as basic personal information or user-generated content was usually returned complete; this was often viewed as mundane and uninteresting, and the focus on these data types in returns was viewed as evasiveness. Facebook, P6 observed, *'give you that kind of descriptive boring data which is mainly all publicly available anyway'* and had omitted *'the stuff that I would consider valuable to them'*.

In general, the data responses did not provide the answers participants sought. Many reported *'still'* not knowing what they wanted to find out. P4 said he remained *'in the dark'* (P4). P7 stated that *'even though I did the process correctly, I still didn't get that much back'*. Concerns held by participants from the outset remained unaddressed, as in P11's case:

> *'I still am quite concerned about how much data organisations have, particularly how they link that other data and how data is bought and sold, and I haven't really got any answers on that.'*

It was not just the data returned, but the process itself, that participants were dissatisfied with; requesting and achieving data access was time-consuming and difficult. *'Jumping through hoops'* was a phrase used independently by four different participants (P4, P5, P7 & P11) to describe the experience. Some found data holders obstructive and unhelpful:

> *'I feel like they give you a response that [makes it so] you cannot proceed intentionally'*—P10

Participants recognised that they had received help and coaching, and that the processes were so tedious that without that, they may not have persisted. P1 suggested that without the provided template, it would be *'a lot harder to get meaningful data out'*, and P7 attributed her sole successful request to the guidance she had received in progressing it. P5, having experienced problems with expiring links, delayed responses and missed emails, had been surprised at *'how difficult it was just to get my data, and the fact that I had to ask them about six different times'*.

Not all requests were this painful, some were handled smoothly. As P11 put it, *'Some companies make it dead easy to get, but then the data is not massively useful.[...] Other*

companies make it a pain in the neck to get it.' Overall the view of GDPR data access was one of disappointment. Participants found GDPR ineffective: P10 said *'Frankly, [GDPR] doesn't have as much influence as I expected'* and P1 commented that:

> *'It's kind of disappointing, because I would have hoped that this process would have levelled the user power versus the organisation power in a way that holds them accountable and [it doesn't] seem to be doing that.'*

5.5.3 Theme 2: People Struggle to Understand, Use and Control Their Data

The Search for Personal Value in Data

Prior to receiving data, participants had anticipated discovering insights about their own lives by browsing and reflecting on their personal data, consistent with personal informatics literature (Li, Forlizzi and Dey, 2010). However, there was a comprehension gap between the useful information they imagined and the actual data returned; data was typically delivered as a bundle of technical files, which were hard to understand and often delivered without explanation. Some felt (in line with *effective access* [2.1.4]) that they lacked the necessary skills or tools to make the data understandable or useable *'for a non-techie person'* (P11). When the researcher guided P7 to *jsonlint.com*, an online formatter, she found her JSON-formatted data more understandable. P2 made the point that data holders must be using tools themselves to make sense of people's data: *'**They're** not just looking at a JSON file, so I would like to have the same visualisation [as them].*'

There was a sense that by sending people individual data files, data had been removed from the environment in which it has meaning, and that the returned data excluded necessary context for interpretation. This was often manifested in the form of internal codes and abbreviations that individuals could not understand. P4 stated of his experience looking at smart-lightbulb data from Philips Hue, that there was *'just so much of it that it's impossible to know [what it all means]… You'd have to spend a few hours going through this and being like, "OK, what does that line mean, and that symbol, and that code?"'*. This lack of context also materialised as a failure to explain decision-making processes: P5 reflected, when looking at driving scores from a car insurer that uses a mobile app to monitor her driving, *'I could see the data; it was the score that was weird for me. Like, it doesn't tell you how it's calculated.'* P1 noticed that although some companies did make some effort to explain the

returned data, this varied substantially across providers. He said that *'it would be nice if these companies had a standardised model of how this information is presented to people, so it [could] be easily understood'*.

One of the greatest obstacles to understanding that participants faced was being faced with a large volume of information and no means to quickly digest or navigate it; either very large files, or complex hierarchies of nested directories containing many separate files. It is clear that there is a need for *summaries* so that participants can quickly get a handle on what is - or is not - present. Returned data *'could be valuable if you knew what the hell [was] in there'* (P4). P1 described one of his data responses as *'almost too much [...] for a normal person to be able to process and understand.'* He said that it *'could do with a document detailing, like, "this is what is in here"'*, and described the disparity across responses as being *'either like death by thirst or death by drowning [...] It would be better to drown, but still not ideal'*. Ultimately it is clear that in general, returned data was not presented in a way that is optimised for understanding.

Another question that our findings were able to shed some light upon, in service of RQ1, was to consider what precisely makes data valuable to individuals. This is especially important given participants did identify the potential to gain personal benefits from their data (as seen in the second set of goals in Table 5.4). An idea that came up again and again was that data is most valuable when it *spans a period of time* and can be related to events in the individual's life over that period. This could potentially provide new insights to participants.

P2 for example hoped to see, or be able to construct, breakdowns and charts that would help him examine his food shopping habits. Through the GDPR process, P10 accessed details of her spending on micro-transactions in the mobile game Pokémon Go that had not been available to her through the app. P11 wanted to derive insights about his train travel by examining the geography, cost, journey length and patterns of his past journeys through data he hoped to receive from LNER. Long-duration data offers the potential ability to identify trends and changes in one's own behaviour over time.

It was these historical parts of their data that participants found most meaningful, offering as it does a means of remembering, with data potentially serving as a *'window into your*

past' (P11). P5 saw value in perusing music-listening data *'just because it's cool to look back on stuff that you've done and you don't necessarily distinctly remember it'*. Generally the longer period the data covered, the more valuable it was deemed to be:

> *'I would actually be interested in last.fm, partly because the data goes back to 2008 … Spotify only goes back about four or five years and not everything I listen to is on Spotify.'*—P11

P6 saw the data accumulated by service providers as potentially forming part of a valuable background context to understanding life events in his past:

> *'I would like to […] build a picture, not just like, "I remember going to Reykjavik", but if there's other data around that time [I could] sort of paint a biography of myself'*

He described some of his data as *'a kind of personal history that has been quantified and sort of datafied'*.

Unuseable Data Formats

This personal value that captured data has the potential to offer shows that it is all the more important that participants be able to understand and make use of their data. Our participants found that the format in which data was returned often meant that it was not only difficult to understand, but difficult to use as well. Using data meant different things to different participants, with imagined uses including budgeting, record-keeping/archiving, or using the data for creative or fun purposes. Some participants (e.g. P5) saw value in potentially combining data from multiple sources, though this did not turn out to be practical. Participants did not know what data to expect, and generally imagined returned data being more useful than it turned out to be:

> *'I think … you could do some interesting mashups, but I don't really know what with until I've got the data. It depends on the data; I'm sure there could be some cool uses of it.'*—P4

Once data was received, participants struggled to interpret and understand it to a sufficient extent to be able to identify the useful data or meaningful information they had hoped for. Returned data formats and response structure were extremely varied. Some reported that there was not sufficient machine-readable data to make use of the data. For example, P4 received a Microsoft Word document full of pasted screenshots from an internal portal as

part of his response from his ISP Virgin Media, and said that its usefulness *'depends on what you want to get out of it, really. If you want to view the data they have about you, it's quite useable. If you want to do something automated[analytical], then it's not*'. P11 found a similar returned screenshot from an internal system to be *'completely non-understandable'*. In other cases, the opposite problem occurred, with data being too technical for the participant to use. P10 said of JSON data: *'For normal people who don't understand programming, I feel it's just, there's no use at all.'* P7 felt she lacked the technical proficiency to make use of the returned data:

> *'They have provided it in formats where I can see that, if I were a developer, I could do things with it, […] but if I was not that sort of person, it might be quite difficult to understand'*—P7

In P5's case, she saw the potential to use the data but felt that what was missing was additional explanation or guidance on how to interpret it:

> *'They did give me the data, but not how it fitted together. It's like being given the bricks to a house, and then they're like "Here's your house". It doesn't really mean anything when it's just bricks, if you don't know how to put it together.'*—P5

P11 highlighted a problem with his Tesco shopping data that was not just a matter of formatting or skill, but the granularity or focus of the data itself:

> *'As a technical person, having a CSV of data is quite useful, potentially, but actually what can I do with that if it's Tesco's internal systems data?'*

On the face of it this finding seems to contradict the previous one—i.e. demands for both more technical and less technical data might seem contradictory. However, what we can infer is that participants collectively need *both* useable technical data *and* easy-to-read information summaries - and that those summaries should cover both the relatable life information encoded within the data *and* the information *about* the data, what it means and how to use it. This idea is explored further in (Bowyer, 2021a).

The Liability of Data You Can't Delete or Control

Having recognised that potential value of data relating to their lives, before or during this research, several participants were concerned about personal data being held. P10 for example said with reference to dating site Bumble: *'Since I found my partner [and therefore*

no longer need a dating site] I deleted my account and I've been wondering, "Are they still keeping my data at the back?"' and with reference to both Instagram and Bumble, expressed a desire to have her data deleted and expected GDPR to play a role in the enforcement or verification of that deletion, something she could not otherwise be sure of. P8 considered the holding of sensitive data to be a liability that she was only willing to tolerate while she was actively using a service, and this was part of her motivation for targeting Natural Cycles:

> *'I now use a different one, but I used, for about a year [their] app to track my menstrual cycle. [It was my] main contraception method, so that's things that this company probably has. Now that I'm not using it any more, I don't know if they delete the things or not.'*

Many participants expressed a desire that data be held only for a short time, and questioned the default practice of data being kept beyond the period where it was needed to deliver a service:

> *'The thing that concerns me is that I haven't used Tesco online for at least four or five years, so why are they hanging on to my IP address from five years ago?'*—P11

He went on to spell out the liability he saw in such apparently mundane data being held, the liability coming from the duration of the data: *'10 years of worth of shopping records… how much would that be worth to a health insurance company, and would [Tesco] succumb to the temptation to sell that on?*' P10, a Chinese citizen, identified long-term sources of personal data as an enabler for future privacy violations, saying that *'in China, [there is a trend] that as soon as someone becomes famous, people begin digging [through] all their past experiences*'.

Most participants described the ability to delete or enforce the deletion of their data as having control over it, and given the current practical lack of such a capability felt that they had insufficient control over data holding. One of the first steps participants identified in gaining control of their data was simply an ability to see it, for accountability, so that they might check the accuracy, security and breadth of collected data and flag any unforeseen concerns. They felt that a deeper understanding might lead to an increased sense of individual safety and data control and facilitate them to make changes in data habits or choice of service provider:

> *'I want to understand how much they're keeping. And what they're doing with it. I'm hoping that by knowing that, I might change my behaviour about all the data I accidentally create.'*—P7

In this participant's case, this hope was unsatisfied, and upon looking back at her experience she remarked:

'I guess that's one of my criticisms of GDPR in general - that although I can understand what data a company holds about me, there's no obligation for them to tell me what they are doing with it.. And sometimes I think my willingness to give a company data might be quite intrinsically linked with what they're gonna do with it.'—P7

In fact, that legal right does exist through GDPR, but as we can see it was not delivered in practice. What participants want is to feel aware and in control of their data; this must begin with better data legibility and explanations of data use, accompanied by clear pathways to enable data correction or deletion.

5.5.4 Theme 3: Poor GDPR Handling Can Damage the Forced & Fragile Trust Relationship

Data Holders Enforce an Uneasy Trust

The lack of visibility and control over personal data discussed in the previous section, combined with a sense of being in the dark [5.1] about data practices, caused participants discomfort before, during and after the GPDR process. This stemmed from a sense of finding themselves facing uncertain risks they feel powerless to change. Many participants, discussing their relationships with providers, expressed a range of emotions ranging from curiosity to anxiety and distrust:

> *'I'm curious… I wonder what they've got on me. […] If it's anything other than the barest minimum that is necessary for them to do their job […] then I get creeped out by that.'*—P11

Participants felt most uneasy about the amount of *'intimate'* (P1,P2) data that providers collect. P1 was uncomfortable about Facebook having information about his social circles. P2 said he felt *'quite vulnerable'* that his Google search terms *'say pretty much everything you have done… the most intimate things you were thinking about'*. P11 singled out ISPs as

having the potential to track everything their customers look at online, noting that *'I don't think you've got much choice about that.'*

Participants also felt that some data holders held so much data that it had begun to resemble surveillance, such as in the case of P1, who used *'an absurd amount of [Google's] services'* and reflected that *'if I'm driving somewhere, I've got Google Maps open, so they know exactly where I'm going, they know how fast I'm going, they know what I'm listening to while I'm driving…'*. Participants saw the potential for abuse, fearing this kind of deeply personal knowledge could be *'used against'* them (P2). P11 felt that Apple had enough data to *'screw me over'*, and P5 considered that her car insurer Direct Line uses data to *'judge'* her, noting that *'it's not like I can contest the data and say "Actually, no, I disagree".'* In a more extreme illustration, P10 shared her fears that data collected by WeChat and Weibu (Chinese services similar to Facebook Messenger and Twitter respectively) would be at risk of abuse against citizens by the Chinese government. In some cases participants were able to identify concrete instances where providers had exploited the personal knowledge they held: in P6's view, Facebook use their knowledge of their users' friendships and relationships to *'hook your attention'* and prevent users deactivating accounts in a *'disingenuous'* manner.

Whether or not data is used nefariously against individuals, thinking about the potential for this caused participants to associate the mass collection of personal data as an acquisition of power over them: *'[Companies that] know a lot about everyone will inherently be able to have power either through persuasion or manipulation'* (P1). P7 saw the holding of data as the source of holders' power: *'when I think about other people having my data [...] the control isn't sitting with me.'*. Others identified the ability of data holders to deny or limit access to data as their key source of power:

> *'If you're not getting what you perceive to be yours back in completion [sic] then you're not in control of your own data and you have fairly little power over it.'*—P1

The view of data holders having more power in the service relationship (mirroring reports described in 2.1.2) was reflected in participants' evaluations of power balance: in 69% of relationships participants felt that the data holder had more power than them (rising to 74% after GDPR), whereas in only 17% of cases (unchanged by GDPR) did participants feel they themselves had more power.

Several participants equated power over a person's data with power over the individual. When asked to define power in the context of data, P8's unprompted comments aligned with my Case Study One findings describing power over data as a proxy for individual participation or inclusion in decision making [Bowyer *et al.* (2018); 4.3.2; 4.5.1]:

> *'For me to have power over my data, I think is a fair and normal thing. But for a company to have power over [my] data means that it's basically a proxy to have power over me.'*

A key dynamic to understand these relationships is that individuals sacrifice their data in exchange for value - that value being the capabilities offered by the services. All 11 participants expressed the idea that the sacrifice of data is something that they have grown to tolerate in exchange for some benefit. P6 tolerates data collection by travel agents because *'they might help me pick a better deal next year.'* P11 said he was happy for Tesco to collect data in order to *'profile me to try to sell me more cheese, fine, whatever,'* though expressed caution that he doesn't *'know what else they're doing with it,'* and more generally was *'deeply concerned'* about unseen data trading. The benefit can be convenience too; P10 had logged into Pokémon Go with her Facebook account, knowing that implied data collection by Facebook, *'because it's much easier'*.

Participants often felt this sacrifice was something they had no choice about, but they did not like it. Unease over the trade-off being made surfaced most often in the context of recommendations; generally, participants valued data-derived suggestions provided they were *'relevant'* (P1, P8) and not too *'intrusive'* (P1, P6). It is clear that data sacrifice is only tolerable up to certain limits: P10 said of Niantic found the sacrifice acceptable provided that *'they don't sell where I live or my daily routine'*; however while Niantic's privacy policy promises data is not sold, it does appear that some level of personal location information *is* accessible in some form to third-party advertisers (Varghes, 2019). P8 said that relevant music recommendations were *'very useful'* but found Amazon shopping recommendations *'very scary'* because *'I don't want to see that I'm predictable'* and felt that *'if someone out there knows [what I want] before you [it's] like taking agency away from me.'*

Permission to collect and use data is knowingly provided by individuals to data holders, but the mechanisms to do so are considered inadequate: P2 felt that permission giving options are *'not granular enough'*, and in P11's view *'it's not a negotiation at all, it's all or nothing.'*

Worse, some participants feel that permission is coerced from them: P10 observed that Niantic *'pressure you into'* giving continuous access to your location data by tying it to the availability of in-game benefits such that *'you don't want to lose out'*.

Such lack of choice or coercion led to feelings among participants of resignation about data collection, seeing it as a Hobson's choice:

> *'I feel like it's inevitable that if you want to access their services at all, in any normal kind of way, that you automatically have to give them your data.'*—P7

Ultimately, participants felt that their data was *'revealing'* (P2, P3, P11) a lot of information about them, and so their only real option to maintain their privacy was to prevent data collection in the first place by not using that service at all (P1, P2, P3, P7, P10, P11), and living with the subsequent lack of service capability.

Perceptions of Data Holders

The discussion of attitudes towards data holders through this study allows some insights to be drawn as to how data holding service organisations are perceived, particularly with regard to data handling:

Factors such as reputation, size and business model were often a major contributor to participants' impressions of companies. For example, P2 described feeling *'more at ease'* with Apple, due to their hardware-oriented business model, than with Google, who *'make money through data'*; in general, where there was a lack of clarity around how a company makes money, or that model was clearly based exploiting sacrificed personal data, there was a greater suspicion, while trust was higher in those companies that offered a paid service:

> *'One of the main things was there [are] no ads. [Natural Cycles is] a paid service, so there's no, like, "you don't have to pay but we use your data to make money".'*—P8

Returning to Apple, P2 also noted that they *'position themselves as a defender of privacy rights'* and along with P11 (another participant who had targeted Apple) held a more favourable view as a result. P10, on the other hand, had been influenced by a documentary she had seen, becoming suspicious of Apple's control over her hardware.

While attitudes to Apple were generally positive, Facebook—which has, and continues to be, the subject of much negative media attention over its apparently cavalier attitudes towards personal data—was held in much lower regard. P6 said Facebook had *'in every shape or form, shown themselves not to be trusted'*, an opinion formed from *'high profile news stories where they have done unscrupulous things and are very willing to just hand over data'*. P9 reported feeling *'slightly dubious'* about Amazon as a result of *'[press coverage] about their ethics that may or may not be true, and just the size of them… and just the level[amount] of data, as well'*. Clearly expectations around data handling are a strong influence on attitudes toward service providers, though sometimes other factors play a role, such as with P8, who was comforted not just by Natural Cycles' payment model, but the values they project: *'This is woman-empowerment-orientated [sic], so in that sense I think I do put my trust there as well.'*

As well as these more impression-based influences, it also became clear that participant's direct experience of interacting with a company affects their feelings toward that provider. P1 found that *'in the same way that Amazon is quite janky [unreliable and awkward to use], Google feels fairly polished and so I trust them more'*. As well as customer/user experience, a perception of receiving a valuable service creates trust: P4 said of Google that *'the amount I trust them is in line with the utility I get from them'.* In the context of data sacrifice, high levels of trust do have an effect on customer behaviour:

> *'When I like the company already, I'm more willing to give them my data.'*—P2

Changed Perspectives Through Scrutiny and Transparency

Longitudinal examination of participant's feelings of perceived individual trust and power across their GDPR experience allows the impact of the experience to be analysed. As illustrated in Figure 5.5 and 5.6, trust in data holders tended to diminish through the data request process. Some of this distrust arose from examination of privacy policies, for example in P5's case who commented that Spotify *'shouldn't need to know that much about me, they should just give me music'.* In most cases the most noticeable decline in trust occurred between Interview 2 and 3 (when the participant experienced the GDPR process) or within Interview 3 (where the returned data was examined), showing that both the quality and coverage of the data return, as well as the execution of the data request process,

often have a detrimental effect on trust. Individuals' perceived power, however, did not undergo a corresponding change:

> *'They've not given me everything back that I thought they'd be collecting, which makes me trust them less. So power-wise, I don't think [anything]'s changed, but trust, I think it has.'*—P1

The absence or sparsity of derived and acquired data and of metadata around sharing and handling [5.4] noticeably damaged trust. P1 directly attributed his reduced trust scores to what he saw as non-compliance [5.5.2] through failure to return all data categories. P5 lowered her evaluation of trust in Spotify further still upon completion of Interview 3 *'because they didn't say anything about what they're doing with my data or where it's going'*. P8's trust score Natural Cycles was similarly reduced *'because I think it's hard to get any sensitive data, and it's not really made clear what they're using it for'*.

It is also clear that poor GDPR handling in itself can damage trust, independently of the data returned. P2 reduced his trust score for AirBNB *'because of the way they've handled [the data request], and the way they've made it hard for me to read the data'*. Similarly, P7 downgraded her score for LinkedIn *'because I feel like they have my data and [they've] not bothered to find my data, and that makes me feel like I shouldn't trust them quite as much'*.

Participants want greater transparency than the current processes are currently providing, and the failure to do so is a direct cause of distrust:

'If someone's not completely open with you, then you're like, well "What are you hiding?", which means you trust them less.'—P4

Despite the impacts on trust, both using GDPR access rights and the wider process of scrutiny and discussion surrounding that process within this study had a positive impact on participants' awareness, offering *'insights into how big companies are actually handling these requests'* (P7) and how to practically use one's data rights, showing that transparency (even in the hampered form of current GDPR handling) has an education benefit. Participants had initially expressed wishes to gain insight into data handling practices in order to increase accountability and inform their decision making on provider loyalty and privacy settings. GDPR offered the potential to compare data expectations with reality—for example P11 was initially *'curious to find out if [Apple's] marketing claims match their reality*

around privacy'. While such broad goals were generally unmet, several participants found the process thought-provoking and reported feeling more aware about what data they were enabling their providers to gather. P4 felt the process *'got me thinking about, like what other things could I try, and what other sources of personal data are there'*. P8 reflected that *'it's a skill and a kind of knowledge that I think everyone should [have]. I don't think it [should be] normal that I felt so clueless'*. Some commented on the value of understanding GDPR itself through the experience:

> *'[I] think the exercise was useful in that I understand what a GDPR request can do and what it cannot do. And there's a lot it cannot do. And I think it might seem that it gives you a lot of power, but really, it doesn't.'*—P2

While considering the negative impacts of the GDPR experience on trust some realised the potential trust-engendering impact that a more transparent response could have brought:

> *'I think the lack of transparency in a lot of these processes has not helped, you know, if Tesco had […] plain English processes for getting the data and you've got the data in a plain English way, that would do a lot to bolster trust.'*—P11

In a small number of cases, this was witness in practice, with a good GDPR response actually increasing participants' trust in certain providers. For example, P5 reflected that she may have been *'a little harsh'* in her initial judgement of Instagram and said she *'actually really liked what they sent… in comparison to the three others, I was genuinely, I opened Instagram's one and I was like "this is really cool"'*. P10 was very impressed with the response from Niantic and after GDPR she trusted them very highly *'because they replied really fast, the data provided is very detailed, and their attitude towards this whole issue is very positive,'* concluding that they are *'a really nice company'* and even indicating an increased willingness to spend money on their product. P6 trusted Sunderland AFC because *'they were really kind of upfront and… I got the data from them first, […] no messing about, the format they gave me just made sense'*.

In these comments, we can see an indication that, although the data requests often did not live up to the hopes of the participants, positively engaging with the process was influential and did affect participants' outlook. In particular, close attention was paid to the willingness of companies to be transparent and forthcoming, with GDPR representing an opportunity to

test organisations on their data practices and assess their integrity and competence as holders of their data.

5.6 Discussion

This study examined the GDPR's effectiveness in improving individuals' access and control over their personal data. The participants' experiences support the existence of a power imbalance over data [2.1] and suggest GDPR largely fails to empower individuals: both objectively (to the extent possible by this limited sample), in that most companies do not comply fully (either by returning insufficient and inadequate data, or by failing to return data on time or at all), and subjectively, in that returned data was often difficult to understand, impractical for use, and raised new questions and concerns. The findings also indicate that swift, transparent, and easy-to-use GDPR procedures can positively impact an individual's perception of an organisation. In light of these findings, this discussion offers insights on how the personal data landscape might be redesigned through policy [5.6.1] and business practice [5.6.2], and how individual action can have important impact too [5.6.3]—all in pursuit of the human-centric empowerment goals described in 5.1, 2.2 and 2.3).

5.6.1 Implications for Policymakers

Compliance, Quality and Ongoing Access

Despite significant and obvious GDPR-motivated investment by service providers in dashboards, processes and bespoke data package production, the findings (while limited by the small number of participants) indicate that inadequate compliance with the GDPR is common. The findings are consistent with literature too: the participants' issues with completeness and compliance echo those first reported within the GDPR's first year (Ausloos and Dewitte, 2018), suggesting completeness and compliance have not improved over this period. However in this study, the focus was on the effectiveness and experience of engaging with GDPR procedures from the individual's perspective. Participants' experiences were overwhelmingly of *disappointment* and *frustration*, with their hopes rarely met. They found that data holders often did not engage meaningfully with the process, and that the responses typically excluded or obscured data that could have provided them with the insights into their data privacy and the organisation's data practices that they sought. Evaluations of perceived power compared to data holders

largely remained the same or worsened after accessing data through GDPR, and participants were not confident in the capabilities of the legislation to shift the balance of power. The process was perceived by some as a *'box-ticking exercise'* that was both frustrating and time-consuming and did not ultimately help them. Even though in 7% of cases participants did feel empowered by the GDPR, *all* participants receiving data were in practice left with the prospect of additional time-consuming and sometimes technically-skilled work to take advantage of or interpret their returned data. This suggests that to improve the situation, policymakers need to make changes towards:

1) Better Compliance Through Enforcement of Complaints. At present, enforcement of the GDPR is uneven; each country has its own DPA (for example in the UK, this is the Information Commissioner's Office or ICO) and complaints are rarely pursued for individual cases. Instead, cases are processed by specific DPAs in a form similar to a class action lawsuit. This means that individuals have little impact when they do raise a complaint, and many GDPR complaints "become lost or resulted in lengthy delay" (Burgess, 2021), or may even be erroneously dropped (Lomas, 2020). Until individuals have a clear and effective means to issue complaints (Baker and Karasti, 2018) that result in enforcement action (or a clear threat of it), it is likely that individuals will continue to have little recourse other than to repeat the request and hope similarly dissatisfied individuals will act on their behalf. Data holders must be held to account when they do not deliver the full set of data that they report possessing, or when they fail to do so within the legally obligated time frame.

2) Policies to Enforce Better Quality Responses. Many participants received data in frustrating formats, including screenshots, printouts or files that were too technical or littered with acronyms. Data was provided in formats too technical to understand, or not technical enough to be useable [5.5.3], showing a demand for both human-readable information summaries and machine-readable data files, where most providers typically provide only one or the other. Policymakers could provide suggested data formats or even propose new standards; this would help data portability, improve effectiveness (Gurstein, 2011) and legibility (Mortier *et al.*, 2014), can reduce costs through common tooling and catalyse the building of tools to interpret and understand data. Such standards are emerging (Morgan, 2020) as they are a technological necessity for data unification, but lack adoption. We note that the European Data Protection Board has published new guidelines

(European Data Protection Board, 2022a) that could help somewhat to improve GDPR responses, though these do not offer new standards, and will not be as effective as legally-mandated policy changes.

3) Policies to Enforce Data Access as Ongoing Support, not One-Time Delivery. A radical redesign of policy is needed to give people the practical outcomes they desire and, according to the GDPR itself, deserve. Data access needs to be seen as more than the one-off delivery of data files. People need understanding of their data and of its handling. Not only that, they need a timely, up-to-date and ongoing view of the changing picture of how they are seen in data, and the occasional 'snapshots' with a 30-day delay that GDPR delivers can never deliver this. Giving people an ongoing awareness and understanding of their data is the measure by which compliance should be assessed. The explanations GDPR mandates are not forthcoming; of the 119 hopes expressed by participants (see Table 5.4), 70 (59%) related to acquiring greater understanding of data practices. 38 (54%) of these were unmet, and a further 15 (21%) were only partially met. By mandating data holders to support individuals with not just the delivery of data, but assistance to understand (and potentially make use of) that data, policies could become more impactful, not least because such understanding is critical to inform judgements around consent, loyalty and compliance.

5.6.2 Implications for Data Holders

Earn Trust by Opening Up Data and Enabling Users

While this study, and the GDPR itself, might seem adversarial to data holders given the goal to reduce their power by imposing new procedures, the findings emphasise the role of personal data in consumer relations. Data holders are likely aware of the paramount role of personal data in decision making, but may not be aware of individuals' perceptions about this. The findings suggest that failure to satisfy users who are concerned about the collection and usage of their personal data risks harms to consumer trust and confidence, at least for those users, and perhaps for others they might influence. In turn, however, this presents opportunities to use the mechanisms of the GDPR for customer loyalty and building better relations.

In 52% of cases, following the process of examining privacy policies and engaging in GDPR data requests resulted in a decrease in reported trust in the data holder. While such impacts

may for now be minimal, as only a small proportion of users read privacy policies (Steinfeld, 2016) and—one can assume—an even smaller number conduct GDPR requests, this is likely to change as issues around data privacy and trust continue to take centre stage in global geopolitics (Véliz, 2020; Zuckerman, 2021). Furthermore, the emergence businesses focused on *'getting your data'* or *'taking control'* ('Whose data is it anyway?', 2019; Dehaye, 2021; CitizenMe, 2021; Gener8, 2021; 'About Us', no date; 'datacy - About Us', no date; 'Ethi', no date; 'Digi.me', no date; 'Exist.io', no date) suggests demand for data access is growing. From the findings, there are three positive takeaways for data holders:

1) Data transparency is an opportunity to increase customer loyalty and trust. GDPR's basic rights provide a starting point for delivering practical data transparency that will allow organisations to demonstrate that they are deserving of trust. By responding clearly and engaging openly and helpfully with GDPR data requests, organisations can demonstrate consistency between their privacy policy and their actions and demystify to their users the role that data holds in their business model. Research has shown that explanations can *'ease humans' interactions with technology […], help individuals understand a system's function, justify system results, and increase their trust'* (Glavic *et al.*, 2021). This was borne out in the results: in 14% of cases, participants felt more trusting of the service brand as a result of their GDPR experience (sometimes even displacing prior apprehensiveness or distrust), citing reasons such as speedy, hassle-free responses, clear and understandable data, providers being upfront and open with data, and staff who exhibited a positive attitude to the request.

2) Data transparency is an opportunity for improved and re-imagined customer relations around data. Beyond the opportunity to improve trust, the mechanisms of data transparency suggested by the GDPR provide individuals with new capabilities for data curation and involvement. By offering individuals the ability to engage in empowering data interactions, data holders have the opportunity to improve engagement with their organisation and their services. If organisations view personal data as a shared resource to be curated and co-owned by the individuals that contributed it, there may be correspondingly shared benefits: for the individual, a sense of agency, influence and negotiability (Mortier *et al.*, 2014); and for the service provider, an incentive for individuals to generate and share more data, an increased likelihood of individuals correcting

inaccurate data, and more reliable and human-centric forms of ongoing consent closer to dynamic consent (Kaye *et al.*, 2015) than today's ineffective models of informed consent (Luger and Rodden, 2013).

3) New customer demands indicate untapped business opportunities. As the 500-member-strong *MyData Global* organization [MyData.org (2018); 2.3.4] shows, there is growing demand for personal data empowerment. People's personal data is splintered and trapped (Abiteboul, André and Kaplan, 2015; Bowyer, 2018b), and they cannot correlate data from different sources in order to reflect upon it, gain insights, and set goals (Li, Forlizzi and Dey, 2010). Due to commercial motivations, service providers generally deliver capabilities within a closed silo, not at the level of one's wider environment (Abowd, 2012). To be better empowered the individual could be the point of integration, the centre of their own Personal Data Ecosystem (PDE) (MyData, 2017). Life-level capabilities (Bowyer, 2021a) and the opportunities that well-designed and well-regulated GDPR-type regulations promise in this regard have not yet been exploited. Thorough, complete and timely data access in standard formats, as mentioned above, will be critical to enabling this vision. As discussed further in 9.3, growing companies such as CitizenMe (CitizenMe, 2021), digi.me ('Digi.me', no date), Mydex (Mydex CIC, 2010), ethi ('Ethi', no date), HestiaLabs (Dehaye, 2021), udaptor (Udaptor, 2021) and exist.io ('Exist.io', no date) as well as larger organisations like BBC R&D (BBC R&D, 2017) and Microsoft (Microsoft, 2021) are already starting to innovate in this space.

5.6.3 Implications for Individuals

Becoming Aware of the Value and Power of Data, and Demanding More

While participants experienced disappointment and frustration in their GDPR journeys, all participants gained new understandings; if not always of their data itself, at least of their target companies' approach to data access requests. This new knowledge was sufficient to re-affirm or challenge existing attitudes or inform judgements—P1, for example, left Facebook soon after the study. Even an attempt to access data can be educational, and even a cursory look at a provider's 'What data do we collect' privacy policy section can provide pause for thought.

Today, individuals remain largely in the dark about the collection, use and sharing of their data through a combination of perceived complexity and effort combined with a lack of clear benefits. Table 5.4, alongside the increased control and insight promised by the PDE movement and platforms linked in 5.6.1 and 5.6.2 above, provide a glimpse of what the future may hold: a world where individuals take more control of their data and gain actionable self-insights. Three key messages for individuals can be inferred:

1) Your data is used to represent you and define your user experience. We hand over our data in exchange for access to services, but providers then use it (usually in aggregate) e.g. to inform product design or decide what content we see. This 'innocent' handover of data is in fact giving providers the means by which we are treated and – at times – controlled. Recognizing the crucial role of data (and our limited influence over it) is the first step to pursuing greater agency and control.

2) Your data contains meaningful and valuable information about your life. Data, as participants found, is dry and technical, but they all sought meaning and value within it [5.2.2]. Within provider-held data lies potentially rich information about one's life and past activity – some of which can even be inaccessible through any other means. This highlights both a risk (that others might gain this insight) and a potential benefit (that we could access this insight ourselves). In this context, data deletion without keeping a copy may be inadvisable. To access the value in data, individuals will need to demand data standards, better access and control mechanisms and insight tools.

3) Self-education and awareness enable accountability and informed choices. The findings highlight a lack of knowledge. Transparency is critical to judging 'to what extent the bargain is fair' (Larsson, 2018). It is not always delivered, but GDPR makes it your right; a right that cannot be fully refused. Through challenging poor GDPR responses and demanding better information, individuals can have impact. Providers are ultimately motivated by public demand—one of the reasons download dashboards exist. Through the public pressure of negative attention, companies can be motivated to improve data access (Dehaye, 2018). With patience, GDPR rights can be exploited to force small changes.

5.7 Summation of Case Study Two

Through a longitudinal study of 10 participants lasting three months, this case study has qualitatively, and to a lesser extent quantitatively, evaluated the human experience of using one's GDPR access rights and of living with data-centric service provider relationships.

The findings, while not statistically representative, suggest that people currently lack awareness of held data and its uses by service providers. By guiding participants on a journey of discovery and careful scrutiny, encouraging them to draw their own conclusions about service providers on the basis of companies' own promises, individuals' legal rights, and participants' own hopes (see Table 5.4), this research has shown that such a journey can be educational and enlightening with regard to increasing awareness, but also can seriously damage brand loyalty and trust in providers if comprehensive and well-explained data is not returned in a supportive and open manner [5.5.4].

The experience of GDPR seems to be an unsatisfactory one for individuals; participants were generally still *'in the dark'*. Serious problems with compliance have been highlighted [5.5.2]: Participants received data that was incomplete, impractical for use, and they failed to acquire desired explanations. By its own aim to enhance individuals' rights and control, the GDPR does not succeed. Participants continued to feel a lack of agency and choice, were largely unable to pursue goals such as data checking, correction or deletion, and their perceived sense of power within the provider relationship was largely unchanged by the experience. Nor does the GDPR allow individuals to adequately pursue their own goals related to accountability, self-reflection or creative data exploration [5.5.3]. Individuals cannot be given power over their data through designing better Human-Data Interaction interfaces alone, but only through redesigned policies and business strategies that consider the sociotechnical context (Baxter and Sommerville, 2011; Bowyer, 2021a).

In order to bring the human-centric 'personal data ecosystem' concept closer to reality, action must be taken to improve both compliance and quality of GDPR responses. Considering these findings, there is cause for radical policy reform, to move away from 'data access as package delivery' and to provide individuals a more effective and ongoing two-way window into their data [5.6.1], providing ongoing awareness, accountability, and negotiability. Data needs to be expressed to individuals in ways they can understand, as

little to no practical impact is currently being achieved by delivery of a one-time snapshot of some technical files; in fact, we have shown such responses can be harmful to customers' perceptions of the data holder in many cases.

For providers, the risk of reputational damage uncovered by this study should motivate them to engage meaningfully with data access requests; but such risk can be averted by redesigning both interfaces and processes to approach data access experiences as an opportunity to educate, and to build trust and loyalty, perhaps even through establishing progressive co-operative data stewardship relationships that truly *involve* the service user [5.6.2]. The pragmatic challenges for organisations wishing to improve relations with their service users through data are explored in 10.2. While the GDPR experience is often disappointing and frustrating, it can provide insights that help individuals to challenge their assumptions, re-evaluate choices, and in some rare cases, feel empowered to act upon their data. Wider assertion of GDPR rights could demonstrate a desire for data holders to be transparent; without such visible demand, little may change [5.6.3].

Considering RQ1 [3.3.1] (the pursuit of a deeper understanding of people's attitudes to everyday data holding and people's wants from that data), this work suggests that people struggle to develop the meaningful relationship with their data that they desire because of the difficulties faced in seeing, accessing and understanding it. They are aware that within data is the potential for value to themselves, but cannot access that value, which in turn causes feelings of resignation, concern, distrust or suspicion towards data holders. What they seek most are two things: sufficient understanding to better judge the value exchange they have signed up for with providers (see goals in top half of Table 5.4), and good quality insights from data that would allow them to understand themselves better, learn from the past, set personal goals, and harness personal data for their individual benefit (see goals in lower half of Table 5.4). This duality of needs around data interaction is expanded upon in (Bowyer, 2021a).

With respect to RQ2 [3.3.2] (the pursuit of a better understanding of the role of that data in everyday service relationships), the findings suggest that personal data, held by providers, as in Case Study One, serves as a proxy for direct user involvement, and is treated as such. Once users have sacrificed their data, or given permission for its collection, they are rarely consulted and most services exclude them from seeing how that data will travel through the

organisation and be used in decision making; this is consistent with the *point of severance* concept (Luger and Rodden, 2013). As a result, the trust relationship between service provider and service user is extremely fragile, highly susceptible to subjective impressions of service brands. As the findings show, discovery of poor data practices or a lack of transparency around data is sufficient to harm that relationship and, in some cases, even motivate individuals to change provider. As discomfort grows and scrutiny occurs, providers can expect customers to lose trust and loyalty. At the same time, this same data could play a central role in a re-invigorated relationship between a provider and a user, one based upon *earned* trust. It appears that providing easy, clear, data access and showing a willingness to respond to questions and explain data usage to users could be sufficient to allay concerns and instil strong customer loyalty. This assumes that the openness offered reveals practices the user finds agreeable. Perhaps this in some way explains why some data-focussed companies (such as Facebook and Google)–whose practices many would find disagreeable upon examination–are apparently less willing to engage in transparency and user empowerment around data.

The general principles of earning trust through transparency, and rethinking data access as a means to involve users in decision making, could be applied in a wide range of service endeavours that are currently very data-centric.

In the next section, the findings from this Case Study will be examined and discussed together with those from Case Study One, in order to look for commonalities that might validate each Case Study's findings against each other and enable more general answers to RQ1 and RQ2 to be formulated.

SECTION III:

DEFINING A NEW RESEARCH AGENDA:

HUMAN DATA RELATIONS

Introduction to Section III

It will be already evident to the reader that there are significant overlaps and parallels to be drawn across the findings and discursive insights in Case Study One (CS1) and Case Study Two (CS2). Section III will take a step back to examine both Case Studies together through the lens of the research questions. The goal here is to produce a unified summary of findings and insights that can serve as a concrete answer to the two research subquestions RQ1 and RQ2. To recap on the research objectives expressed in 3.3, these two RQs are:

- **RQ1: "What is the human experience of personal data, and what do people want from their data?"**
- **RQ2: "What role does data play in people's service relationships and how could relationships involving data be improved?"**

In the first discussion chapter, Chapter 6, these RQs will be answered, drawing references from both Case Studies and from prior literature. A standard participatory research thesis would then conclude at this point, but the dual track approach of this PhD [3.2.2; Figure 3.2] allows the research to not only specify what people want from their data and relationships, but to explore how those goals might be actively pursued through innovation, activism and design. Therefore, in Chapter 7, the uncovered wants around data and relationships involving data are formulated into a defined research agenda, laying out the objectives and context for the practical pursuit of these goals. This new research agenda, named *Human Data Relations (HDR)*, provides the basis for the remainder of the thesis as well as a framework for others to continue to pursue these data-related wants in future, especially through the approaches laid out in 9.3, 9.4, 9.5 and 9.6; and through the Principles for Generative Action provided in 10.1.

Chapter 6. Discussion - An Understanding of What People Want in Data Relations

"For a participatory system to work, you need: a plausible promise, an effective tool, and an acceptable bargain."

—Clay Shirky
(author and journalist focused on the social
and economic effects of Internet technologies)

The answers to the two RQs [see Section III Introduction above] can be more usefully expressed as an understanding of individual ***wants*** relating to data. The word 'want' is used here in a broader sense than its everyday meaning, referring to the *lack of something that would be beneficial* (which may or may not be accompanied by conscious desire). Framing the accumulated understandings from the Case Studies in this way both exposes both the problem (the things that individuals do not have or cannot do) and also provides clear goals that any imagined solutions or improvements to the status quo would need to address. It logically follows that any solution that better delivers on individual *data wants* will lead to improved relations between individuals and their data. These wants also allow us to understand what 'better' means, serving as a yardstick against which progress can be measured. The concept of *data relations* will be developed further in Chapter 7, but for now can be understood as two kinds of relationship people have with their data:

- ***direct* interaction** (such as through an interface in an app or website) and
- ***indirect* interaction** (through interacting with services, providers or individual representatives who themselves have access to personal data about the individual).

RQ1 [3.3.1] and RQ2 [3.3.2] map cleanly onto these two types of data relations. To answer RQ1 we must identify what people want from direct data relations–which is the focus of 6.1. To answer RQ2 we must understand what people want from indirect data relations–which is the focus of 6.2. Once those two questions are answered, these new understandings will then be contextualised [6.3] as an answer to the overall research question, in conclusion of the participatory track [3.2.2] of this thesis.

6.1 Answering RQ1: What do people want in *direct* data relations?

By comparing and grouping elements of the findings from CS1 (see 4.3) and from CS2 (see 5.4), three distinct data wants are evident when considering *direct* data relations. Based on the findings and discussions thus far, we can conclude that, according to an individualist perspective, all data about individuals should be:

1) **Visible**: People want to have knowledge of data about them and an ability to see it and effectively access it;
2) **Understandable**: People want to be able to interpret this data to extract meaningful information from it (and about it), including through visualisations and summaries; and
3) **Useable**[11]: People want to be able to act upon this data, including exploring it, asking questions of it, using it to serve their own goals, and gaining personal value from it.

These wants are detailed in the following subsections.

6.1.1 Visible Data

Data matters to every individual now, in a way that previously it did not. As described in 2.1.2, the role of data in our everyday lives has changed; data has become a material used by businesses to shape our world. In 2.2.2 and 2.2.4 I outlined how this change has practically manifested itself in people's lives. Where once data was viewed as *ours* for our personal reference and use, the emergence of complex multi-party data ecosystems has meant that personal data management has become a sociotechnical & societal problem, not a practical individual one. In the past, when businesses were more local, more personal and less data-centric, the data they held about us was minimal and much less significant to our lives than the human relationships we had with them. Businesses grew and data began to be considered as a resource to be processed at scale for customer insight and marketing exploitation. Though we didn't realise it, our need to understand those processes to protect one's own interests began to grow. In the past, you didn't need to become aware of data storage and use, because it had little effect. Now, data has become a substitute for direct communication with the individual being served, as my research in both public sector [PS; 4.4.3] and the private sector [5.5.3] has shown. In both domains, people do not have awareness, let alone access, to the extent of data that exists about them. In 5.5.4 we saw

individuals feeling that companies forced them to hand over data in exchange for service access, and then subsequently maintaining power over them through holding that data, using it to make decisions, and denying them access to that data or even be clear about what data is held.

Data sacrifice is now required for many services, putting individuals at risk. Be it the personal financial, health and lifestyle data collected on an Early Help (EH) assessment form when a family signs up for EH support [4.1.2], or the contact details, payment information and preferences provided when individuals register with commercial service providers such as insurance providers or streaming media platforms [5.5.4], supplying your personal data is required to access services. Consent to hold and use this data is enforced upon signup, through waiver forms or Terms & Conditions agreements. Service providers in both sectors see the acquisition of more personal data as beneficial to their operations, be it support workers wanting to gain more data about families' lives [4.3.3; 4.3.5], or commercial providers using trackers to gain more insight about users that they can exploit for advertising (Binns, 2022). In both domains, this sacrifice is seen to have an emotional effect on people, ranging from curiosity to fear and distrust [PS; 5.5.4]. Such fears are well-founded, with mistreatment through incorrect data known in both settings [4.3.2;5.5.4]. While data holders almost certainly do not *intend* to cause harm, data *can* be 'used against you' (P2's quote in 5.5.4) (Kröger, Miceli and Müller, 2021; Strohmayer *et al.*, 2021). In providers' eyes, people are now represented through data [2.1.2]. Despite the fact that data is never truly objective (Gitelman, 2013; Taylor *et al.*, 2015) and a recognition (at least on the public sector side) that a data record can never tell the full story [4.3.5; PS], the data record becomes the object to be administered, rather than the individual (Cornford, Baines and Wilson, 2013; Zuboff, 2019), and this in itself creates risk - through mishandling or inaccuracy. Given the data record is seen as a source of truth [4.1.2; 5.6.3], it is vital that such information remains *fair*, and *accurate*; this is especially important in the commercial sector, where people are only just beginning to become aware of data misuses and data exploitations that are actually happening (Chang, 2018; McNamee, 2019; Zuckerman, 2021; 'Data Brokers: Everything You Need to Know', 2022). Ensuring fairness and accuracy of held data cannot be verified without individuals' awareness of data held about them.

Once data has been sacrificed, it enters a closed and opaque ecosystem, where the individual loses access and becomes unaware of that data's storage and use (Luger and Rodden's 'point of severance' (Luger and Rodden, 2013)). What was previously available for individuals to see becomes inaccessible and invisible. In the EH context, this manifested as families having a lack of awareness or direct access to data held about them and having to rely on support workers as gatekeepers to choose to inform or show them aspects of their data [4.1.1; 4.6]. In the commercial context, the situation is perhaps even worse, as not only is there rarely any kind of data viewing interface, there is not even a gatekeeper who might make people aware of their data and its use (given DPOs rarely engage individually). Even if someone becomes motivated to gain awareness, the GDPR leaves them in the dark; in 62% of cases, the data that companies own privacy policies stated they collect, was not returned, and data that was returned was complete in only 22% of cases [5.4.2]. In both contexts, no awareness is gained unless the information is actively sought. This means that the vast majority of people, busy and unaware, remain so. This is problematic, because people cannot judge data accuracy or protect themselves from risk, because they may not even be aware of certain data's existence and use, or be able to access it even when they are.

People want to see data which is hidden from them. In the SILVER project [see Section II Introduction; 4.3.2], PS, and in CS1 [4.3.5; 4.4.2; 4.5.2], families wanted to see what data was held about them (*'what they've got on me'*). They wanted to be actively kept informed and to have the ability to see if data was fair and accurate. In the commercial context, the same feelings were found; participants expressed a great desire to see and know what companies are storing about them, especially data collected or inferred about them without their involvement. This is not just a desire, but a need, given that data can have impact on their daily lives as it is used to inform decisions on how content is presented and recommended to them, and what services they are advertised, offered or can access [5.6.3]. We also see from CS2 that awareness is not a binary. It includes having an appreciation of why the existence and use of certain data is significant and what its implications might be. Article 13.2.f of the GDPR ('Article 13: Information to be provided where personal data are collected from the data subject', 2020) states that, at least in the case of automated decision making, people are entitled to meaningful information about the significance of the processing of their data, yet such explanations were typically not given to participants in CS2.

Effective access to held data is required for visibility. Having gained awareness of data held and of the significance of it, people want that to be accompanied by meaningful access to the data itself. In CS1 [4.4.2; 4.5.2], we saw families and support workers recognise the need to accommodate the differences in families' digital literacy, mental and physical handicaps, and technology skills while providing them access to the data held about them. This mirrors Gurstein's call for *'effective access for everyone'* (Gurstein, 2011). Four aspects in particular are relevant here: the content and formatting of the data (which should support different levels of linguistic and computer literacy), the capabilities made available in terms of software, hardware and Internet access (sufficiently powerful, sufficiently available and affordable), and skills (ensuring that individuals are able to interpret the data). In CS2, we saw several participants feeling that data was delivered in too-technical formats [5.5.3], or that they lacked the skill to properly interpret the data [5.5.3]. Effective access and interpreting data goes beyond visibility of data and includes understandability, which is explored in the next section.

Visibility of, and access to, data must be timely, and ongoing. Given the ever-changing nature of data (and indeed of the lives of the people it represents), occasional or one-off access is not sufficient. In HDI theory, this concept is described as having *negotiability* (Mortier *et al.*, 2013, 2014): the ability to re-evaluate data and associated decisions as contexts change externally. It is also mentioned by Gurstein, who points out that time-limited access to data would not be effective (Gurstein, 2011). In the EH context, families wanted access to their data outside of support meetings; this implies the availability of self-service interfaces that you can use *in your own time*, rather than being reliant on the support worker as gatekeeper. People wanted to see all data about them directly, through a personal interface, as reflected in their workshop designs [4.4.2]. This echoed PS findings that continuing rights and visibility of data over time are needed in order enable vigilance over keeping data accurate and meaningful as life changes. Timeliness also implies that access to an up-to-date view of the data does not require special and ongoing effort by the individual, it is always available. Both support workers and supported families saw value in notification feeds about changes to data records [4.4.3], so that changes are discovered and can be acted upon without having to wait for the next support meeting. In the commercial context, we can see that GDPR provides a form of access that is not at all timely. The 30-day delay on request processing guarantees that data will be out of date by the time it is viewed,

and individuals must repeatedly make GDPR requests to maintain an up-to-date view (and in doing so, they risk the imposition of charges as GDPR states that requests should not be excessive and that fees can be levied for additional copies). This lack of timeliness in the design of GDPR data access motivates my third suggestion to policymakers in 5.6.1, that they should offer ongoing access rather than the one-off delivery of data packages.

For held data to become visible, systemic support is needed, including governance, advocacy and assistance. Offering access to data is not solvable at a purely technical level. Even a well-built data interface with 24/7 access would not provide the depth and breadth of visibility people want. As observed in CS2, even those companies that provide instant data access portals such as Google and Facebook did not provide participants with all the data they desired, nor all the answers they sought [5.5.2], and most companies offered negligible follow-up support after data had been delivered [5.4.3]. Further investigations into data access conducted as part of the *digipower* investigation [See Section IV Introduction] confirmed CS2's findings that SAR requests and data portals rarely provide insight into some of the most desired types of data including derived and acquired data and data transfers. Effective access and visibility also requires advocacy (Gurstein, 2011): people require support and training to make use of their data. Furthermore, given the insufficient breadth of returned data from companies [5.4.3; 5.5.2] and near-total lack of access to data on the public sector side [4.4.2;4.4.2], it is clear that external governance (Gurstein, 2011) to ensure effective access is needed. Without the sort of pressure on data-holders that only policymakers can exert, organisations will not be compelled to provide richer responses or better information-access support [5.5.2; 5.6.1], and while small improvements can be achieved through individual action, people generally lack the means to effectively *demand* the increased visibility required [5.6.3]. The impact of this lack of governance is most keenly felt in the PDE/MyData space [2.3.4], where emergent actors seek to encourage data-holding organisations to enable greater information access so that they might build better data access tools for individuals, but are hampered by a lack of top-down governance supporting their requests as well as a lack of funding and investment by data-holders in data advocacy.

If data is not visible, this can lead to subjection, alienation and exclusion. Throughout both Case Studies, we have seen the negative psychological effects of people not being able

to see their data. Families in both my earlier work with families [the pilot study (Bowyer *et al.*, 2018), hereafter 'PS'] and CS1 were caused significant worries by not being able to see their data. People do not want to be treated like *subjects* (in either sense of one being subjugated [PS], or as a topic being discussed) and reducing people to a set of assertions in data causes them to become, in effect 'objects to be administered', which is harmful and disempowering [4.3.3; 4.4.4; Cornford, Baines and Wilson (2013)]. Supported families felt helpless and resigned to being judged through data and sometimes suspicious of those holding or using that data [4.4.4]. This led in some cases to withholding of information or distrust of support workers, harming the effectiveness of a relationship that is designed to empower [4.5.1]. In that same section (and in (Bowyer *et al.*, 2018, p. 7)) I outlined how ongoing individual access to data has the potential to transform attitudes, remove dependence and a feeling of being a subject, and could empower families to help themselves. Such fears and worries about unseen data were echoed in CS2, with participants exhibiting great unawareness of held data [Table 5.4, 5.4.3], and concerns over data being held out of their sight for long periods of time [5.5.3] as well as similar feelings of resignation or lack of choice [5.5.4; 5.7]. Denying access to held data was seen as a key source of holding power over individuals [5.5.4], and visibility of data is a key part of assessing *'to what extent the bargain'* (of data sacrifice for value as described above) *'is fair'* [2.1.4; 5.6.3; Larsson (2018)]. It is an inherent consequence of representing people through data and then using that data to make decisions [2.1.2; Cornford, Baines and Wilson (2013); PS] that individuals become sidelined and excluded [2.3.3; Crabtree and Mortier (2016)]. Without visibility of data, consent is not meaningful, and individual needs are more easily ignored or overlooked.

6.1.2 Understandable Data

Visibility and access to see data is not enough, people need to be able to interpret it. Data is only valuable in so much as it enables us to access the information which it encodes [2.1.1]. People need to be able to make sense of it. When humans look at data, we inevitably attempt to interpret it to see what it can tell them; in EH, support workers try to learn more about people's lives by examining data about them [4.3.3; 4.4.3]. In doing so they apply their own knowledge and expectations in an attempt to extract facts. Similarly in the context of everyday digital life data, individuals search for value and meaning in that data,

they reflect upon it and try to relate it to their own lives [5.5.3]. While EH staff receive training on how to understand families' data, individuals struggle to understand their data without sufficient support, as discussed above. Returned data from GDPR requests is often dry and technical. It may contain codes, internal notations or abbreviations that a layperson cannot understand [5.5.3]. Raw data is rarely sufficient to provide clear, unambiguous and unbiased information to the reader (Gitelman, 2013; Neff, 2013). In line with one of the three core principles of HDI, *legibility*, data should be understandable by those it concerns (Mortier *et al.*, 2014). In both Case Studies (and my prior work), individuals shared a desire to not just be aware of, but to *understand* what data was held about them and how it was used [PS; 4.4.2; 5.5.2]. People are only just beginning to understand the significance of a data-centric world that uses data to make decisions that affect their lives [2.1.2; 2.1.4; 5.6.3].

People need understandable summaries of information content and context. It was very clear from the findings of both Case Studies that all humans looking at data need *summaries* to help them digest and locate key information. In CS2 participants were often overwhelmed or *'drowning'* (P1) at the volume or technical complexity of the data returned from access requests, *'so much of it that's impossible to know what it all means'* (P4) [5.5.3]. These feelings were mirrored in CS1, by support workers who feared the liability of having to *'trawl through'* large volumes of data and know all the relevant and important facts about a family so that they do not make mistakes [4.4.2]. Participants on both sides talked of needing help to see the whole picture, something that is hard to achieve from individual datapoints or sets of files. In both cases, summaries of data would help comprehension. However, the task of creating a summary is not straightforward and places power in the hands of the summary-maker, who can decide what is relevant, how the data is framed and what is omitted from the summary. People look at information for different reasons, to answer different questions, so the question of who decides what is relevant or most important within a body of data is a critical one. Different summaries would be needed for different audiences. As Mortier reminds us, effective legibility requires a recognition that individuals' viewpoints of data can and should differ (Mortier *et al.*, 2014). There is a question about who decides what the viewer of a summary 'needs to know' [4.4.3; 4.4.4]. This is further complicated by the fact that the data itself is not neutral (Gitelman, 2013); in the EH context it was clear that opinions as well as facts are recorded (Bowyer *et al.*, 2018,

p.6 and from SILVER project), and a focus on the recording of data most helpful to the support worker. Commercial data holders record data in ways that are optimised for their existing systems and processes, as seen through the presence of internal codes, system screenshots and filenames in returned data [5.5.3]. In CS2 most participants' comments on returned data indicated that it had not been not presented in a way optimised for understanding [5.5.3], failing to support *sensemaking*;

> *"Information presentation should be as clear as possible so that people can interpret their data and extract meaningful information from it."*—Gurstein (2011)

Rather than raw data, people need information and visualisations, arranged and optimised for understanding. Data, by itself, is not meaningful. In order to be able to answer questions and acquire knowledge, people need information [2.1.1]. Access to raw data files or database records or spreadsheets does not satisfy this, and prior research in the civic data context states that this would be inadequate and limiting (Cornford, Baines and Wilson, 2013). To comprehend the meaning of data, visualisations and explanations can help [4.5.2]; as one support worker in CS1 observed, some families might find data tables too technical, *'I think sometimes it's easier to do it in pictures'* [4.4.2]. Participant-designed interfaces in CS1 included pie charts, graphs, spider diagrams and timelines, all designed to convey information more intuitively [4.4.2]. In CS2 and prior GDPR requests, it was often the case that companies often returned data not in understandable forms that were less useful than the apps or websites those service providers offer. For example, run-tracking apps such as Nike+ and Strava return route log information in XML-based TCX files which are meaningless without some analysis tool or visualisation. JSON files, a commonly returned data format, often use a timestamp format that is just a long number, not understandable by humans without extra work. Data was often returned in formats that were more a reflection of internal systems (e.g. screenshots, table dumps or exports) than being optimised for understanding [5.5.3], which some participants found useless. As P5 observed, *'It's like being given the bricks to a house… It doesn't really mean anything when it's just bricks, if you don't know how to put it together'* [5.5.3]. It is clear that visualisations are key to accelerating understanding (and are also subject to the same challenges of selection and bias as summaries). Furthermore, visualisations of data can function as powerful *boundary objects* and *things to think with* [PS; Brandt and Messeter (2004); 3.4.2; 4.5.2]. In

the care context the use of data representations as a focal point or evidence for was thought to be more productive and empowering [4.4.2; 4.5.1].

Information becomes most meaningful when it is recognisable and relatable and can be mapped back to life experiences. Across both case studies, there was a clear search for meaning in data, which manifested as a desire to use that data to build a fuller picture of the individual (or family)'s life. In CS1 and in SILVER, it was evident that support workers seek a broader view into supported families' lives by reviewing data about them [4.4.2; 4.4.3]. Both supported families and staff saw value in seeing all the data about each individual in a common place - in other words, structuring the information around the individual's life, rather than the information silos of different agencies [4.4.2]. In CS2, where participants were asked about the value they saw in the data companies held about them [5.4.3], it was clear that data was most valuable to participants when it was recognisable and relatable to events in their lives. This is particularly important for data that participants have never seen before such as derived, acquired or metadata - without a way to connect it to one's own life, it is impossible to relate to [5.4.3]. Echoing goals surfaced in personal informatics literature (Li, Forlizzi and Dey, 2010), participants sought insights about themselves in data, and so valued data that spanned a longer time period such that they might use it to spot patterns in relation to events in their lives [5.5.3]. They valued the opportunity to use data as *'a window into the past'* [P11 in 5.5.3]. Long term data was seen as a liability [5.5.3], and when it came to considering attitudes to providers, participants were most concerned about data collection by the larger data companies like Google and Facebook who, through their myriad apps, websites, devices and other means, had many touchpoints into their lives and thus had a broader picture of their life activity [5.5.3; 5.5.4]. It is clear that to transform dry, technical data into meaningful information, it should be contextualised in relation to events in the life of the individual it describes.

In practice, ongoing human support is needed to facilitate the understanding of data. In both case studies, participants felt that they had questions to ask of data holders, especially when the data was difficult to understand but also more generally when the answer to a given question could not be found in the data. In PS and CS1, participants emphasised the ability to talk to someone about their data [4.4.2], and included such features in their designs - both to understand and receive help in understanding, but also to

feedback and provide additional information, explanation or context so that they might be better understood [4.4.3]. In some cases, families needed additional support for reasons of accessibility or technical literacy. Human communication channels for data support need to be available at convenient times too; in CS1 individuals felt constrained by needing to wait until their next meeting with a support worker, and wanted a communication channel they could use in their own time. In CS2, participants were regularly frustrated by data they could not fully understand and could not ask about: sometimes literally, when internal codes were not explained or technical formats were used, and sometimes when they could not understand how the information had been derived, such as insurer's driving scores [5.5.3] or Instagram's inferred interests. The general pattern of GDPR handling by organisations was to deliver data, often handled by a back-end team with no customer face or means to ask follow-up questions. Where questions were asked, they were typically hindered by delays and middlemen preventing an effective conversation [5.4.3]. Considering the systemic changes toward data-centricity that the world has undergone, as described in 2.1.2 and 2.2.4, it is no surprise that human support has reduced. Across both studies we see the cost of that shift toward dealing with data instead of dealing with people; individuals get left behind, without the means to understand or ask about their data.

If data is not understandable, distrust can arise. In both Case Studies, the costs of individuals not being able to understand their data and being left in an unsatisfactory position of being unable to resolve concerns or ask questions is evident: Without understanding, comes distrust. In CS1, participants were concerned when they saw assertions on their records that they disputed or could not identify the source of. In CS2, privacy policies that were too vague in their explanations of data (such as Google's) caused participant distrust because they seemed broad and lacked tangible examples, and large volumes of technical data caused suspicion. It is clear that understandable information about what is represented in data, and about the context and use of that data, can help individuals to trust the data holder. In the minority of cases in CS2 where the GDPR had a positive impact, the reasons given included understandable data and helpful human responses, for example P10 cited Niantic's detailed data and positive attitude, while P6 described Sunderland AFC as upfront, and said that their data *'just made sense'*. P7 specifically distrusted LinkedIn because she felt they had not bothered to adequately

engage with her [5.5.4]. The importance of trust will be discussed in more detail in [6.2.1] below.

6.1.3 Useable[11] Data

People need to be able to explore and interrogate data to ask questions of it. In both case studies, many participants showed they are aware that their personal data contains insights and value they cannot access. In CS1 and PS this manifested as concern over what unseen incriminating judgements might be storied in their records, and an awareness that the only data supported families can actually see is whatever subset of data their support worker chooses to share with them [4.4.4]. In CS2, this manifested as feelings of not having the technical skills necessary to explore their returned data [5.6.3] and that without better tools individuals were not getting the same view of data that service provider staff have [P2 in 5.5.3]. What we can see from these observations is that even visible, understandable data is not enough to meet people's needs. People need to be able to interact with their data, in order to explore it, gain insights or answer questions. There is a need for tools not only to access data, but to help people make sense of (often technically formatted) data [5.5.3], and explore it in different ways to answer questions - which implies the need for operations such as filtering, searching, comparing and orienteering (see 2.1.4 for others), in order to understand history, context and patterns in the data, as suggested in self informatics (SI) literature [2.2.3]. Several participants in CS2 hoped to receive data in formats they could visualise, *'mash up'* and play with, but did not get this - not only was data not optimally

---

[11] The words *'usability'* and *'usable'* (spelt without an 'e') most commonly refer to a judgement of the degree to which a website or user interface is easy to use (Nielsen, 2012). Throughout this thesis, I deliberately use the alternative word spellings of *'useability'* and *'useable'* (Collins English Dictionary, no date a, no date b) respectively, to clearly distinguish from this ease-of-use concept and to denote that I am referring a different meaning: the more literal definition, i.e. *'the quality or state of being convenient and practicable for use'* (Merriam-Webster Dictionary, no date a, no date b). Any usages without an 'e' can be taken to refer to the interface ease-of-use concept.

formatted for such use, it lacked sufficient explanation to enable individuals to pursue such goals for themselves [5.5.3].

Data needs to be useable - correctly formatted and explained in a portable and standardised form. As discussed in 5.7 and 5.5.3, people have multiple needs around data: For understanding, people need *understandable information*, both the facts and assertions encoded within the data, but also information about the data itself - its context, history, use, and significance. But there is a distinct need for *useable data*, which is somewhat orthogonal to this. A PDF containing a screenshot from an internal system might potentially be sufficient for understanding, but is useless for exploration or visualisation-building. Similarly a technical log file might contain rich data that can be queried and visualised given the right tools, but without those tools or an informational summary, is of no immediate value [5.5.3]. This dilemma was alluded to by P4 in CS2 thus: *'If you want to view the data they have about you, it's quite useable. If you want to do something [analytical], then it's not'* [5.5.3]. This dichotomy of needs is discussed in 5.6.1, where the introduction of standard formats is proposed as a means to catalyse the building of data insight interfaces. As P1 stated, *'it would be nice if these companies had a standardised model of how this information is presented to people'* [5.5.3].

People need to be able to interact with data, which means interfaces are needed. By themselves, even standardised files as described above are not practically useable. As discussed in 4.5.2 and consistent with effective access (Gurstein, 2011) participant data designs in CS2 remind us that simply providing data is not sufficient: to be meaningfully able to act upon data requires some form of interface not only for visualisation and interrogation as mentioned above, but also so that any physical, cognitive or accessibility needs can be met. In CS2, some participants wanted not just for data access, but for tools to help them find insights from their data [5.6.3; 5.7]. While several companies are starting to create interfaces for data access, most of these are still focussed upon file delivery, with the notable exception of Google Timeline and Google My Activity, which provide a glimpse how an interface to explore data could be more useful than providing a bundle of files. Both of these examples also re-iterate the value of unifying data around an individual's life, as discussed in 4.4.2 and above in 6.1.2 above.

Data needs to be explorable from a temporal perspective. Another aspect of useable data that goes beyond what a data file can offer is the ability to view it over time. The importance of this temporal capability, as identified in literature cited in 2.1.4 (practical information access) and 2.2.2 (temporal PIM systems), and my prior writing (Bowyer, 2011), was evident in both Case Studies. In SILVER and in CS1, being able to access historical data for a full picture regularly surfaced as a desire in discussions - even though the exact bounds and mechanisms for achieving this were contentious [4.3.5; 4.4.2; 4.4.3]. In CS2, as Table 5.4 shows, 26% of participant goals related to SI-type reflection [2.2.3] on one's past to enable self-insights, nostalgia and creative uses of data. No participant was able to achieve this, and this was in large part due to the lack of temporal data exploration capability, meaning that significant potential value (the value of a long-term dataset as described above) remained locked away and inaccessible [5.4; 5.7]. CS2 also suggested a lack of thought to this temporal perspective from data holders, who delivered data as a one-off snapshot that was already out-of-date when delivered. Indeed the GDPR explicitly discourages excessive data requests, rendering an ongoing view of data as proposed in 5.6.1 near impossible. Having better temporal data exploration capabilities would enable people to understand themselves and their data ecosystems better, informing both personal self-improvement goals and better decisions about personal data practices and provider choices [5.7]. Other exploration perspectives that could be powerful include location-based views or person/company/relationship-based views.

People need to be able to access the valuable knowledge locked within their data. People need to see the complete picture of their data in order that they might access at least glean as much value and insight from it as data holders do [5.5.3]. For self-improvement and improving one's situation (a key goal of EH), access to metrics visible in data are extremely important, so that one might measure progress [4.4.2]. While support workers do endeavour to provide this and sometime show data extracts, this can never be as empowering as having full capabilities to explore that data *'in your own time, at your own pace'* [PS; 4.4.2]. In commercial service provider relationships, data interfaces present data in ways that are configured to reflect the profit motives of the organisation (explored further in my work with Goffe *et al.* (Goffe *et al.*, 2021)), and so accessing a relevant view of one's own data and having the means to access the knowledge within it is similarly difficult. Participants in CS2 found returned data unuseable [5.5.3], and wanted visualisations that

would allow them to discover patterns and insights, and tools to explore their data [5.5.3; 5.6.3]. If data holders can make data not just visible but useable, individuals can learn about the valuable knowledge that data holders are currently extracting from their data, and hopefully how to access that knowledge for themselves.

If holders do not make data useable, this is a barrier to individual agency and power. In Human-Data Interaction terms, people need not just legibility, but *agency* - the ability to act upon one's data (Mortier *et al.*, 2014). Personal data contains valuable and actionable information about individuals and their lives [5.6.3]. The SI field has identified that there are already many practical barriers to working with one's personal data effectively, including not just access but challenges in integration, sensemaking and goal-tracking—the 'barriers cascade' (Li, Forlizzi and Dey, 2010)—which hinder the ability to use one's data for personal benefit. As observed in CS2, an inability to access the value in your own data can lead to feelings of resignation, concern, suspicion or distrust [5.7]. Even if one can see and understand one's data, an inability to act upon it can reinforce feelings of being passive and uninvolved; without this, any opportunity to feel engaged and motivated is lost (Bowyer *et al.*, 2018, p. p8). Being able to use data for one's own purposes is a critical ingredient of empowerment and rebalancing power (Hoffman, 2014a). Without data usability, individuals are in effect digitally impaired, leading a less functional society where innovation and growth is limited [Abiteboul, André and Kaplan (2015); 2.1.4].

6.2 Answering RQ2: What do people want in *indirect* data relations?

By comparing and grouping elements of the findings from CS1 [4.4] and from CS2 [5.5], especially from an individualist perspective of individuals' relationships with care providers and digital service providers respectively, three distinct data wants are evident when considering *indirect* data relations:

1) **Process Transparency**: People need to know what data is being collected or held, and critically how it is being used, for accountability and safety and in order to have trust in data holders;
2) **Individual Oversight**: People need the ability to affect what data is held and how it is used, including reacting to changing circumstances, deleting data or withdrawing consent for certain uses; and

3) **Involvement in Decision Making**: People need to be invited and involved in decision making based upon their data, so that they are not misrepresented and their needs are not overlooked. This can be aided by collaborative use of data, giving individuals a human point of contact, and consulting the person not just the record.

These wants are detailed in the following sections:

6.2.1 Process Transparency

People need a window into how their data is used; this means transparency of processes not just of data. It is well-established that there is currently extensive use of personal data by service providers and other parties that is beyond an individual's view (Hoffman, 2011), forming an ecosystem of data use based upon one's data, which is currently not centred on or visible to the individual concerned [2.3.4]. Decisions made based upon personal data directly affect people's lives through policy decisions (in the care context) or business/functionality decisions (in the commercial context). People need to understand the value created by the use of their data and how (if at all) they are compensated for this (Hoffman, 2011). Even with full data access, understanding and useability, individuals cannot see into this opaque world of data use; the data is just an artefact produced and shaped by unseen processes. Like an archaeologist trying to infer the customs of lost civilisations through ancient relics, observation of the data can only reveal so much. Andrew Cormack, writing before GDPR, observed that *'it is more important to know how information is processed than the actual values involved'* (Cormack, 2016). The SILVER project found that families had very little awareness or understanding how their data was used, and that consent was therefore not meaningful because consent had been given without processual understanding [4.3.2]. In CS1 participants agreed that people need rights to see how their data is used [4.3.5]. CS2 revealed a clear desire for awareness of how data is used, how decisions are made, and how this might affect them [5.5.2], with over 74% of goals in pursuing GDPR requests relating to wanting greater insight into personal data use practices [5.4.2]. 70% of participants wanted to understand what providers infer from their data and this was unmet in 73% of cases and fully met in only 7% of cases [5.4.2].

Process transparency is required to enable accountability. In CS2, participants recognised that organisations had collected data about them which could be exploited, and wanted to understand the extent of that capability [5.5.3]. Data access can provide a window into collection capability, but only process transparency can reveal the extent of data use capability. Many participants expressed a desire to assess the trustworthiness of their service providers; they had curiosity, suspicion and unanswered questions that only transparency could address [5.4.2] and sought to judge whether data use practices were "appropriate" [5.2.4]. In the EH context, all data processing is hidden from individual view and no access or questioning capability except through their support worker (functioning as a selective gatekeeper) [4.1.1; 4.3.1; 4.5.1]. Returning to CS2, there was evidently some transparency available in the form of the ability to make a GDPR request, but many participants found GDPR responses inadequate for holding providers accountable [5.2.4]. Nonetheless, data access request handling is itself a data process, and so in this sense, the GDPR process did offer some ability to judge the trustworthiness and integrity of providers in data handling, in part informed by the breadth and quality of data returns [Table ARI5.1] but perhaps more by the experience of the GDPR process. Many participants formed or revise their perceptions of companies, with perceptions of providers having a lack of care and making access difficult, or of providers being helpful and open having a strong impact on participants' attitudes toward them [5.5.4].

There is no accountability, processes are not transparent, and thus power remains imbalanced. Across both case studies, the lack of process transparency is clear. EH services have no obligation to describe or share their data use practices with supported families, and apparently only even attempt to do so at the point of initial onboarding and consent collection [PS; 4.3]. This equates in practice to a complete lack of accountability over data practice [4.5.1; 4.6]. Meanwhile in the commercial context, some companies failed to respond at all to GDPR, which is a barrier to accountability service providers. Many routinely failed to adequately meet the transparency rights stated by GDPR, without repercussion or consequence [5.5.2]. In both sectors, data holders' freedom to collect and use data without adequate transparency or ability to be held to account can be seen as an exertion of power over individuals. The power imbalance (Hoffman (2011), Hoffman (2014a)) and the dominance of data holders over the individuals about whom data is held, is reinforced by a lack of transparency.

People face an incomplete picture of their data ecosystem, even after using all available means to achieve transparency. In SILVER and PS, supported families expressed concern that data could not adequately represent the complexity of their lives, a view supported by literature (Cornford, Baines and Wilson (2013); Gitelman (2013)) and sustained by support workers & staff in CS1 [4.3.5; 4.4.3]. Therefore, transparency is vital in order to ensure data is fair and accurate to them, yet it is not available [4.4.4] so they have no means to ensure this. Similarly in CS2, the most popular GDPR goals around understanding inferences made from data about people [5.4.3] remained unmet. Data was incomplete, delayed, or inaccessible [5.4.2] and the potentially most informative type of data when it comes to understanding processing—metadata, derived and acquired data—were typically absent. Apparently broad responses were discovered to be very limited when viewed through the lens of privacy policy commitments and GDPR transparency rights [5.4.4].

Trust in data holders is needed, and gaps in transparency create distrust and a risk of broken expectations, harming relations. Individuals need a functional understanding of their data and its handling, and this is crucial to trust. Good explanations (as were often found lacking in CS2 [5.5.3] can deliver some of this needed understanding and subsequently increase trust (Glavic *et al.*, 2021), as observed in a minority of cases (14%) where a good GDPR response led to the participant's trust in their provider increasing. Conversely we found that incomplete data (or a general lack of transparency/difficulties of access [5.5.4; 5.7]) can harm trust, as in the majority of cases (52%) can harm trust [5.4.4], leading to thoughts such as 'what are they hiding?' [5.5.4]. Privacy policies that contradicted expectations or lacked sufficient explanations also led to distrust [5.4.4]. Trust in the independence and integrity of data holders is essential (Dijck, 2014), and this was often a concern in the EH context, where trust between support worker and supported family is especially critical in order for the support relationship to be effective. Earlier work found that families wanted to be confident that their data would be handled sensitively and fairly only by those with a need to know, and believed that greater visibility of data processing would allow them to trust that that was the case [4.3.2; PS]. When families felt alienated from their data, trust was absent [4.5.1], and a lack of transparency and accountability makes it hard for families to maintain trust in the system [4.6]. These issues of transparency and trust are inherent in a data-centric operating model:

> *"A crisis of trust is developing, stemming from the use of personal data in ways that are inconsistent with individuals' preferences or expectations."*—World Economic Forum (Hoffman, 2014b)

Information facilitates trust; transparency therefore offers an opportunity to earn trust & improve relations. In both studies, the findings indicated that increased transparency from data holders / service providers would improve trust; in CS1 support workers and organisations were recommended to be as open as possible about data handling and sharing [4.4.4]. CS2 highlighted the potential benefits of increased consumer loyalty that greater transparency might bring [5.6.2], as well as the need for policymakers to legislate in favour of increasing individuals' understanding of data practices [5.6.1]. This is a level of transparency that goes beyond current GDPR practice, and even beyond current GDPR policy. In order to redistribute power, GDPR needs to deliver meaningful transparency, not just the *'box ticking'* delivery of unhelpful files that our participants sometimes observed [5.6.1]. Crabtree describes meaningful transparency by saying that it cannot be a *'one-way street'* that reduces individuals to *'being spectators'* on how their data is used; he says that it involves *'making the whole ecosystem transparent, not just the front end'* (Crabtree and Mortier, 2016). Access to good information about practices is the most effective way to earn trust [5.5.4], and both studies' findings suggest that a proactive attitude can do just that [4.5.1; 4.4.4; 5.6.2; 5.7].

Initially, transparency may cause distrust, but only where practice is problematic; this is accountability becoming real and catalysing better data practices. It is important to note that in CS2 we saw the transparency of GDPR cause elevated distrust [5.4.4], however this does not mean it should be avoided. The reasons cited for distrust arising were invariably due to the discovery of practices that participants did not approve of. This is a clear illustration of the link between transparency and accountability; the transparency reveals the non-consensual or unsatisfactory practices that providers must change if they wish to maintain trust and loyalty, such as unclear data practices, data over-use or data sharing that the individual would not have consented to had they been asked. This shows that in some cases trust is fragile, where unfavourable practices are hidden and only the individual's unawareness is keeping the relationship intact [5.5.4]. Data holders should not only be transparent, but should follow this up by acting upon subsequent feedback, improving practices that individuals discover and challenge [5.5.4]. By shining a

light, accountability becomes real and change for the better can occur. Ultimately, increasing transparency can help providers uncover exactly what they need to do to earn greater trust [5.7].

Without transparency of data and processes, individual action is blocked and power remains imbalanced. Across the two studies, a clear pattern emerges: transparency can increase trust, enable accountability, empower individuals, and (provided organisations respond favourably) actually tilt the power balance back toward an equitable and fair relationship where data is collected and used in clear sight of the individuals it concerns, where they might hold those organisations to account and immediately challenge any unsatisfactory practice, unauthorised processing/sharing or inaccurate data. Thanks to GDPR, individuals are now able to take direct action to educate themselves and pursue greater transparency, and utilise their rights to motivate incremental changes from data holders [5.6.3]. Without transparency, data holders will continue to hold the balance of power, and individuals will lack agency and accountability.

6.2.2 Individual Oversight

Data visibility and process transparency naturally leads to a desire for individual oversight. If you see something that is 'not right', you are motivated to want to fix it. And therefore, people want something more than data and process transparency, the natural next step is the ability to make decisions about what happens to their data. Participants' goals in CS2 included curiosity, suspicion and a desire to shed light on specific incidents [5.4.3], mirroring the desires families in PS exhibited to be able to know and see what data was held and used about them. In both cases, individuals wanted to have a say over what happens. Current models of informed consent have been found to be inadequate, with the initial handover of data acting as a *point of severance* (Luger and Rodden, 2013). This was echoed in the experiences of families in the EH context, who gave consent at the point of initial onboarding, but lost all ability to influence what happens to their data thereafter [PS; 4.3.2]. GDPR aims to adhere more to a *dynamic consent* model (Kaye *et al.*, 2015; Williams *et al.*, 2015) by giving people an ongoing set of rights, including the right to be informed about the use of your data, the right to object to certain data uses, and the right to get your data corrected or deleted (Information Commissioner's Office, 2018). In line with the *accountability principle* (Article 29 Data Protection Working Party, 2010; Crabtree and

Mortier, 2016), this in effect would allow people to act as overseers or regulators over their own data: watching how it is used, and demanding action or change to practice when they see data use that goes against their wishes.

People need agency and negotiability over held data about them, in order to ensure fairness and accuracy and reduce risk. As PS showed, there is a strong desire to ensure data is *fair* and *accurate*, because that data is used to inform judgements and make decisions that can directly affect the individuals concerned. Data in CS2 showed a clear problem with the accuracy of unseen data: while in 92% of cases volunteered data (which by definition, has been seen by the individual) was found accurate, derived and acquired data (previously unseen by the individual) was found inaccurate in 50% and 80% of cases respectively [5.4.2]. Being able to ensure fair and accurate data goes beyond being able to see and understand the data, but requires also *agency* (the ability to act within a data system, such as to delete or correct data, or withdraw consent) and *negotiability* (the ability to continue to have a voice and make changes as circumstances change) (Mortier *et al.*, 2014). People need a relationship with their data [2.1.5]. In both CSs and in PS, individuals perceived tangible risks both of data being held beyond their reach, but also of potentially inaccurate data being used to make decisions. Risk factors identified in the EH context included facilitating or encouraging crime, causing social and psychological harm, and enabling medical mismanagement or welfare support failures. In CS2, participants felt that held data about them that is not visible or controllable was a liability that might lead to privacy violations, commercial exploitation, and an increased risk of data leaks [5.5.3]. Clearly people feel that for their data to be safe, they must be able to see and verify its storage and use for themselves and enforce action when something is not right. As early as 1980, when the world was less data-centric, it was already recognised that individuals would need the ability to challenge data use, as the OECD observed in their guidelines:

> *"The right of individuals to access and challenge personal data is generally regarded as perhaps the most important privacy protection safeguard."*—Organisation for Economic Co-operation and Development (1980)

Individual oversight capabilities must be supported by governance, so that individuals can effect desired changes. As Gurstein notes, a key element of effective data access is governance, that is, mandating data holders to support individuals in accessing

their data and respecting their wishes over what should happen to that data (Gurstein, 2011). Individuals need to be able to give instructions, make changes and express permissions that have weight; they need to be listened to, so that they can meaningfully effect change [4.5.2]. Bakardjieva, examining the use of data about others in a different context (research), identified that individuals whose data is used need the ability to influence not only the data about them, but the actual decision making that occurs based on that data: both the data *and* the decision making should become objects that the individual subject can manipulate (Bakardjieva and Feenberg, 2001). At the time of writing (August 2022), much of the focus on GDPR has been about access to data, perhaps because this is more tangible, and very little about GDPR's other rights that can influence decision making [5.1.2]. This was backed up by participant experiences in CS2, where desires to influence or change practices or delete data were either not actionable or ineffective [5.5.3]. Governance over individual data rights has two elements. First, to support individuals in complaints or challenges, which are currently unevenly enforced [5.6.1]. But more importantly than this, given the extensive use of data by organisations and the great potential for misuse or harm, individuals need to be able to trust that systems are in place that mandate the behaviour of data holders to be trustworthy; to compel organisations to maintain good data practices such as data security and dynamic consent in the first place [5.5.4]. In the GDPR context bodies that can do this already exist–the Data Protection Authorities. In the public sector/care context, the picture is less clear. Participants identified a need for oversight bodies to compel good practice, identify appropriate access rules, and to provide independent oversight in contentious cases [4.4.4]; this is particularly difficult given that no organisation can see the full picture of an individual's civic data.

Individual oversight would bring individuals back to the centre of their personal data ecosystem as an active participant. As outlined in 2.3.4, the ideal has been established that individuals need to be at the centre of their own personal data ecosystem, overseeing and controlling their data selves as easily as their physical selves. This aligns with the concept of *individual self-determination*, which is gaining ground in PDE circles (Fisch, 2015). Currently, as seen in both contexts, data functions as a proxy for their direct involvement [PS; 5.5.4]. Decision makers consult data first, as the primary source of truth [4.1.2; 2.1.2], and the individual second (if at all). For transparency to be meaningful, data flows need to open up to include individuals as part of the loop, changing them from passive

spectators to *active participants* (Hoffman, 2014a; Crabtree and Mortier, 2016) in the processing of their data. Examples of specific oversight abilities desired in CS1 were the ability to explain or annotate datapoints [4.4.3], to be able to check data together with support workers, with the record of that check becoming part of the data [4.4.3], or to have granular access controls over precisely which data could be seen by whom [4.4.3]. In CS2, a clear picture emerged that what participants want is the ability to make *choices*. They want control over the data they are forced to sacrifice to companies [5.5.4]. To avoid the *point of severance*, data sacrifice should be a loan or sublicense, not a taking-possession-of.

Given the changing and complex nature of human life, data is inadequate and consent is never complete, so longitudinal participation and oversight is needed. Too often, data is treated as a static source of truth (see above). Attempting to represent people as data in order to require less human contact is a reasonable goal from an organisational efficiency or cost-saving perspective, but any representation will never be complete or adequate [Cornford, Baines and Wilson (2013); PS]. In CS1, the findings showed the need for numerous efforts to augment data in order to combat its inherent inadequacy, such as support workers seeking to understand the people behind the records [4.4.3] and maintaining a constant attitude of seeking to understand more deeply than the data record can allow [4.4.4]. Even if a data record can be corrected or completed, it will still be inadequate, because human lives change continuously: people move, start and end relationships and jobs, marry, divorce, have children, pursue new interests, become incapacitated, or die. The passage of time can radically change the context or relevance of data [PS]. A one-time informed consent upon data collection is inadequate in this ever-changing context [4.3.2; 4.3.5; 4.4.3]. If consent needs to be ongoing (and in order for it to be meaningful) this means that engagement with the individual concerned, and that individual having a view of their data and its use, need to be ongoing too. In order to avoid storing or using data beyond its need, ongoing data access is needed, in order to enable ongoing individual oversight [4.3.2; 4.4.2]. Systems and processes must treat data as dynamic [4.4.3], as something that will become inaccurate without sustained engagement. CS2 noted that GDPR data access is currently based upon viewing a one-time snapshot of your data, and does not take this need for negotiability (Mortier *et al.*, 2014) into account at all [5.6.1] (though some companies now offer download dashboards that come closer to providing ongoing access). Ongoing access, consent and participation do carry cost

implications for providers, and effort implications for individuals - but these can be improved over time: the former through automation, standards and education, and the latter through holistic approaches to personal data ecosystems; these mitigations are be explored further in Section IV.

There is scant individual oversight available today. Governance is lacking. If people cannot make choices about their data, they will remain powerless. Participants in both contexts faced an inability to see the full picture of how their data is processed and used. Families in CS1 workshops spent time designing interfaces for seeing and correcting their data and changing permissions [3.4.3; 4.4.2; 4.4.2], but no such interfaces exist. The entirety of their data access and influence is limited to what can be achieved verbally with their support workers [4.3.4; 4.5.3]. Without transparency and dynamic consent mechanisms, those families lack accountability. They are excluded with no ability to oversee or participate in the life of their data. In CS2, of the 41% of participant goals that concerned gaining insight into and control over the use of their data, 66% were unmet. Participants saw no clear pathway to use their rights to control over data. Only 1 of the 10 cases where a participant wanted to delete their data was successful [5.4.3]. At the time of writing, only one company (Apple) has a privacy hub that offers clear routes to access data rights other than access. Participants also reported in some cases being unable to check the accuracy of their data, or to investigate specific incidents where they had concerns [5.4.3]. The general view was one of widespread disappointment, that despite the promise of GDPR it did not confer any power to the individual to influence data use [5.2.4], leading in some cases to a reluctance to submit GDPR requests in future. Access requests were also rarely seen as useful in the care context. They typically only occur in the case of complaints. In the GDPR context, the inability to restrict data use or delete data was seen as a lack of control, and the retaining of data against their wishes as a liability [5.5.3]. Ultimately oversight means having choices, which is essential in the data-centric world. The Case Studies' findings show that, in general, participants felt they had been forced to sacrifice data to access services, and offered *no* practical choices or control over that data. Without individual oversight, there is no choice and people remain powerless.

6.2.3 Involvement in Decision Making

Data represents people. But people are more than can be encapsulated in records. There is a need to engage the human behind the data, as people can never be fully represented in data. Intrinsic in the move towards data-centricity has been a move away from human involvement. In the commercial sector this is due to cost-saving (call centres and web portals being cheaper than individual customer interactions) [2.1.2]. In the care context it is similar but there is also a desire to create a society that functions at large without individuals requiring special handling and support [4.1.1]. Both case studies' findings, consistent with literature (Abiteboul, André and Kaplan, 2015; Crabtree and Mortier, 2016), reveal myriad problems created by the exclusion [PS] of people from matters that affect them - from feelings of alienation or disengagement [5.5.4] to actual harms caused by erroneous or unfair judgements [4.3.2]. Service providers holding data need to contextualise data as an incomplete view into the complex human world, and seek greater understanding [4.4.4] while looking for positives in data [4.4.4]. It is interesting to note that the Troubled Families programme was created help find the human situations of people slipping through the cracks of the system. This highlights the inadequacy of purely data-based decision making [4.1.1].

Consent to access and use data needs to be dynamic and meaningful, which can only happen through ongoing involvement. As established in 4.5.1 and 6.2.2, ongoing data consent is essential, and this is especially important where that data is used to make decisions [4.5.3]. One-time consent is ineffective and meaningless [4.6]. Asking individuals for consent and subsequently less involved in decision making reinforces a hierarchical, rather than an equitable, power relationship (Bakardjieva and Feenberg, 2001). Without ongoing consent, the power imbalance is amplified [4.5]. In the commercial context, companies view data as their asset to exploit (Hoffman, 2011; Toonders, 2014), and the simple fact of having the ability to collect or access to data about people has proved in practice, sufficient to enable a variety of practices which would be likely to be refused consent if made visible to users (Melendez and Pasternack, 2019; Evans, 2021; Claburn, 2021). Individuals feel forced into a one-sided arrangement of sacrificing data in exchange for service benefits; with no choice upfront on signup, and minimal practical choices afterwards, their only choice is Hobson's choice ('Hobson's Choice', no date). Consent has

become commoditised, and from a corporate perspective the focus has become constructing a *legal justification* for using an individual's data rather than practically engaging with them and verifying if they approve (Woods, 2022). This can only happen when the individual about which data is held is excluded from data handling processes.

A human channel for conversation is wanted, to enable explanations, questions, and consultation. In PS and CS1, all participants viewed that individuals should be able to talk to someone about their data [4.3.5; 4.4.2], in order to ask questions or explain datapoints. In CS2, participants had questions about their data that they wanted to answer [5.4.3], yet these questions remained unanswered [5.5.2]. Participants regularly experienced painful and ineffective processes when trying to answer their questions, found that GDPR responses, often unhelpful, provided no backchannel for follow-up questions or further communication [5.5.2]. They were left *'in the dark'*. This highlights the need for a human support channel, which is not mandated by GDPR, not just to better understand the data itself [6.1.2], but to enable ongoing consent negotiation within the relationship with data holders.

Individuals should be consulted in decision making. This improves accuracy, perspective, and fairness and reduces consent liability. A common theme in the findings of both PS and CS1 was the idea that data is more likely to be fair and accurate if in the individual has the opportunity to express their perspective on it. Families advocated checking data together to identify gaps [4.4.3], valued the prospect of making their own contributions to the data to *'tell their own story'* [4.4.3] and wanted a *'right to explain'* or annotate their data [PS]. In a PDE context, this idea has been taken further by *Self-Sovereign Identity (SSI)* (Hori, 2021) which explores the idea of people owning and creating their own identity data. In CS1, support workers recognised the need to work with families to understand their situation better [4.4.2], rather than just look at data. These findings show it is important to give the human data subject a role in data creation or checking, and that if they are involved, to ensure a fairer and more complete view can be obtained than the limited view presented by the data record [4.4.3]. Without involvement, agency (Mortier *et al.*, 2014) will always be limited. Data is not neutral (Gitelman, 2013; Neff, 2013), and this means all stakeholders should be given a role (Bowker, 2005) in order to avoid errors, harm or disempowerment [4.5.1]. Involvement is negligible in the care context (being

limited to the TAF and the support worker relationship). In the commercial context, there is no such involvement, and the results of this can be seen in the low accuracy of previously unseen data [5.4.3] and in the low trust ratings [5.4.4] given to many providers after seeing data returns. It seems that data and knowledge would become more accurate when it is closer the data is to the individuals concerned [4.5.3; 4.6]. Data created and handled far from the family would intuitively be less accurate and less likely to have been considered from their perspective [4.3.2]. Organisations on both sides value data accuracy [4.3.3; 2.1.2] so greater involvement could help achieve this. Data holders would also benefit from involving individuals because responsibility for consent would become shared, resulting in a higher 'buy-in' from individuals and a reduced liability. Provided communication is effective and without barriers, involved individuals would be inclined to speak up if they see something they do not agree with [4.5.1]. Of course, this call for involvement in decision making is not absolute. Businesses need to be free to exercise their expertise. Patients are not best placed to decide what medicines they need, and over-involvement of users in product design can result in mediocre products that suffer from *design by committee*. In fact, the attitude conveyed toward the individual is critical, This was borne out in CS2 findings [5.5.4; 5.5.4]. Edwards and Elwyn, in their paper on shared decision making, argue that *feeling* involved is actually more important that actually contributing to decisions (Edwards and Elwyn, 2006).

Effective collaboration can be achieved by coming together around the data, using it as evidence (of facts or of opinions) and as a boundary object. Workshop C in CS1 specifically explored the prospects of shared data interaction, which had emerged from SILVER and phase 1 (Workshop A/B [4.2.3]) findings as a possible way to meet the needs of both supported families and support workers. Shared values were identified [4.3.5] and a model for shared data interaction builds upon the findings [4.5.3]. Central to this is the idea that evidence-based decision making can be more effective (both from a rapport-building perspective but also in terms of the quality of decision made) as guidance has advised [OFSTED (2015); Department for Education (2018); 4.1.2] and as seen in the findings [4.4.2; 4.4.3; 4.5.1]. Families and staff saw potential benefits from checking data together [4.4.3], using specific datapoints in discussions as evidence [4.4.2], or to help families open up [4.5.2], on top of the simple benefits of making all evidence equally visible [4.4.2; 6.1.1]. Throughout such shared data interactions, representations of data perform an important

function as a boundary object (Star, 2010; Bowker *et al.*, 2015); it provides a common focus to discussions that is relatable to both parties, *things to think with* (Papert, 1980; Brandt and Messeter, 2004). This helps improve legibility by surfacing the (perfectly valid) differences in different parties' perspectives (6.1.2, Mortier *et al.*, 2014). In PS I observed that gathering around data representations facilitated a less confrontational interaction than an across-the-table interview would.

Being involved means being able to learn and take action at any time, including on one's own and away from official contact or interactions with service representatives. Both Case Studies reveal that people want an ongoing window into their data and its use. Limiting data access and process transparency to specific times spent with a gatekeeper, as in the EH case [4.1.1], or to a process that only provides a one-time snapshot of data [5.6.1] reinforces the hierarchical power imbalance. Giving access to data and abilities to explore and ask questions *'in your own time'* [4.4.2] can unlock new individual benefits from data [5.5.3; 6.1.3] and provide more immediate feedback that can help people measure their progress towards improvement goals [2.2.3; 4.5.2; 4.5.3], as well as enabling ongoing individual oversight [4.5.2] and dynamic consent [4.4.4]. In the face of providers making decisions based on unseen data using processes that cannot be observed, people feel excluded and powerless [5.5.4; 6.1.1]. To ensure effective access (Gurstein, 2011), accountability and trust, there must be ongoing involvement, data access and transparency [5.6.1; 6.2.2].

Indirect data use enforces an uneasy trust; services that use data need a human face or point of contact, in order to grow understanding, earn trust and improve relations. There is a coldness to data. Facts, judgments and mistakes appear in print with equal weight, without explanation or context, seemingly a complete set of objective facts. In both studies, the value of human contact accompanying data access was evident; in CS2 participants valued the GDPR responses that felt most human [5.5.4] and lamented the inability to discuss data or resolve questions [5.4.3; 5.5.2]. In PS participants wanted *'to have a conversation'* about their data and in CS1 participant data interface designs included buttons to chat to their support worker or ask questions [4.4.2]. People do not want to be severed or alienated from their data, yet they feel they have no choice but to relinquish access and involvement [4.5.2; 5.5.4]. Ideally, all data-using organisations would have a

human face or point of contact that individuals can address questions to and in whom their trust of the organisation can be embodied [4.5.2].

Without involvement, people cannot take a full and equitable role in processes that affect their life. When data is used by organisations, this inherently serves as a proxy for their involvement [PS; 5.5.4]. People have consequently lost control and agency (Crabtree and Mortier, 2016), creating a crisis of trust and a power imbalance (Hoffman, 2014a, 2014b). Exclusion from data handling and decision making inevitably reinforces a hierarchy, with the individual destined to have less say and influence over services that affect their daily lives. What people need, even more than rights to transparency over data, is for better decisions to be made where data is used (Edwards and Veale, 2017, 2018). Taken together, an ongoing involvement with the data held by service providers and the processes that use that data would change indirect data interaction [6.2] into direct data interaction [6.1] and transform the individual experience of service use. Of course such a shift is difficult and costly to offer, for both small and large companies—the data-centric world emerged in part as a means to reduce costly human interactions and facilitate large-scale scaling up to serve larger user bases—but what we see suggests that the balance has tipped too far towards exclusion of individuals, and that giving them even a limited role in reviewing data, consenting and contributing to decisions, and just being informed, could carry significant benefits for both individual empowerment and organisational reputation, and trust.

6.3 Summation: Empowering Individuals with Better Data Relations

In this chapter, the separate Case Study findings and insights from Chapters 4 and 5 have been synthesised to identify six specific capabilities that people need in their relationship with their personal data - both from the direct perspective of RQ1 [3.3.1], where people want for **visible, understandable and useable[11] data**, and from the indirect perspective of RQ2 [3.3.2], where people want for **process transparency, individual oversight and decision-making involvement**.

6.3.1 Empowerment

In line with the pragmatist, individualist outlook of this thesis [3.1], these six wants amount to **a desire for empowerment**, motivated by a desire to pursue one's own happiness and

self-interest. Empowerment is defined as *'the process of gaining freedom and power to do what you want or to control what happens to you'* (Cambridge Dictionary, no date). In essence, empowerment is achieved when the individual, in consideration of the question *'what can I do?'*, judges that they can do more. Power, it transpires, is a double-edged concept. The power imbalance over personal data [2.1.2] encompasses both *power to* but also the concept of *power over*. These are best considered as two sides of the same coin or two perspectives on the same set of facts: organisations have *social power* over individuals, that affects what both parties can do. An organisation having social power does not imply a desire to dominate. Impacts on the individual's agency are merely material effects of the power holder's attempts to influence the behaviour of individuals in pursuit of the power holder's own desired outcomes (Pansardi, 2012). Concepts of power are explored further in 8.4.1.

6.3.2 Disempowerment

Not only do these six wants lead to individual empowerment, there is a clear correlation. **The more that the six data wants can be addressed, the more empowering it is to individuals.** From the Case Studies it is clear that the status quo of the data-centric world [2.1] is that the lack of these six capabilities reduces individual capacity to act, individuals are **disempowered**. People are treated indirectly through data (Cornford, Baines and Wilson, 2013). They are excluded and not involved [4.6; 5.5.2; 5.5.3; 5.5.4]. In both Case Studies many of the findings are based on the opinions of participants as to what they believe *would* be desirable, preferable or more successful. In the case of data visibility [6.1.3] and process transparency [6.2.1], however, there is clear evidence that these *do* have an impact on individual's subjective assessment of their own empowerment: In CS2, 45% of cases saw people experience a change in power after examining privacy policies and scrutinising GDPR responses. 29% felt a decrease in power, and 17% an increase. Data visibility and process transparency seem to be key first steps towards empowerment, as they allow a more accurate assessment of one's own capability. In 52% of cases, CS2 participants felt more distrustful, having discovered unsatisfactory exertions of power by providers [5.4.4]. Once visibility and transparency have been achieved, it is clear that this should be followed soon after with understanding [6.1.1], agency [6.1.3; 6.2.2] and involvement [6.2.3]. Any one of the six wants can improve data relations for the individual,

but the combination of all six is likely to produce more than the sum of its parts—an empowered digital citizen.

6.3.3 Towards A Better Society

Considering the societal level, these six wants show **how society should be reconfigured to improve data relations**. Giving people a role in influencing the life of their own data is a key ingredient in and of more progressive digital citizenship (Bridle, 2016). Shifting data interaction interfaces and processes to a more human-centric [2.3] model where people are controllers at the centre of their own personal data ecosystem [2.3.4] would be progressive and transformative, and not without cost, education, deployment and uptake challenges. Nonetheless, participants in both Studies could easily imagine more human-centric interfaces and more empowering service interactions, and demanded those improvements. And it is possible. Just as some human-centric practices were beginning to emerge among support services in CS1 [4.4.1], so some companies targeted in CS2 are already beginning to move in this direction. Some offer privacy hubs and explanations of data practices, while others offer interfaces solely for accessing data and exerting data rights [5.1.1]. In both Case Studies, there was a clear demand for these initial forays by service providers to be expanded: be it through more pro-active data practices in the care context [4.4.4], or through new models of data involvement in the commercial context [5.6.2]. If the locus of decision making [4.5.3] could be shifted towards individuals through such reconfigurations of existing practices, this would give them a role to play as *agents in the life of their own data*. This would allow them to curate their own data self, the representation of them used in decision making, so that it is fair, accurate and representative [PS; 4.5.3; Hori (2021)]. The principles of involvement, effective access and shared data interaction could be applied in many domains - education, health, democracy and commerce, and an emphasis on individual sociotechnical capabilities is a useful mindset to apply to both business process design and data interface design.

Empowering individuals with better data relations should lead to a better future. Taken together, the pursuit of the six data wants allow us to envisage a new, data-empowered future for individuals, who would reap the benefits of being able to gain insights and feedback from their own data in real time, while also co-operating with service providers in the stewardship of their data and involved in decisions. They would have

agency, influence and negotiability in an ongoing manner. In this future, there are opportunities for organisations to reimagine customer relations and the role of data in service provision, leading to increased accuracy and consent, reduced liability, greater trust and loyalty [5.6.2], and potentially to capitalise on new demands for data insight tools [5.6.3]. The ideal is that individuals would have a *'permanently open two-way communication channel [with data holders], allowing a dialogue to take place'* (Bakardjieva and Feenberg, 2001), delivering dynamic consent and ongoing involvement. Such a change in thinking, moving away from legal box-ticking (be it capturing informed consent [4.1.2], or satisfying GDPR requests [5.6.1]) and towards delivering an ongoing understanding of the collection and use of one's data and a voice in what happens to it, would empower individuals and begin to move us towards that idealised future. In such a society, the positive impacts of transparency would be evident. The general principle that transparency allows trust to be earned (which can in turn benefit organisations themselves) could be applied in a wide range of service endeavours that are currently very closed. This would allow innovation in data access and process involvement to flourish.

The vision of these new empowered individuals, given new inclusive roles in the lives of their data, therefore, is the ultimate answer to the research question [2.4] this thesis initially set out to answer through its participatory track [3.2.2]. This data-empowered and process-involved future is what better data relations would look like. The next chapter frames these wants as a new research agenda, which underpins the embedded research track that has been running alongside the Case Studies. The focus for the remainder of this thesis will now shift to a more practical research question, exploring how the identified wants might be achieved in practice.

Chapter 7. Defining a New Research Agenda: Human Data Relations

"What drives and drags the world are not machines, but ideas."

—Victor Hugo
(19th century poet)

7.1 Foundations for Practical Exploration of Data Wants

The participatory research track of this PhD research [3.2.2], which has been the focus of the last three chapters of this thesis, sought to discover what people want from data, and from relations that involve or use their data. We now know, from findings common to both Case Studies, six 'wants' that people have in their relationships with data: *visible*, *understandable* and *useable data*; *process transparency*, *individual oversight* and *decision-making involvement* [Chapter 6].

It is clear, given the complex sociotechnical and power-imbalanced reality of today's data-centric world [2.1.2; 2.3.3], that if positive change towards these wants is to be pursued, there is more to understand. We need to learn how people can design pragmatic, viable approaches to effect change. There is an opportunity to move from qualitative research into activist research and adversarial design [3.2; DiSalvo (2012)]. This has been the focus of the parallel, embedded research track of this PhD—to go beyond the RQs [3.3.3] and explore how people's relationships with data might be changed in practice.

The Case Studies prioritised a participatory and investigatory approach, but there is a need for specialist design innovation that cannot always arise from working with everyday users. The approach for this track, therefore, is more UCD than PD [3.2.1]. The established findings become material for myself as an activist researcher and adversarial designer, informing ideas for technical and societal changes that can bring about better data relations. Bringing to bear my experience as a designer and software developer, I can advance the exploration of this problem space farther than a participatory researcher could. It has been my good fortune to have been embedded, over the duration of this PhD, in various research, design

and development capacities, on four practical projects, that can be seen to relate very much to the pursuit of personal empowerment and human-centric data that I and the participants have envisaged.

Before explaining the different projects undertaken (which will be done in Section IV), it is important to establish a more precise framing for this track of the research. In short, the RQ for the practical track of the PhD can be considered as:

> **"Having understood what relationship people want with their personal data, how might these better data relations be achieved?"**

When looking for an existing discipline around which to frame the exploration of this question, Mortier et al.'s Human-Data Interaction (HDI) (Mortier *et al.*, 2013, 2014) is an obvious place to start; its principles relate closely to the wants uncovered in Chapter 6:

- HDI's *legibility* is very similar to the want for *understandable data*, and in part highlights the want for *process transparency*.
- HDI's *negotiability* could be seen to imply both the want for *individual oversight* and *decision making involvement*.
- Along with that want for involvement, the want for *useable data* could also be mapped onto HDI's concept of *agency* (and to Gurstein's *effective access* as well).

While HDI is clearly a foundational school of thought that underpins this work [2.3.2] and aligns with its findings about what people need from data, its value in finding out how we might pursue those goals is limited. To explain why this is the case, it is useful to consider again the perspective of information theory introduced in 2.1.1, specifically the difference between data and information.

HDI focuses on the way people interact with data, in a very general sense. Mortier *et al.* defined the field of HDI without making the important distinction between data (the digital artefact stored on computer) and information (the facts or assertions available by interpreting that data, or facts about the artefact of data), and their conception of data interaction seems to encompass both. This was a significant oversight. As further evidence for the importance of making that distinction, it should be noted that *Human Information Interaction (HII)* originated in library sciences to consider how humans relate to information without regard to the technologies involved (Marchionini, 2008).

Coincidentally, Jones *et al.* had around the same time called for a new sub-field of HII in an HCI context[12], highlighting the need to focus on information interaction:

> *"*[HCI can] *unduly focus attention on the computer when, for most people, the computer is a means to an end—the effective use of information."*—Jones *et al.* (2006)

DIKW theory [2.1.1] highlights that interpretation of data to obtain information is a discrete activity. This was borne out in the findings of Case Study Two, where it became clear that participants have distinct wants from data, and from information [5.5.4.2]. Access to data *and* information is critical to both understanding and useability[11] [6.1.2; 6.1.3].

Drawing on DIKW theory with consideration to these findings allows the identification of three distinct things that people can have relations with:

- **data** - the stored digital artefacts held by organisations for algorithmic processing and human reference, copies of some of which can be obtained using data rights;
- **information about individuals** (which I will call *life information*) - facts and assertions about the individual and their life, obtained through human or algorithmic interpretation of stored data or analytical inference; and
- **information about data** (a.k.a. *metadata* [Table 5.2; 5.4.1] or what I will call *ecosystem information*) - stored facts about data, such as storage location, access history, means of collection, contextual meaning, or sharing records.

These new terms introduced here of *life information* (information **within** the data, that says something about the individual) and *ecoystem information* (information **about** the data, where it is held and how it is used) are critical to the subsequent sections of this thesis, as each must be considered separately. These concepts are unpacked more in 7.4.

By differentiating and highlighting the different wants around these concepts, this thesis enriches existing understandings of HDI. There is however another aspect of HDI that makes it insufficient for our needs: it offers understanding of the problems of today's data

---

[12] The HCI panelists involved (excepting Fidel) were seemingly unaware of the existing HII field in library sciences, as they positioned the publication as a call for a 'new field'.

relations, but it does not go far enough in exploring, accepting and confronting the real-world barriers to agency, legibility and negotiability that exist. A number of researchers within the HDI field have identified the need to develop HDI into something broader: Crabtree and Mortier wrote about the need to examine HDI in a relational sense in 2015 (Crabtree and Mortier, 2015), and in 2016 identified the need to examine the issues of diminishing user control as being worthy of multidisciplinary study (Crabtree and Mortier, 2016). Deborah Lupton acknowledges the social entangled nature of personal data in (Lupton, 2016) which begins to explore HDI more through a societal lens. Sailaja *et al.* outlined the need to embrace HDI from a practical and grounded designerly stance, asserting that *"the emergent nature of [HDI] calls for refinement of the theoretical tenets [of legibility, agency and negotiability] to help them translate into practical and tangible responses that are embedded in the technologies we create"* (Sailaja *et al.*, 2021).

As a response to and a formalisation of these calls to extend HDI, this thesis proposes a new research agenda to elevate HDI to the sociotechnical level. From the individual data needs of HDI and of the Case Studies it defines a clear, actionable model against which efforts to operationalise HDI and other data wants can be aligned. This new agenda provides a framing that embraces and confronts the practical realities of pursuing better data interaction capabilities and better relationships that involve data. This new research agenda is named *Human Data Relations (HDR)* and is explained below.

7.2 'Human Data Relations': A Definition

The section above shows there is a need to distinguish this wider focus from HDI as conceived such that design and research work can be appropriately focused. The world of humans and machines can in effect be designed for at three levels, each with a broader and more human-centric focus than the last:

- *Human-Computer Interaction (HCI)* focuses on the interaction between people and machines.
- *Human-Data Interaction (HDI)* focuses on the interaction between people and the data stored within those machines.
- *Human Data Relations (HDR)* focuses on the *relationships* that people have with that data and the relationships people have with those who hold such data about them.

Building upon HDI but embracing the sociotechnical, HDR encompasses all the ways people and organisations can and should relate to and through data, not just direct data interaction. Through its greater focus on relationships and ecosystems, and approaches that address today's data-centric power-imbalanced reality, it offers a more actionable research agenda for the world of the 2020s. This different focus also highlights why HDI cannot just be expanded by adding new tenets; the tenets are quite similar to the data wants identified in this thesis, but the lens through which we view them is different, being as it is more practical and grounded in social, legal and technical reality. This makes HDR something fundamentally different than HDI and justifies the expression of a new research agenda.

In naming this new research agenda, concepts of 'human-technology relations' and later 'human-data relations' which have been the subject of some minor study in the contexts of philosophy, embodied interaction and the performing arts (Ihde, 1990; Hogan, 2012; Windeyer, 2021) were repurposed, to produce the name of **Human Data Relations** (without a hyphen), or **HDR** for short.

HDR's definition draws upon three distinct connotations or readings of its name:

A Definition of Human Data Relations (HDR)

The research agenda for better data relations encompasses all the ways in which humans and human organisations relate to, and with, data, specifically:

- *Human-Data Relations*: users' direct relationships with data, interacting with it to understand and use it, similar to HDI, in service of the direct data wants [6.1] of visible, understandable and useable data.
- *Human "Data Relations"*: individuals' relationships with organisations that hold data about them, in service of the indirect data wants [6.2] of transparency, individual oversight and involvement.
- *Human/Data Relations*: how organisations manage their customers with respect to personal data. Similar to *public relations* or *customer relations*, organisations choose how present their data practices (so as to build trust), and whether they will involve users with data, and provide support to understand data to their users. Organisations can empower individuals and build more effective customer relationships through HDR [4.5.1; 5.6.2; 6.1.2].

HDR uses the concept of 'relations' in a similar sense to that described by Peter-Paul Verbeek, who in his 2015 interactions article unpacks the complexities of designing *'not only the design of technological objects that allow for specific interactions, but also the design of the human subjects who interact with these objects'*. He highlights three approaches to considering human-technology relations–where technology functions as *extensions* of the human, as *dialectical* constructs that can oppose us and give us a relation to ourselves, and where human and technology function in the world in *hybrid* fashion. He further breaks down the different aspects of relations including *types*, *points of contact*, *influences* and *mediation*. Such detailed analysis of relations is more granular than this thesis will go into, but his framing could be useful to practitioners wishing to undertake a structured exploration of the ideas of this thesis.

Having scoped HDR through the definition above, we see that 'better' HDR can be achieved by working to empower individuals [6.3] through pursuit of the identified six wants for data relations. However, as 7.4 will explain, HDR is motivated in two distinct ways, to which

those wants apply differently. As background understanding, let us examine more closely what role data plays in people's lives.

7.3 The Role of Personal Data

Today, almost anything can be encoded as data. Many previously analogue objects and activities now have digital equivalents, so the concept of data has become broad and hard to pin down. Underlying HDR is a need to recognise what roles data can play in people's lives—what it *is* to people. This research has identified eight distinct lenses to explain how people might relate to data—including as property, as memory and as creative work. These are modelled in Table 7.1.

Way of thinking about data	**Explanation & Implications**
Data as property	Data can be considered as a possession. This highlights issues of ownership, responsibility, liability and theft.
Data as a source of information about you	Knowing that data contains encoded assertions about you and can be used to derive further conjectures enables thinking about how it might be exploited by others, but also how you can explore and use it yourself for reflection, asking questions, self-improvement and planning. It invites consideration of the right to access, data protection, and issues around accuracy, fairness and misinterpretation / misuse.
Data as part of oneself	A photo or recording of you, or a typed note or search that popped into your head could be deeply personal. This lens on data highlights issues around emotional attachment/impact, privacy, and ethics.
Data as memory	Data can be considered as an augmentation to one's memory, a digital record of your life. This lens facilitates design thinking around search and recall, browsing, summarising, cognitive offloading, significance/relevance, and the personal value of data.
Data as creative work	Some of the data we produce (e.g. writing, videos, images) can be

Way of thinking about data	Explanation & Implications
	considered as an artistic creation. This lens enables thinking about attribution, derivation, copying, legacy and cultural value to others.
Data as new information about the world	Data created by others can inform us about previously unknown occurrences in our immediate digital life or the wider world. This lens is useful for thinking about discovery, recommendations, bias, censorship, filter bubbles, and who controls the information sources we use, as well as who will see and interpret data that we generate and what effects our data has on others.
Data as currency	Many data-centric services require data to be sacrificed in exchange for access to functionality, and some businesses now explicitly enable you to sell your own data. This lens highlights that data can be thought of as a tradable asset, and invites consideration of issues of data's worth, individual privacy, exploitation and loss of control.
Data as a medium for thinking, communicating and expression	Some people collect and organise data into curated collections, or use it to convey facts and ideas, to persuade or to evoke an emotional impact. This lens is useful to consider data uses such as lists, annotation, curation, editing, remixing, visualisation and producing different views of data for different audiences.

Table 7.1 - Eight Lenses on Personal Data.

People may think of their personal data through any or all of these *lenses* [Karger *et al.* (2005); 2.2.2] at any given time. Any data interaction process or interface design should take these into account. Different informational representations might be needed at different times (Lindley *et al.*, 2018), bringing different aspects of the data to the forefront. Looking across these lenses, there are four specific *roles* that data can serve:

- Data has a role as an **artefact of value** to your life;
- Data has a role in **informing** you about yourself, the world, and the prior or recent actions of others that may affect you;

- Data has a role as **a useable**[11] **material with which to effect change** in your life;
- Data has a role as **a means to monitor changes** in data holders' behaviours, in digital influences upon you, or within your life.

7.4 The Two Distinct Motivations for Better Data Relations

Considering these two types of information in the context of the six wants [Chapter 6] reveals two very different reasons why people might want better data relations:

(i) to acquire *information about your data*, so that you might exert control over where the data is held and how it is used, in order to be treated fairly and make informed choices about personal data. This will be called **Personal Data Ecosystem Control (PDEC)**.

(ii) to acquire *information about yourself*, so that you might gain insights into your own behaviour, and gain personal benefits from those insights or make changes in your life. This will be called **Life Information Utilisation (LIU)**.

Figure 7.1 shows processes individuals might go through in pursuit of these motives. PDEC is a process of holding organisations to account and managing *what happens* to personal data, often regardless of what it means. LIU is more concerned with *what the data means* and its inherent personal value, regardless of where it is stored and how it is used[13]. This novel motivational model was first proposed in (Bowyer, 2021a).

[13] There is some overlap. Organisations hold data to enable interpretation (usually algorithmic) to inform decision making. In this way, organisations are doing LIU for *their* benefit. This grey area is situated as part of PDEC, because from the individual's perspective, how organisations understand you through information informs decisions that affect your life. As such, it is more likely to enable you to exert control over use of your data than to pursue personal LIU goals.

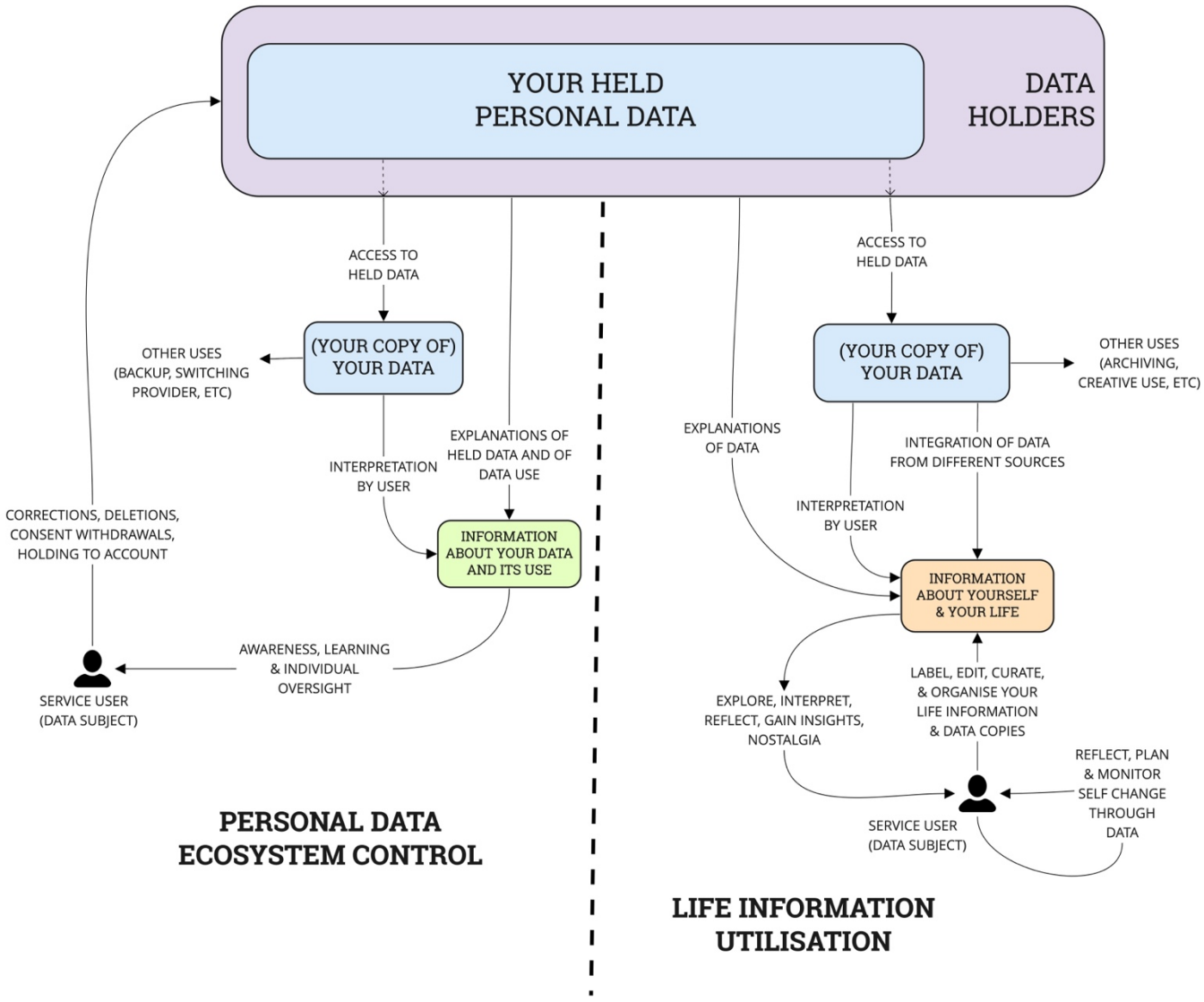


Figure 7.1: The Two Motivations for HDR: Controlling Your Personal Data Ecosystem and Utilising Your Information About Your Life (with 'idealised'[14] processes illustrated)

7.4.1 Life Information Utilisation (LIU)

Life Information Utilisation is a superset of *Self Informatics (SI)* [2.2.3], including all purposes relating to self-monitoring and self-improvement through personal data, but also

[14] The illustrated processes incorporate existing data access processes such as GDPR, where the only access is through provision of a copy of one's data. This is *not* ideal, as it creates divergent versions and will quickly become out-of-sync, however for the sake of simplicity that inefficiency is ignored [see 5.6.1 for improvements to copy-based access].

other uses including creative expression, evidence gathering, nostalgia, keeping, and sharing. Many such desires were expressed in Case Study Two [Table 5.4], and also hinted at in the Early Help context [4.5.1]. While the existence of digitally-encoded information clearly unlocks new possibilities, LIU has existed in some form throughout human civilisation, as seen through analogue processes such as storytelling, journaling and scrapbooking.

The most relevant of the six wants to LIU are *data understandability* [6.1.2] and *data useability*[11] [6.1.3], which relate closely to the HDI concepts of *legibility* and *agency* respectively.

7.4.2 Personal Data Ecosystem Control (PDEC)

Unlike LIU, *Personal Data Ecosystem Control* is a *new* individual need, arising as a result of the emergence of the data-centric world [2.1; 2.2.4]. Only when organisations began to collect and store facts about people as a substitute for direct communication and involvement did it become necessary. The more data is collected about individuals, and the more parties doing so, the greater the need for individuals to understand these acts so that they might influence them (or risk their lives being affected in unexpected or unfair ways). PDEC is a direct response to the power imbalance between data holders and individuals [Hoffman (2014a); 2.1.2].

Several of the six wants are important to PDEC: visible data and transparent processes (referred to collectively as *data ecosystem transparency*), and individual oversight and involvement (referred to collectively as *data ecosystem negotiability*, drawing on the HDI concept of *negotiability*). These grouped terms are used below.

7.5 Four Objectives for Human Data Relations

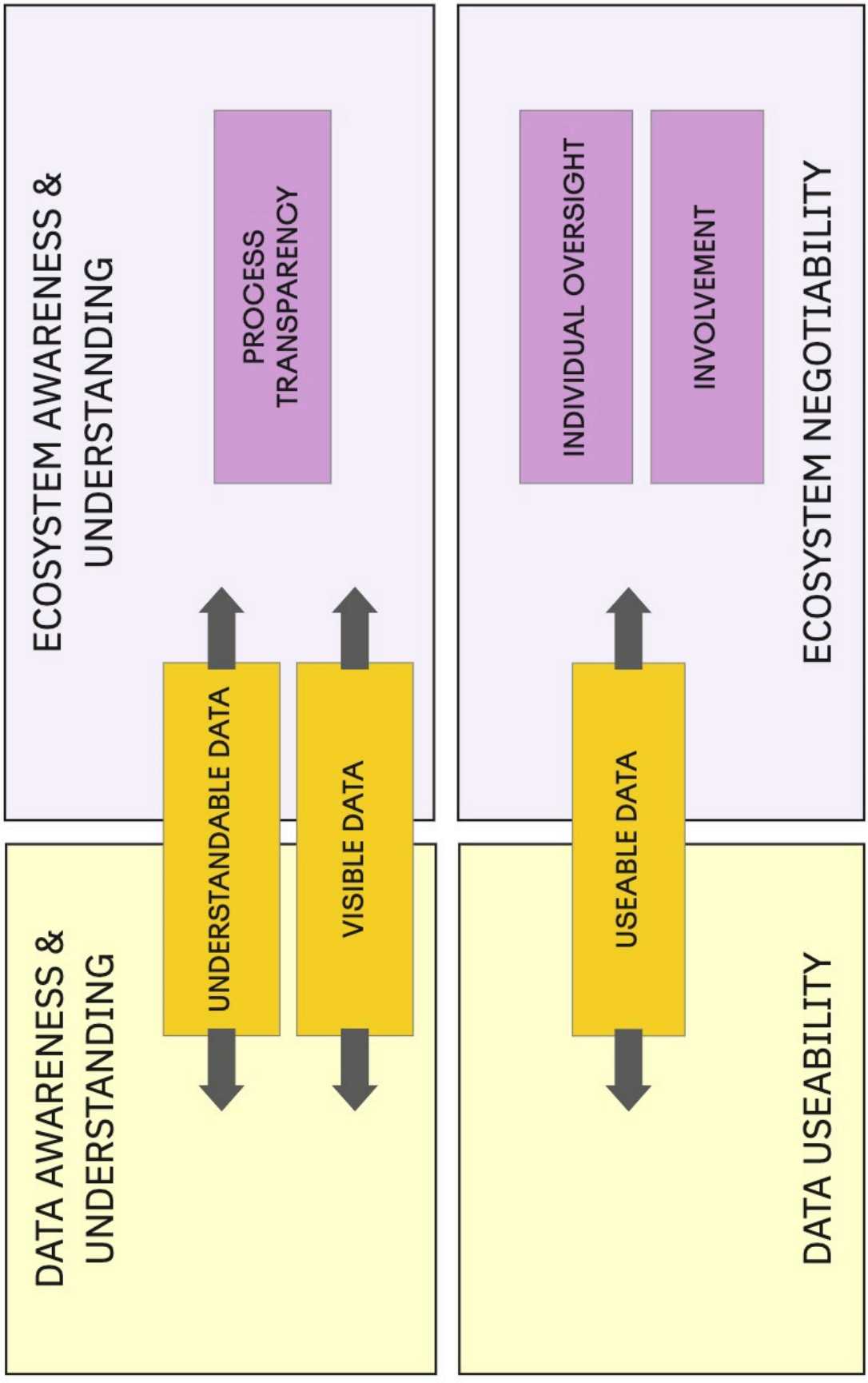


Figure 7.2: Mapping the Six Wants into Objectives for the HDR Opportunity Landscape

To offer future value to future researchers, activists and innovators, this chapter contributes a map of the HDR opportunity landscape. This map is expressed in abstract here, and explored in more depth in Chapters 8 and 9. First, the six wants [Chapter 6] are transformed into four simple *landscape objectives* which shape the ultimate goals for effective HDR in this landscape of opportunity:

- Data Awareness & Understanding;
- Data Useability[11];
- Data Ecosystem Awareness & Understanding[15] and
- Data Ecosystem Negotiability[15].

As Figure 7.2 shows, the need for data to be understandable, visible and useable applies to all types of data, whether it is interpretable as *life information* or *ecosystem information*.

7.6 Better Human Data Relations as a Recursive Public

Let us revisit the stance from which we approach the goals of better HDR. This PhD is grounded in participatory research and experience-centred design [3.2]. Using a *Digital Civics* (Vlachokyriakos *et al.*, 2016) approach to understand people's unmet needs, we can model how the world should change. Such research is political [3.2.1], seeking to correct an imbalance in the world through *adversarial design* (DiSalvo, 2012). Where the Case Studies' track embraced participatory investigation, the embedded research track moves forward from the perspective of an activist researcher, exploring how individuals and groups can actually change their world to meet the established understanding of what should change. By taking this role, we can consider ourselves (those who pursue better HDR, or *HDR reformers* as a shorthand) as a nascent *recursive public* ('Recursive Public (Discussion Page)', no date). This term originates in the free software movement to describe:

> *'a collective, independent of other forms of constituted power, capable of speaking to existing forms of power through the production of actually existing alternatives'*– Kelty (2008)

[15] To avoid overly cumbersome wording, subsequent sections will drop the 'Data' prefix from 'Data Ecosystem Awareness & Understanding' and 'Data Ecosystem Negotiability'.

Being a recursive public means using various means at our disposal to seek to modify the systems and practices we live within in pursuit of our goals. These methods might include participatory research, experience-centred design, software prototyping, rights exertion and campaigning.

This idea of reconfiguring society in this way has been conceived as *civic hacking* (Crabtree, 2007; Levitas, 2013; Tauberer, 2014). The collective around HDR reform does not yet exist as single named and identifiable *public* (Le Dantec, 2016) but there is an emergent trend towards better HDR. HDR reformers today congregate around interconnected and overlapping movements including (but not limited to):

- the *MyData* community [MyData (2017); 2.3.4];
- personal data lockers (CitizenMe, 2021; Sharp, 2021; 'Digi.me', no date);
- digital rights campaigners ('Open rights group: Who we are', no date);
- gig economy worker rights (Kirven, 2018; 'Worker info exchange', 2022);
- privacy by design (Cavoukian, 2010);
- privacy activism (Davies, 1990; 'Bits of freedom: Our focus', 2000);
- data justice (Taylor, 2017; Crivellaro *et al.*, 2019);
- critical algorithm studies (Gillespie and Seaver, 2016);
- adversarial interoperability (Doctorow, 2019);
- the 'right to repair' movement (Svensson *et al.*, 2018);
- 'makers' and digital DIY enthusiasts (Gauntlett, 2015; Altsitsiadis, 2021);
- humane technology (Harris, 2013a);
- design justice (Costanza-Chock, 2020);
- self-sovereign identity (Hori, 2021);
- user liberation (Edwards, 2004);
- explainable AI ('Explainable AI: Making machines understandable for humans', no date); and
- responsible AI (O'Sullivan, 2021).

7.7 Summation: HDR—A Landscape Ready to Explore

The commonality to so many groups [7.6] suggests HDR reform is an emergent cultural phenomenon, whether or not a single identifiable public coalesces. Time will tell whether *Human Data Relations* as laid out in this thesis is sufficient to give form to that phenomenon. At the least, HDR offers a descriptive umbrella term. The breadth of research, innovation and activism validates the need *and* the desire for such a recursive public around HDR reform to exist. In fact, it already does exist, whether named and unified or not. There is a growing desire to take back control of technology for the betterment of humankind. Therefore, in the next chapters this thesis takes an unashamedly critical view of the status quo, favouring the disruptive societal changes these movements seek. The HDR research agenda poses (from a data-oriented design standpoint) an important question of society:

"How can we change the world into the one we want?"

This chapter has established a clear scope and motivation for the new research agenda of Human Data Relations. This sets the stage for the next two chapters, which describe this researcher's efforts to pursue the HDR agenda in practice as part of his role in embedded academic and industrial project work. The next chapters will take the reader on a journey of exploration through the HDR landscape, recognising the design implications of law and the commercial realities of design and considering the above question in more detail. By highlighting the pitfalls and opportunities that exist, the objective is for this thesis to provide actionable principles and approaches for HDR reform practitioners that can help nascent movements toward HDR reform to flourish.

It is important to note that the Human Data Relations research agenda, while it views a change from the data-centric status quo as positive, does not imply that to be an HDR reformer one must be an activist. While adversarial design [9.5.2] and collective activism [10.1.10] are clearly useful approaches to progress HDR, systems can also be changed from within. Designers and thought leaders within data-centric organisations and governments could embrace HDR's objectives themselves. The fact that so many of the HDR reform communities mentioned in the previous section are somewhat activist in nature, is more of a reflection of the reality that the goals of HDR cannot easily surface within the dominant

digital commerce paradigms of today, and so people - pursuing both individualist [3.1] and collective agendas [10.1.10], operate at the grassroots level, and may seem to many to be activists. While this thesis leans more into the activist angle of HDR research, motivated by the researcher's background [1.1.1], balance is sought through the fact that the embedded project work track has included work in projects both inside (BBC R&D, SILVER) and outside (Hestia.ai, DERC) of the system [see Section IV Introduction].

What follows in Chapters 8 and 9 is deliberately broad and open-ended[16]. It does not provide a complete answer to this question or the extended research question [7.1], nor could it. This thesis does not pretend to be complete or definitive in its interpretation of the outlook for HDR. Alternative interpretations and schools of thought than HDR exist. What follows is not a strict roadmap, but rather a preliminary model based on a sampling of ongoing work, identified challenges and known opportunities, that can be understood through the remainder of this thesis, leading to the major contribution of this thesis as an actionable research agenda and useful reference material for the pursuit of better human data relations.

---

[16] Some of the challenges and opportunities described in the next two chapters are covered in greater detail than others. This corresponds only to this researcher's proximity and depth of engagement with those ideas, rather than their relative merit, complexity or impact potential. Given the broad aim to chart a new area of research, it is considered to be more useful to introduce a range of applicable ideas, even if some are only lightly detailed, than to document just a few.

SECTION IV:

OPERATIONALISING THE RESEARCH AGENDA THROUGH DESIGN & PRACTICE

Introduction to Section IV

Section III concluded the documentation of the participatory track of the thesis, bringing together findings from both Case Studies to identify the six wants that people have from data, and then formulated these into a new research agenda, Human Data Relations (HDR), with specific objectives to improve awareness, understanding, useability and negotiability. As Figure 3.2 illustrated, alongside and throughout this PhD I was conducting embedded project work in the roles of developer, designer, and researcher, in projects that also relate to the exploration of improved data relations, supporting a growing understanding of not just *what* people want from data, but *how* such desires might be achieved. The goal for Section IV, therefore, is to document the insights and learnings from that practical work, providing the reader with both a more detailed understanding of the landscape of HDR, but also with practical examples, approaches, insights and principles that can be taken forward to operationalise the agenda and take actions that might bring about better HDR in society.

Within this section introduction, I will introduce the four projects I participated in—namely CHC's SILVER project, BBC R&D's *Cornmarket* project, DERC's Healthy Eating Web Augmentation Project and Hestia.ai/Sitra's *digipower* project. Through field experience in these projects, I have understood constraints and opportunities that affect data interaction system and process design. Concurrently, I have fed research learnings about HDR back into those projects, creating practical impact. These projects were not formal Case Studies from this thesis' perspective, and thus are not documented at length. Rather, they are described only in sufficient detail to contextualise the learnings acquired through participation in them (which are presented in Chapters 8 and 9. Nonetheless, they have provided me with the opportunity to undergo a process of acculturation to the world of practical system building and project operation in the PDE space [2.3.4]. Through design, technical prototyping and pushing the boundaries of existing systems, I have developed knowledge and gained expertise which allows me to draw conclusions with confidence about how the discipline of Human Data Relations should proceed in its future R&D to best serve individual and societal interests.

This section has two chapters: Chapter 8 focuses on mapping the HDR landscape. Using the four HDR objectives as a frame, the barriers and obstacles to progress against each

objective are explored and explained, as well as a set of general HDR obstacles. Chapter 9, cognizant of these obstacles, outlines four specific approaches that could be adopted, including both activist and conformist approaches to research, design and innovation in this space.

While the Case Studies were useful for surfacing problems that users experience in today's digital world (as with GDPR access in Case Study Two, for example), their ability to explore the practicalities of change was limited. In identifying and confronting practical challenges and exploring feasible approaches, this Section aims to enhance the HDR research agenda offered by this thesis, so that it is not only a direction of travel, but also the preliminary outline of a roadmap that can be used to begin to design a better future, identifing actions that may be possible within the constraints of the commercial, legal and technical realities of today's digital landscape.

Embedded Research & Design Settings

Concurrent to this PhD, I took a major role in two industrial research projects (1 & 2), and two academic research projects (3 & 4):

- **BBC R&D's Cornmarket Project** (Sharp, 2021), which explored through user experience design, technical prototyping and participatory research, how individuals might interact with data through a Personal Data Store interface;
- **Sitra/Hestia.ai's *digipower* Investigation** (Härkönen and Vänskä, 2021), a successor to Case Study Two, in which European politicians examined companies' data practices through exercising data rights and conducting technical audits;
- **Connected Health Cities (CHC)'s SILVER Project** (Connected Health Cities, 2017), where I, along with a backend developer and a team of researchers, developed a prototype health data viewing interface for Early Help support workers; and
- **Digital Economy Research Centre (DERC)'s Healthy Eating Web Augmentation Project**, which explored the use of web augmentation techniques to modify the user interface of takeaway service *Just Eat* to include health information, in support of healthy eating.

Further details about these projects and my involvement in them, sufficient to understand these learning contexts, are provided below. See also the end of this introduction for a note about the attribution and origin of the ideas presented in Section IV.

BBC R&D's Cornmarket Project

I took a three-month sabbatical from my PhD in the summer of 2020, during which I was remotely embedded within a full-time research internship at **BBC R&D** - the British Broadcasting Corporation (BBC)'s Research and Development (R&D) department (British Broadcasting Corporation, 1997), working with specialists, designers, researchers and developers on an exploratory research project codenamed *Cornmarket*. I continued this involvement as a part-time research consultant and critical friend for a further 5 months after the conclusion of the initial three-month placement.

As part of its Royal Charter, one of the BBC's lesser-known obligations is to maintain a centre of excellence for research and development in broadcasting and electronic media. To this end it employs over 200 researchers around the UK in its R&D department, looking at everything from AV engineering and production tools to new forms of media, virtual reality, digital wellbeing and human data interaction (British Broadcasting Corporation, 1997). The Cornmarket project was launched in 2019 and derives from the BBC's earlier Databox/BBC Box project [9.2]. Cornmarket is a BBC-internal human-data interaction research project which explores a possible role for the BBC as it moves beyond broadcast television, using its public service responsibility to guide citizens to a position of empowerment within today's digital landscape–encompassing not just entertainment but health, finance and self-identity. Due to its unique funding from UK-wide TV licensing and its duties to not only entertain but to inform and educate the general public, the BBC is uniquely placed to take a more human-centred approach than commercial innovators in this space as it needs only to deliver value, not profit. Technically, the project explores the use of Solid [Berners-Lee (2022); 9.2] technology to build a working Personal Data Store (PDS) prototype [2.3.4] while also developing, iterating and trialling user interface designs for PDS data interaction and conducting participatory research interviews and activities to explore what for a BBC PDS might take and what features its potential users might value.

The Cornmarket team produced a prototype, internally called *MyPDS*, which would allow people to populate a PDS with personal data from APIs and data downloads from a variety of services including BBC iPlayer, Netflix, All4, Spotify, Instagram, Strava, Apple Health, banks and finance companies, as well as social media companies such as Facebook, LinkedIn and Twitter. MyPDS would use these combined data sources to create personal *profiles* for Health, Finance, and Media (i.e. entertainment) as well as a Core profile. Within these profiles various data insights, visualisations, capabilities would be delivered. One feature the work explored in depth as potentially valuable to users is the ability to include and exclude certain datapoints from the imported viewing history data in order to present a more accurate, curated view of oneself that could then be fed back to other applications such as BBC Sounds to give better content recommendations.

With a cross-disciplinary team of around 20 people including architects, developers, user experience designers, product designers, innovators, participatory researchers and marketers, and funding to outsource public engagement research to agencies, this project represented, at the time of the internship in 2020, a significant player in the emerging personal data economy [2.3.4]. As such the Cornmarket project is a fertile ground in which to learn more from practitioners in the PDE space and to test the learnings of this thesis in practice while also finding deeper insights in response to my research questions - in particular the extended RQ concerning the building of more human-centric personal data interfaces in practice.

Much of the work I did during this extended internship can be seen in the designs within 9.4, as well as the research report I wrote (Bowyer, 2020a) and internship writeup (Bowyer, 2020b). My work with the Cornmarket project can be seen as the concluding part of one of several action research cycles within the PhD [3.2.2].

A number of articles relating to the Cornmarket project have been published:

- Personal data stores: building and trialling trusted data services - BBC R&D (Sharp, 2021)
- Building trusted data services and capabilities (explainer videos) (Sharp and Bowyer, 2022)
- The BBC's radical new data plan takes aim at Netflix (Wired) (Orphanides, 2021)

- Sir Tim Berners-Lee and the BBC stage a very British coup to rescue our data from Facebook and friends (The Register) (Goodwins, 2021)
- BBC and Sir Tim Berners-Lee app mines Netflix data to find shows viewers like (The Times) (Kanter, 2021)
- Stronger Together: Cross Service Media Recommendations (research paper) (Ricklefs *et al.*, 2021)
- BBC wages war on online echo chambers with 'unbiased' tech (The Telegraph) (Woods, 2022)
- Richard Sharp (BBC Chairman) discusses Data management & privacy (video) (Tim Davie (BBC Director-General), Richard Sharp (BBC Chairman) and Clare Sumner (Director of Policy), 2022)

In April 2023, I rejoined the BBC as a full-time employee and Senior UX Designer, ready to work on the design of both staff-facing and public-facing data experiences in the BBC's non-commercial context.

Hestia.ai, and Sitra's *digipower* Project

Following the conclusion of the funded period of my PhD, I took up a near-full-time position as Project Leader and Personal Data Coach at **Hestia.ai** (Dehaye, 2019)), a startup based in Geneva, Switzerland. Hestia.ai is a company conducting research, developing technologies, and delivering training, in the emergent MyData/PDE space [2.3.4]. In essence, the company's mission is to help individuals and especially collectives to more easily obtain and understand data held about them, and to help them visualise, aggregate and make use of that data. It is an example of a **data access and understanding services** company as described in 9.3.4.

I was specifically hired to co-lead the *digipower* project (Härkönen and Vänskä, 2021), for Hestia.ai's client, **Sitra** (Sitra, 1967). Sitra is a non-profit organisation in Finland, funded by the Finnish Parliament and accountable to the Finnish people. The goal of the digipower project was to guide 15 European politicians, civil servants and journalists through the process of obtaining and exploring their own data. The participants were high-profile VIPs, including the former Prime Minister of Finland and former European Commission Vice President, Jyrki Katainen. The goal was to empower those individuals to better understand

the workings of the data economy, so that they might be able to influence others and effect change. One of Sitra's goals is to establish a fairer data economy (Sitra, 2018). Methodologically, the project drew heavily on my own Case Study Two [Chapter 5], adopting a similar method of guiding individuals through the process of making GDPR requests and scrutinising the returned data; I was employed on the project for this expertise. Where it differed from my own Case Study was that the focus of the research was outward, on the data economy and the practices of service providers, rather than inward, on the lived experience of the participants. Other differences included the building and use of software interfaces to provide participants with data visualisations, the use of TrackerControl software to audit mobile phone apps [Principle 12–10.1.12], and the direct analysis of participants' retrieved personal data by the Hestia.ai research team (whereas my Case Study explicitly avoided handling participants' personal data). The project resulted in three reports:

- Sitra's official project report (Härkönen *et al.*, 2022); and
- Two technical research reports by Hestia.ai:
  - A high-level interpretation of models of power and influence in the data economy (Pidoux *et al.*, 2022); and
  - A detailed auditing of provider practices, evidenced by examples from participants' data (Bowyer, Pidoux, *et al.*, 2022).

I continued to be employed by Hestia.ai until February 2023, working on the research, design and development of tools to help collectives [Principle 10–10.1.10] with data, make data easier to understand [6.1.2; 7.5; 9.3.4], and exploring methods to help people 'hack the seams' of digital platforms and services [9.5].

Where the BBC internship see above helped me to understand the practicalities of connecting people with their personal data in pursuit of Life Information Utilisation [7.4.1], my work with Hestia.ai helped me understand the practicalities of how people might acquire greater Personal Data Ecosystem Control [7]. In this sense, both industrial projects were highly complementary to developing an overview of the pursuit of HDR in practice.

Developing a Health Data Interface within CHC's SILVER Project

The SILVER project was a government-funding 'care pathways' project funded as part of the Connected Health Cities programme between 2017 and 2019, and focused upon developing interventions, including new data interfaces for the viewing of health data by support workers in an *Early Help* context. For more details on this context and this project at a high level, please refer to Section II Introduction and 4.2.1. A large part of my role was as a researcher, which enabled Case Study One Chapter 4's participatory workshops to take place. I was also the front-end developer within the team, taking forward the research findings from SILVER interviews and workshops and using them to develop a prototype health data viewing interface. As part of this, I worked (mandated by the project to adopt a more data-centric stance than the rest of this thesis, which provided valuable perspective, as discussed in 4.3.3) with back-end developer colleague Stuart Wheater of Arjuna Technologies to architect and design the system. This is where my role in the project becomes relevant to the exploration of the extended RQ. The work involved many discussions with technical staff and care professionals within local authorities, as well as technical and architectural design work. All of this contributed to the evolving understanding of obstacles to the building of human-centric data interfaces in practice. Videos of the prototype, which was tested with local authority staff and received positive feedback, are available online (Bowyer and Wheater, 2017).

DERC's Healthy Eating Web Augmentation Project

As a software developer, I have been aware for a long time that one of the biggest challenges in building new data interfaces is to gain programmatic access to the necessary data. As part of the trend towards cloud-based services and data-centric business practices, it has become increasingly difficult to access all of the data held about users by service providers. Application Programming Interfaces (APIs) are a technical means for programmers to access a user's data so that third-party applications may be built using that data. Unfortunately, as a result of commercial incentives to lock users in and keep data trapped (Abiteboul, André and Kaplan, 2015; Bowyer, 2018b), much of users' data can no longer be accessed via APIs [8.4.2]. While GDPR data portability requests do open up a new option for the use of one's provider-collected data in third-party applications, this is an awkward and time-consuming route for both users and developers. **Web augmentation**

provides a third possible technical avenue for obtaining data from online service providers. It relies on the fact that a user's data is loaded to the user's local machine and displayed within their web browser every time a website is used, and therefore it is possible to extract that data from the browser using a browser extension; this as another **seam** that can be hacked—see 9.5.2 and Principle 12 [10.1.12]. Similarly, once loaded into the browser, a provider's webpage can be modified to display additional data or useful human-centric functionality that the provider failed to provide.

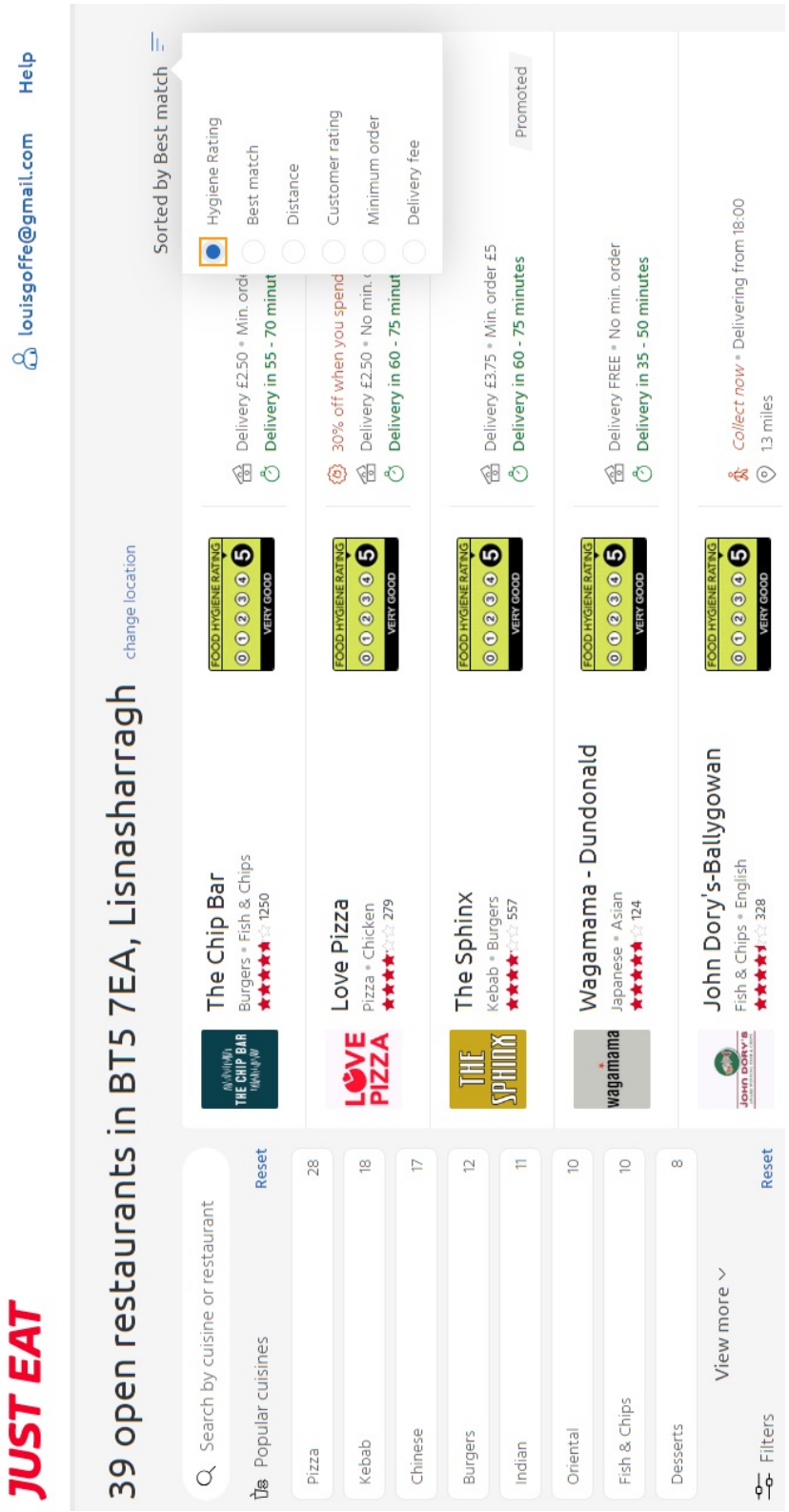


Figure IV.1: Screenshot from a Web-Augmented version of the Just Eat Website, showing hygiene information and offering additional sorting

In order to better understand what is and is not possible using this technique, I participated part-time from 2018 to 2020 as the sole software engineer in a DERC (Digital Economy Research Centre) project. This project was using the web augmentation technique to explore how researchers could improve the information given to users of Just Eat, a takeaway food ordering platform in the UK. Hygiene Rating information for each outlet was added, as well as a feature to enable user to sort by hygiene rating, as shown in Figure IV.1. The theoretical basis for this research was published in (Goffe *et al.*, 2021, 2022). While this particular use case does not concern personal data, the technology and techniques being used by the project to exploit the browser seam were considered highly relevant to the exploration of HDR-improving possibilities, and the goals of the research project were also human-centric, and consistent with this thesis's research goals - tackling the hegemony of service providers in order to better serve individual needs. Like the work with Hestia.ai, this project also embodies an activist approach to improving human data relations, because it explores methods that could be undertaken without the buy-in of incumbent data-holding service providers.

Special Attribution Note for Section IV

This is a note about the attribution of insights within Section IV, as the ideas originate quite differently than from the rest of the thesis.

This thesis is my own work. All ideas synthesised in the project track (which did not involve research participants) are original. *Some* of the specific details, theories and ideas presented in Chapters 8 and 9 arose or were developed or augmented through my close collaboration, discussion and ideation with other researchers, designers and innovators both alongside and prior to the PhD timeframe, including:

- Jasmine Cox, Suzanne Clarke, Tim Broom, Rhianne Jones, Chris Gameson, Alex Ballantyne and others at BBC R&D;
- Paul-Olivier Dehaye, Jessica Pidoux, Francois Quellec and others at Hestia.ai;
- Stuart Wheater of Arjuna Technologies and Kyle Montague of Open Lab during the SILVER project;
- Louis Goffe of Open Lab on the DERC Healthy Eating project;

- earlier innovation work with Alistair Croll and Sarah Severson at Rednod, Montréal, Canada (circa 2011); and
- earlier innovation work with Megan Beynon and Dean Upton at IBM Hursley, UK (circa 2006).

Due to these collaborations and the ongoing and parallel nature of many of these projects to my PhD research, it is impossible to precisely delineate the origin of each idea or insight. In practice, ideas from my developing thesis and own thinking informed trajectories and thinking of projects I was involved in, and vice-versa. These ideas would not have emerged in this form without my participation, so they are not the sole intellectual property of others, but equally I would not have reached the same conclusions alone, so the ideas are not solely my own either. All diagrams and illustrations were produced by me, except where specified, and the overall HDR synthesis and framing presented in Sections III-V is entirely my own original work. Where this section includes material from the four embedded projects, that material is either already public, or permission has been obtained from the corresponding individuals or project teams for it to be included here.

Chapter 8. Mapping the Human Data Relations Landscape

"There are certain things you do not in good conscience do to humans. To data, you can do whatever you like."

—Nikhil Sonnad
(data journalist and technology commentator)

This chapter begins to engage with the expanded research question laid out in 7.1. Considering how better HDR might be achieved in practice generates further questions: What is it that inhibits the building or adoption of human-centric technologies? What are the barriers and what opportunities might overcome them? How can we catalyse progress toward *MyData*'s human-centric agenda (MyData, 2017) and the HDR agenda laid out in Chapter 7? What challenges are faced when attempting to build human-centric technologies for today's world? Building on an understanding of human experience of the data-centric world, can we more provide an outlook for PDE design & development and define a *research agenda* for the next step of tackling the PDE challenge?

Informed by my work within the embedded projects [see Section IV Introduction], I have identified specific *obstacles* to the four HDR objectives [7.5]. Analogous to Li's *barriers cascade* [2.2.3; Li, Forlizzi and Dey (2010)], these are the challenges that individuals or system designers must be empowered to overcome. These obstacles are documented in the following sections. Figure 8.1 depicts an HDR-specific barriers cascade: a route of overcoming obstacles (shown as blue boxes) through which individuals might be empowered and by which organisations might become more HDR-friendly. The concepts introduced in this figure will be refined and explained across this chapter and the next. The latter blue boxes (in the 'solution space' box) are more pervasive obstacles that apply to all four HDR objectives. In uncovering the obstacles shown in the figure, I have developed several design Principles or insights for HDR (shown by clouds). These help obstacles to be overcome, and could help adversarial designers or strategists to facilitate better HDR. These constitute the 'preliminary roadmap' mentioned in the introduction above. These insights

are mentioned where relevant here but further detail for each principle can be found in section 10.1.

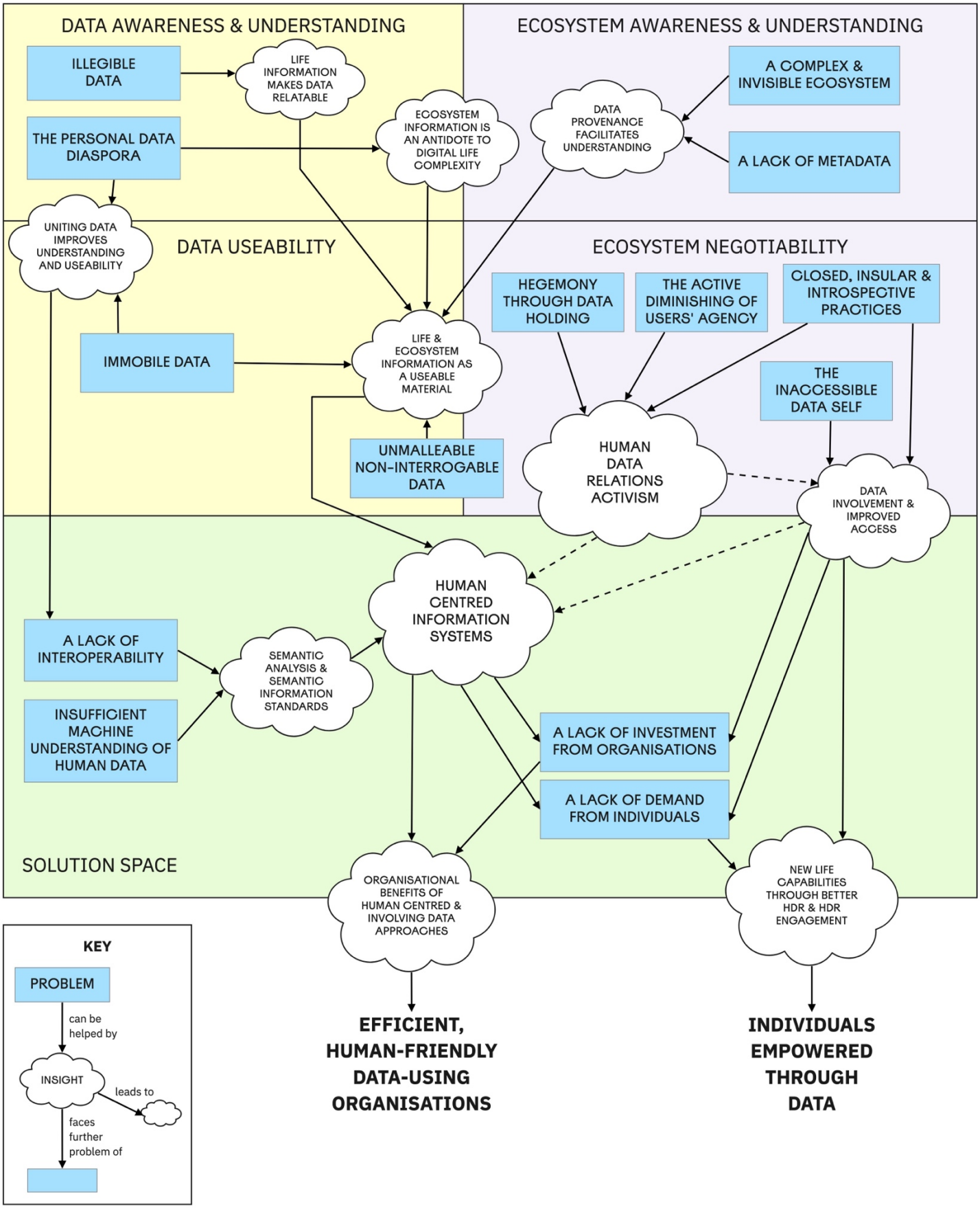


Figure 8.1: Obstacles and Resulting Insights in the HDR Opportunity Landscape

8.1 Obstacles to the HDR Objective of Data Awareness & Understanding

8.1.1 Invisible, Inaccessible or Unrelatable Data

In pursuit of visible, understandable data [6.1.1; 6.1.2], the first obstacle encountered is that most personal data is **invisible**, **inaccessible** or **unrelatable**. It is trapped in service providers' databases, or on different devices or hard drives, or by software limitations and proprietary file formats (Abiteboul, André and Kaplan, 2015; Bowyer, 2018b). My research participants spoke of 'not knowing' what data exists and of being 'in the dark'. Case Study Two showed that even where data is accessible, it is not **relatable** (*legible* (Mortier *et al.*, 2014); 2.3.2). The objective here is to ensure that people not only have awareness of their data, but can understand ('make sense' [Gurstein (2011); 2.1.4]) of what it means. Principle 1 [10.1.1] can help address this: interpreting data as *life information* makes it more relatable.

8.1.2 The Personal Data Diaspora

Another important obstacle to consider here is what I call the **Personal Data Diaspora**[17]. As illustrated by Imogen Heap's quote opening Chapter 1, an individual's personal data is typically very widely dispersed, and there is no central, holistic view of one's data. For example, if I consider just my movement tracking data, I have over time accumulated activity logs from walking, running, cycling, and driving which are stored by Nike+, MyFitnessPal, Strava, Google Fit, Fitbit, Apple Health and Google Maps, not to mention the records remaining on my different smart watches, smartphones, hard drives and insurer black boxes. This is the SI problem of *Integration* [Li, Forlizzi and Dey (2010); 2.2.3]. As well as the challenge of managing one's data ecosystem [2.2.4], this makes it impossible to view physical activity history in one place, to spot patterns over time or make comparisons. To overcome this, we need interfaces that recognise the scattered reality of each individual's personal data, and begin to make that ecosystem visible and understandable 8.3. Data awareness and understanding is a problem of representation. Invisible data should be

[17] The word *'diaspora'* is typically used with reference to populations, but is an apt term, derived from the Greek 'diaspeirein' meaning 'scattered about' or 'dispersed'.

visibly represented. All data should be represented as contextually relatable life information. This leads to Principle 2 [10.1.2]—that access to *ecosystem information* can help to de-complexify our digital lives.

8.2 Obstacles to the HDR Objective of Data Useability[11]

8.2.1 Immobile, Inaccessible or Unmalleable Data

To improve the useability of data, let us consider what properties of data make it hard to use. Most personal data is immobile, inacessible, unmalleable and not interrogable.

Data is **immobile** in that is very difficult to move it out of its environment. Most data exists in organisations' internal databases, where it is tightly coupled to technology stacks, interfaces and business processes. Separating one's data from the service that holds it is difficult and often impossible. It is **inaccessible** to individuals (in the sense of *effective access* (Gurstein, 2011)). Data access requests such as GDPR are typically satisfied by creating a copy of the data, creating problems of delay, divergence and understanding. Even then, returned data is incomplete [5.4.2]. Its accessibility is also hindered by the technical nature of data. Data is often stored in complex proprietary structures which are designed for the algorithmic efficiency of the specific business operations rather than for general-purpose re-use.

People need to be able to ask questions of their data [Table 5.4; 4.4.2]. But data is **not interrogable**. It must stand for itself, yet there is no obvious way to ask a question about the meaning of the data or its ability to answer a particular question. To ask questions of data requires either the co-operation of the data holder or advanced technical skills in data querying and analysis (assuming the data is complete and contextualised). Data needs to be **malleable**—capable of being broken down, looked at from different perspectives, and reconstituted in different ways. This goes beyond visually representing the data, and implies an ability to interact with the data to produce new interpretations and insights to investigate specific questions.

To overcome these obstacles, data must be freed from its current constraints and moved into environments where it can be freely examined and reconstituted without restriction.

This leads to Principle 2 [10.1.2]—that data should be united and unified so that it can be interrogated holistically and without artificial constraints of device, format or platform.

Furthermore, as Principle 3 [10.1.3] explains, data will only become useable once we change its nature. Since the 1970s, drawing on the then-common metaphor of a filing cabinet, computers have considered *files* as the basic material that users will interact with; largely static objects which are optimised for reference rather than use. Where we do interact with data as information instead of files, that information is typically presented in limited contexts within certain products or apps [Principle 1–10.1.1]. To move up the DIKW pyramid [2.1], we need smarter computer systems, that move beyond files (Bowyer, 2011). We need systems whose basic material is not files, but pieces of human information. **We need a human information operating system**.

8.3 Obstacles to the HDR Objective of Ecosystem Awareness & Understanding

8.3.1 Complex and Invisible Personal Data Ecosystems

Crabtree and Mortier highlighted that users need their whole personal data ecosystem to be visible (Crabtree and Mortier, 2016). As established [2.2.5; 2.3; 6.2], HDR cannot be made effective without a sea change in the way that individuals are able to interact with the complex ecosystem of personal data they each inhabit. Our PDEs are incredibly complex and largely invisible. For example, it is easy to allow a handful of messaging and social media apps to access your contact list. Before you know it, you have created a complex and unmanageable network of connections that silently sync and propagate your addresses and phone numbers across the Internet. And there are deeper layers which are even less evident to users: networks of data brokers, advertisers and digital cookie companies exchange user identifiers, activity data, inferred interests and personal information about you while you browse or use apps (Pidoux *et al.*, 2022). The ability to build up a meaningful picture of your personal data ecosystem is completely absent [4.4.4] or severely limited. People remain 'in the dark', leading to fear (Bowyer *et al.*, 2018), overload [2.2.4] and resignation [5.5.4]. Managing one's personal data ecosystem is an **overwhelming, unmanageable task** that even personal data experts are not fully able to get a handle on. We do not feel 'in control' [Teevan (2001); 2.2.2]. The ability to provide a user with

ecosystem transparency is hindered by the complexity and multiplicity of data relationships they have been encouraged to set up. People lack tools to provide a meaningful, or indeed any, view of those relationships. In both Case Study contexts, we saw that no one individual or organisation has the ability to see the whole of a user's data ecosystem [4.4.4; Cornford, Baines and Wilson (2013)]. There is little commercial motive to try and solve this problem, as each provider focuses on their own apps, websites and services. Making one's ecosystem visible, transparent and understandable is therefore an essential ingredient of any design for better HDR, as Principle 4 [10.1.4] identifies.

8.3.2 A Lack of Metadata

As we start to consider *what the data is about*, new possibilities are unlocked. A PDS-type system could be built that is not only a repository of personal data, but (using proxy representations), a collection of ecosystem information and *contextually-situated* life information too. This could include information about relationships with data holders or other entities. Builders of such a system would face a further challenge—**a lack of metadata** [2.2.2]. Typically, most data on our hard drives lacks context about its origin, and how it relates to the individual in a holistic life or ecosystem sense. Where data access rights are executed (or data is personally shared [4.4.2]), the attention is on the data itself: what it says. But as Case Study Two showed, some of the most desired information was not the data itself, but handling information and inferences—information that can only come from metadata, which was rarely forthcoming [Table 5.3]. Metadata could include many facets that could be quantified and recorded, as illustrated in Figure 8.2, which I created at BBC R&D:

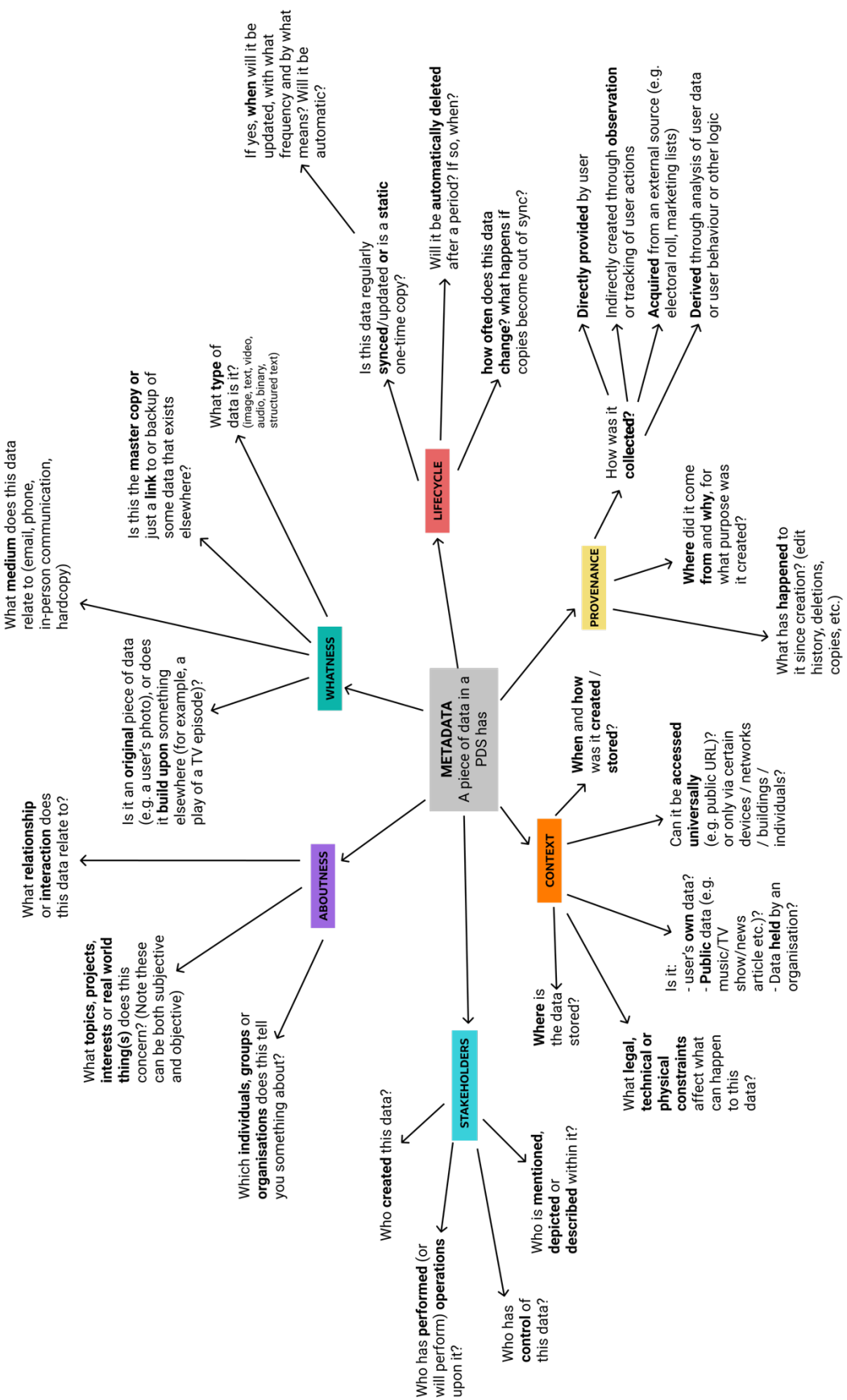


Figure 8.2: Some of the Many Aspects of Metadata that Might Exist About a Datapoint or Dataset

These facets can be mapped back to the 5 W's that collectively make up the user's *context* [Abowd and Mynatt (2000); 2.2.2]. Many of these facets are not explicitly recorded today, or would take significant work to capture. Nonetheless, this exploration shows how data can be better contextualised, supporting contextual and associative approaches [2.2.2]. This leads to Principle 5 [10.1.5], which says that we must know data's provenance if we are to be able to take good interpretations from it and make good use of it. Paying attention to ecosystem information, metadata and provenance facilitates a new space that, at the time of writing in 2022, almost no-one is building for. For people to manage their digital world, **they need a map**. This is the first step on the road to giving individuals oversight of their personal data ecosystem.

8.4 Obstacles to the HDR Objective of Ecosystem Negotiability

There are three distinct obstacles to ecosystem negotiability:

- the intrinsic structures that give data holders power [8.4.1],
- the trend of actively diminishing user agency [8.4.2], and
- the intractable data self [8.4.3].

8.4.1 Hegemony through Data Holding

It is in the pursuit of oversight [6.2.2] and involvement [6.2.3] that the impact of the power imbalance [2.1.2] becomes most clear; unlike the other HDR objectives, individuals cannot act to claim ecosystem negotiability for themselves. Negotiability means having the power to act, and in the context of systems and interfaces owned and designed by service providers, **that power can only be given**. The hegemony of data holders is therefore is the greatest obstacle to this objective, so it is vital to examine the nature of that power if we are to confront it. Where does it come from?

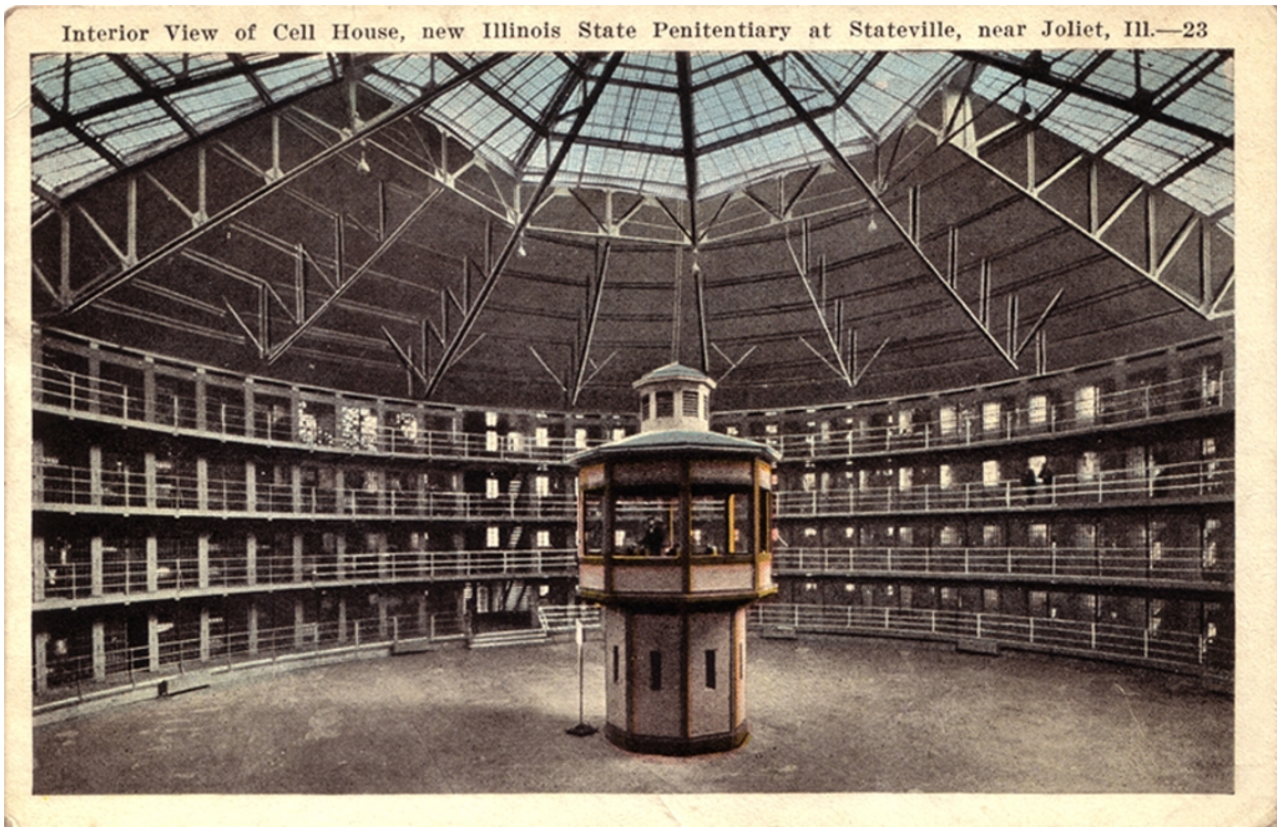


Figure 8.3: The Panopticon Structure of the Illinois State Penitentiary

A helpful analogy for the relationship between provider and user can be seen in the design of Jeremy Bentham's *panopticon* (Bentham and Bozovic, 1791), a real-world version of which is pictured in Figure 8.3. The panopticon is an 18th-century prison architecture that elevates the power of the (hidden) prison guards to observe all the prisoners easily at any time while removing prisoners' privacy and providing them no ability to observe those in power. As in Orwell's *Nineteen Eighty-Four*, individuals are unable to know when they are being watched (in this case, because the guards are hidden from view by one-way screens). This enforces compliance. Structuralist philosopher Foucault interpreted the panopticon as a political design, recognising that human environments can be configured to influence or regulate behaviour, in order to defend the power of the ruling class (Foucault, 1975). Such designs embody his four principles:

- **Pervasive Power**: the guards see everything all the prisoners do, all the time
- **Obscure Power**: the guards can see into any cell at any time, but the prisoners can't know when, how or why they are being observed

- **Direct Violence Made Structural**: the structure motivates the prisoners to self-regulate their behaviour without being coerced (through beating or punishment)
- **Structural Violence Made Profitable**: having been made compliant by the structure, the prisoners can be put to work for the benefit of those in power, as it is the only option available to them.

We can see at least three of these traits in modern Internet platforms such as Facebook today. These platforms monitor users' behaviour without their knowledge (pervasive power), and without accountability (obscure power). Interfaces are designed to offer only those actions that benefit the platforms, e.g. clicking ads, sharing content or spending more time on site (structural violence made profitable). This has happened through the processes of *platformisation* and *infrastructurisation* (Helmond, 2015; Plantin *et al.*, 2018) which have supplanted the Web 2.0-era promise of a free, open Internet that could have been more empowering to individuals.

Through the control of data and of interface design—the only channels through which they can be observed—**service providers and platforms assert a structural power over the digital landscape**. Just as the design of the panopticon regulates the behaviour of the prisoners, so the configuration of platforms, apps and service interfaces we use regulate and limit our behaviour as users. As Lessig wrote, *'code is law.'* (Lessig, 2000). This infrastructural power is explained further in [Principle 6–10.1.6], which arose from research conducted in Hestia.ai and Sitra's digipower project [See Section IV Introduction].

Structural power is not the only form of power which modern-day data-centric service providers hold. Jasperson *et al.*'s extensive review of types of power in the context of technology organisations (Jasperson *et al.*, 2002) identifies 23 different power paradigms, of which at least 13 can be, and are, asserted by data-centric organisations today:

- ***authority***: ownership of technology or infrastructure (e.g. of websites, servers and code)
- ***resource control***: controlling the flow of resources (in this case of information/data)
- ***systems/structural power*** structural manipulation of others (as detailed above)
- ***rational power***: controlling decision-making processes (such as banning users)

- ***disciplinary power***: using an influential position to affect others' mental models (e.g. positioning location tracking as theft resilience)
- ***zero sum power***: winning a battle for ownership/resource control at the other party's expense (e.g. losing control of your sacrificed data)
- ***behavioural influence***: persuading others to carry out the desired behaviour (e.g. restricting features to motivate subscription payments, or promoting certain content or actions)
- ***interpretative influence***: determining how reality is externally represented (e.g. Facebook determining the way in which your social network is represented to you)
- ***network centrality***: becoming an indispensable hub of a wider ecosystem (e.g. Facebook/Google dominance in online ad-brokering)
- ***processual power***: changing processes for competitive advantage (e.g. platforms offering preferential APIs or rates to compliant partners)
- ***socially shaped power***: influencing a wide audience to settle upon a preferred interpretation (e.g. using dominant market position to dominate debates e.g. about privacy norms)
- ***interpretive power***: creating the internal representations of reality within an organisation (e.g. presenting unpopular attitudes to data privacy to staff as normal/acceptable/beneficial for business)

The model described in Principle 6 [10.1.6] describes a simpler model with four key mechanisms through which digital platforms exert their infrastructural power.

8.4.2 The Active Diminishing of User Agency

The second major obstacle to ecosystem negotiability is that platformisation and power exertion are not a one-off transition, but rather an ongoing process. Today's platforms exhibit **a continuing trend of actively diminishing individuals' agency**, especially in the last decade. When software was sold in a box, manufacturers competed based upon which product would let the user take home the greatest range of features and capabilities. New releases with more or better features drove new product sales. But in the cloud computing era, a smaller set of core features done well is sufficient to guarantee an ongoing

subscription revenue from a user. Cost savings in development and support costs can be made by reducing feature sets. Constrained, compliant users are easier to manage. The relentless pursuit of increased profits and further cost saving sees products lose, not gain, features. Interfaces are reshaped to serve businesses' interests first and foremost. Providers' focus on making user behaviours constrained, predictable and profitable, more than meeting their needs or providing maximal value [2.3.5]. Plantin *et al.* describe the particular harmful influence on the ecosystem of Facebook's power exertions:

> *"Facebook is a formidable force in a profit-motivated platformisation which is beginning to eat away at the Open Web. This entails moving away from published URIs and open HTTP transactions in favour of closed apps that undertake hidden transactions with Facebook through a Facebook-controlled API."*—Plantin *et al.* (2018)

Here are just a few examples of the ways in which users' agency has been, and continues to be, diminished:

- Facebook closed their RSS feeds, and later parts of their APIs, meaning that users could no longer consume their friends' posts in any other environment than the ad-filled and algorithmically manipulated Facebook main feed. Later, they eliminated feeds of friends' posts and favourite pages (Perez, 2018), removing users' ability to compartmentalise their content viewing to certain friends groups. The 'Friends' page on Facebook currently shows a list of recommended new friends. To access your current friend list requires an extra click. Encouraging users to grow their networks is prioritised over user convenience.
- Twitter closed the parts of its APIs that allowed real-time notifications and access to one's home feed, killing off primary functionality for a healthy ecosystem of third-party Twitter clients that increased user choice (Newton, 2018). TweetDeck, a major third-party Twitter client was acquired, and later shut down, as was Twitter's own desktop client. Eventually, the only option left to users was to use the web interface. (Gayomali, 2015; Hatmaker, 2018; Siegal, 2022)
- Apple has been diminishing users' agency for a long time. Users cannot open up iPhones even to change the battery without invalidating their warranty. Apple have removed disk drives, headphone ports, SD card slots and other ports. Certain parts of the hard drive on macOS devices are now read-only and non-writeable by users.

- Facebook recently announced they will no longer store users' historical location data (though they will still use location information) (Pegoraro, 2022). This means users will lose the capability to access historical location records. I would argue this makes it harder for users to see how their location data will be used in future, as there will be no historical log to examine. Data-centric companies can change their practices to limit agency and reduce accountability.
- Online news and discussion site Reddit has removed content access for non-logged in users, and uses deceptive techniques to present advertisements that look like posts from users, and to discourage users from appearing offline. These patterns have been described as *disrespectful design* (Regoje, 2021).
- In an example from the public sector, through my work on the SILVER project [see Section II and IV Introductions] just prior to the introduction of the GDPR in 2018, I heard whispers in at least one local authority of plans to 'shift from getting data collection consent from supported families towards simply informing them of our practices' (in other words, removing citizens' choice). The instinct to further organisational interests over those of the individual appears not to be limited to commercial data holders.
- In a similar vein, TikTok announced in 2022 that it would rely on *legitimate interest* rather than consent when it comes to using users' activity data to personalise the app experience. This removing users' ability to withdraw consent to such use. This plan has subsequently been paused after warnings that this might breach GDPR (Lomas, 2022).

Unchecked, trends to reduce users' agency and further providers' interests at the expense of human autonomy are likely to continue. Today's data-centric systems suffer from a lack of consideration to individual welfare. Data centricity encourages neglect of the human end user perspective, creating potential for harm, as the quote atop this chapter illustrates.

The trend to diminish users' agency is needs explicit targeting if data interfaces are to become more free-flowing (Bowyer, 2018b), and if ecosystem negotiability is to be realised. Somehow, the trend needs to be halted, before it can be reversed. The TikTok example suggests this may only be achievable through regulatory changes.

8.4.3 The Intractable Data Self

The third obstacle to ecosystem negotiability is **the intractable data self**. Data about individuals serves as their *proxy* [Bowyer *et al.* (2018); 5.5.4]. This is their *data self* [4.5.1]. If it is incomplete, inaccurate or unfair—highly likely given the difficulties of representing people in data (Martin, 2007; Cornford, Baines and Wilson, 2013)—this can cause harm (Bowyer *et al.*, 2018; Crossley, 2022). Yet currently, although some legal rights to data correction exist (Information Commissioner's Office, 2018), people cannot practically modify or assert control over this most important version of themselves—the version of them that exists in data. Even when data can be seen, people lack the ability to **exert influence** over their data self [5.6.2; Cornford, Baines and Wilson (2013)], which is necessary for *individual self-determination* (Fisch, 2015). To address this obstacle, HDR reformers should explore giving people a role in the curation of their data self [4.5.3; 5.6.2; 6.3].

The sociotechnical complexities of the concept of the data self has been explored in some depth, for example in the work of Deborah Lupton (Lupton, 2020). Research and innovation on ecosystem negotiability, however, has been very limited. It is easier to find business models and research funding for narrow and well-defined contexts. Without a business motive, only non-profit socially-focussed research organisations such as BBC R&D and Sitra have found themselves well-equipped to explore this problem space. Nonetheless, there is an urgent societal need for individual oversight over one's data self [6.3]. People need to reclaim their data selves, and be given control over their digital lives at the broadest level.

8.5 Obstacles to the HDR Objective of Effective, Commercially-Viable and Desirable Systems

The previous four subsections considered the obstacles to the HDR objectives [7.5]. However, through pursuit of these objectives, and through observation of public and business responses to human-centricity, I observed additional obstacles that affect *all* efforts to make progress towards improving HDR. The main challenge is around building such disruptive systems that are so different from the status quo:

Businesses and individuals will not readily invest time and money in HDR, because it is unfamiliar.

8.5.1 A Lack of Individual Demand

Customers are not demanding HDR capabilities in their lives, and, all but the most socially-responsible businesses do not see value in an approach that runs so contrary to current business models, based as they are on data accumulation and the constraining of customer experiences.

Data is overwhelming, complex, and 'sounds boring'. Engaging with your personal data economy to any degree more than that of passive consumer is hard work. People routinely accept data sacrifice, click through T&Cs and cookie banners and are unwilling (or in some cases lack sufficient technical literacy, comprehension or skill) to do the work of asserting control over their digital lives. There is not a clear demand for holistic digital life management and control. Research in this this and at Cornmarket suggests that even if human-centric information systems and more inclusive service interaction practices emerged, people would not be inclined to use them in great numbers. It could seem like hard work or not worthwhile. Just as some people (who can afford it) hire an accountant to manage their finances, we can imagine that some would prefer not to have to manage their own data. This obstacle affects all HDR improvement approaches. Indeed, this is why many companies in the emergent PDE economy [2.3.4] struggle to find a business model. There are clear benefits, but better HDR does not appear to something a mainstream audience will pay for. This should not deter disruptive innovation nor diminish the potential value for such tools. As automobile pioneer Henry Ford famously said, *"If I had asked people what they wanted, they would have said faster horses."* Nonetheless, it is a clear overarching obstacle, which Principle 7 [10.1.7] attempts to confront. Whatever human-centred information systems we can design, they must, above all else, serve human values and relieve pain points. To be appealing enough to drive adoption, they must offer people new digital life capabilities that are not available in any other way.

8.5.2 Closed, Insular and Introspective Practices

The kind of life-spanning, unifying interfaces described in Principle 7 [10.1.7] are nothing like the interfaces that are built today, as they span across different providers' data and services. This highlights the secondary obstacle that all HDR system builders will face, whichever objective they wish to target: **closed, self-interested organisations** with **a lack**

of interoperability. Building an HDR system will necessarily involve connecting to systems of different providers that have different touchpoints into an individual's life and world. Yet most companies act in closed, introspective and non-cooperative ways to further their own interest. Companies like Apple, Amazon, Microsoft, Facebook and Google (the so-called *'big five'*) build **proprietary, incompatible silos** or *'walled gardens'*—sub-Internets that pretend that the alternatives do not even exist, in order to encourage a flow of money and attention to their own products and services. In doing so, they fail to recognise users' holistic needs [2.3; 6.1.3]. In fact, large companies such as these often have, through a combination of apps, devices and platforms, multiple touchpoints into individuals' lives, giving **them** a more holistic view that they can exploit for advertising, without making it available to us [see Facebook's *world2vec* system in 9.4.3 for example].

Commercial motives encourage digital service providers to get users to spend time in their own proprietary spaces (so that resultant ad revenue can be captured and further activity data can be collected) and in order to maintain subscription revenues it is in providers' interests to make it hard for individuals to leave or switch providers. In effect, providers build for a world that does not exist, where every individual is imagined to only interact with that single company's interfaces. I would argue, for example, that Google's venture into social networking with Google+ did not succeed because it failed to build for a reality where most people and their friends were already on Facebook.

8.5.3 A Lack of Organisational Investment in HDR

One can understand why companies are not motivated to build holistic, open experiences. There is little incentive to open up the ecosystem when the free flow of information and of users might result in loss of income. Users with negotiability would be more able to leave. And this also encourages keeping users in the dark [5.5.2]. The less agency and negotiability that users have, the more freedom the provider has to do exactly what they want with their data. In this context, users are *'docile bodies'* (Foucault, 1975) or *'pathetic dots'* (Lessig, 2000).

The tendency of organisations to work in closed, introspective ways and to be resistant to opening up data or services is not solely motivated by commercial reasons: the public sector has a vastly complex, closed and fragmented ecosystem [Pollock (2011); Copeland

(2015); 4.1.2]. Our efforts to build a system to share health data with support workers for the SILVER project [see Section IV Introduction] proved hugely challenging. Sometimes the challenge was a technical one–incompatible data formats that are hard to reconcile, or data being stored in legacy systems with no public API that would allow programmatic access to that data, or issues around licensing. But data sharing agreements also have be established, especially in the public sector which is by its nature more liable to scrutiny and accountability. On top of these technical or procedural issues, there was *resistance to change* data processes and an unwillingness to share data between agencies, often motivated by a fear of legal repercussions. **Data-centrism encourages insular thinking**: it encourages organisations to codify the world into their own systems, processes and formats for their own use (e.g. Figure 9.12).

8.5.4 A Lack of Interoperability

Yet, for effective HDR, **data needs to be separable from services**. The more users' data is tightly coupled to specific services, the less agency users have and the harder it is to build life-centric systems. In BBC R&D's Cornmarket project, attempts to build an interface for users to import data from multiple popular Internet services proved to be a hugely complicated endeavour, requiring access to many different APIs or manual exports and imports of data by users. There needs to be greater interoperability and greater establishment and adoption of **standard formats for exchanging human information** (as distinct from establishing standards for data or service-specific APIs). As mentioned above, platformisation breaks the Open Web (Plantin *et al.*, 2018). To overcome this, companies must be persuaded that human-centric thinking, interoperability and transparency has not just social benefits, but business benefits too. In the absence of such openness, a subculture of *adversarial interoperability* has arisen, where activists, facing a lack of support, force connections to providers' systems in ways that are not approved (Doctorow, 2019).

8.5.5 Insufficient Machine Understanding of Human Information

At an abstract level, the technical obstacle is one that has always faced the tech industry, which is that there often is no universally agreed way to represent important concepts–in this case human-centric information concepts such as events, social media posts, website visits, location history information, app activity, etc. And any entity that does create a

standard then faces the challenge of trying to persuade others that their standard is the best one to use. In general, standards work best when established by non-commercial industrial standards bodies (for example the World Wide Web Consortium (W3C) or International Organisation for Standardization (ISO)) and then mandated through policy such as European Union law. Such standards much be established with input from industry experts.

Even after addressing the obstacles of end-user buy-in and the technical complexities of building human-centric systems, data-driven corporations, motivated as they are by profit and business success (and smaller online organisations too) need to be persuaded of the business value of transparency, interoperability and human-centricity. This is explored further in 9.6.

In summary, whichever of the above four HDR objectives are targeted, all HDR reformers involved in building HDR systems must:

- create, adopt and co-ordinate around **new standards** for human information storage and management
- invest in systems that elevate computers from data-processing machines to **human-information-processing machines**, and
- make a persuasive case to both businesses and individuals that the new approach offers **tangible, previously unavailable value**.

Principle 8 [10.1.8] highlights that what is needed is a different way of thinking about how we build computer systems: Computers need to be taught to understand how the data the hold relates to the human world, so that they can more easily unify, unite and draw inferences across datasets.

8.6 Summation of Chapter 8: From Obstacles to Opportunities

This chapter has presented, in effect, **a map** of the HDR landscape. It has described the major obstacles to better HDR including invisible, inaccessible, scattered, immobile, unmalleable, or unrelatable data; the complexity of current personal data ecosystems; a lack of metadata and machine understanding; the ongoing exertion of power by introspective data holders to diminish user agency; and a lack of demand and investment in

HDR. Figure 8.1 showed an overview of how these different obstacles might be understood to relate.

This chapter can enable HDR reformers to 'hit the ground running' with an understanding of some of the challenges that exist, and insights that may suggest possible strategies to tackle them. A good high-level understanding of the landscape combined with some specific ideas should be valuable for anyone working in the HDR space. Therefore, Chapter 9 expands further on these understandings and insights, presenting four specific and detailed strategic approaches to tackling the obstacles in this chapter.

Chapter 9. Practical Approaches to Improve HDR

"The world is working exactly as designed. And it's not working very well. Which means we need to do a better job of designing it."

—Mike Monteiro, author of
Ruined by Design: How Designers Destroyed the World, and What We Can Do to Fix It

The objective for Chapter 9 is to expand further on the map of the HDR landscape in Chapter 8 so that we might chart a course towards the desired goal of improving HDR in society. Unlike its predecessor, this chapter is solution-focused, and considers the nuts and bolts of *how* we might begin to pursue HDR's objectives [7.5] of data and ecosystem awareness, understanding and negotiability.

Having established a baseline understanding of key obstacles to the pursuit of HDR in the last chapter, Chapter 9 describes what I have learnt about how the HDR agenda [Chapter 7] might be operationalised in practice, bearing those established obstacles in mind and building upon the design insights and Principles theorised in Figure 8.1 and detailed in 10.1. In developing and modelling specific approaches to the operationalisation of HDR, through the discussion in this chapter, I draw from multiple sources:

- the findings from the participatory track [Chapter 6];
- self-experimentation, background research and explorations in the Case Study contexts (SILVER and GDPR);
- prior design, modelling and innovation work on understanding and changes and diminishing agency in the digital landscape;
- my practical experiences from the embedded projects [see Section IV Introduction];
- the theoretical framing and research agenda of HDR laid out in Chapter 7; and
- the established understanding of key obstacles in Chapter 8 and the proposed Principles to tackle them in 10.1.

Across these different activities, I have informally adopted numerous stances, including but not limited to: power user, challenging customer, policy analyst, adversarial designer, legal rights tester, social activist, technical strategist, prototype developer, social commentator, information architect and user experience designer. Of course, my expertise and knowledge varies in these different capacities. Thus, the work leading to the approaches suggested below can be see as multi-disciplinary, speculative work, rather than an empirical or deterministically-reproducible output. Nonetheless, as much as possible I have attempted to ground the approaches in the practical experiences of the projects or other activities I have been involved in and convey examples from my own or others' work to the reader so that the approaches are well-justified.

In the following sections, I present four different 'flavours' or approaches to HDR reform which I have observed:

- **Discovery-Driven Activism** - see 9.3; informed by: Hestia.ai & Sitra's digipower project
- **Building the Human-centric Future** - see 9.4; informed by: BBC R&D's Cornmarket project, prior design & innovation work, and networking with PDE companies like ethi, CitizenMe, and digi.me.
- **Defending User Autonomy and Hacking the Information Landscape** - see 9.5; informed by: self-experiments, DERC's Healthy Eating Web Augmentation Project, and prior personal research.
- **Teaching, Championing and Selling the HDR Vision** - see 9.6; informed by: all four embedded projects, Case Study Two, past software industry experience and prior personal research.

As stated in the conclusion of Chapter 7 [7.7], the approaches listed here are not intended to be definitive, exhaustive or prescriptive. Instead they serve only as examples of what could be done in the pursuit of better HDR. The approaches are not mutually exclusive; elements of them could be combined and many other approaches might exist. Approaches 1 and 3 are more activist and grassroots-based, while approaches 2 and 4 are more conformist and work within established patterns for pursuing organisation and social change.

These approaches, and their underlying Principles [10.1], are documented in sufficient detail such that they, or variants of them, could be subsequently exploited or used as a starting point for action planning by HDR reformers and practitioners, and as such Chapters 7 to 10 constitute a rich, clearly-defined and actionable research agenda to drive future research and innovation for a better, more human-centric digital world. As with Chapter 8, in this chapter I continue develop new Principles for pursuing HDR in practice, which are mentioned here and detailed in 10.1.

The four approaches are described in sections 9.3 to 9.6. Each of these sections begins with a diagrammatic representation of that approach as a trajectory of change. These diagrams use a model known as *Theories of Change (ToC)* which is explained in 9.1 below. Also provided in [9.2] is some additional background information about PDS projects including *Databox*, *Solid* and *Hub-of-All-Things*, which are relevant to Approach 2.

9.1 Additional Background: Theories of Change (ToC)

To provide a structure for cataloguing the insights conveyed by this section, I use a *Theory of Change (ToC)* framing. ToC is a set of methodologies is commonly used by philanthropists, educators and those trying to improve the lives of disadvantaged populations (Brest, 2010). The theories can be used in different ways including planning, participatory design and field evaluation of the effectiveness of new initiatives. There are many different implementations, but common to most of them is a focus on explicitly mapping out desired outcomes (Taplin and Clark, 2012) with a clear focus on **who** is acting and whether the change being brought about is a change in **action**, or a change in **thinking** (Es, Guijt and Vogel, 2015). In this chapter, ToC theory will be used in a very limited way, not as a methodology but simply to provide a visual and structural frame for proposed changes. Using ToC to perform evaluation of the effectiveness of proposed change approaches in action in society would be well beyond the scope of this thesis. The frame of ToC is however a useful way to map out the different approaches to changing the world in pursuit of better HDR.



Figure 9.1: Theory of Change [ToC]: The Four Dimensions of Change[18]

[18] Diagram used here unchanged from *Hivos ToC Guidelines* (Es, Guijt and Vogel, 2015, p. 90) under a CC-BY-NC-SA 3.0 license, whose authors state that this diagram was adapted from earlier work by Wilber (1996), Keystone (2008) and Retolaza (2010, 2012).

Figure 9.1 illustrates the aspects of ToC thinking that are important when using this frame. Specifically, desired changes can be broken down into:

- ***Internal changes***: changes in thinking, feeling, reasoning, understanding, attitudes or identity.
- ***External changes***: changes in actions, behaviour, interactions, structure, policy, technological capability, processes and the external environment.

At the same time, desired changes can be broken down into:

- ***Individual changes***: changes to individual thought or actions
- ***Collective changes***: changes to the thoughts or actions of groups of people together, or to the systems, practices and norms of society at large.

These two splits produce four dimensions of change, and form four quadrants representing different types of change, which are shown in Figure 9.1 and described here:

- ***Individual/Internal (II)***: This top-left quadrant represents changes to what individuals know and understand, and to how they think, feel and plan to act.
- ***Individual/External (IE)***: This top-right quadrant represents changes to how individuals' relationships with others; acting (or being enabled to act) differently in their daily lives and when interacting within society.
- ***Collective/Internal (CI)***: This bottom-left quadrant represents changes in the shared knowledge of groups of people or to the collective identity or values of social groups.
- ***Collective/External (CE)***: This bottom-right quadrant represents changes to the structures and procedures within which people operate, including technology, law, societal norms and communications.

Key to ToC thinking is the idea that making changes in one quadrant can stimulate change in others; for example, collective learning about data attitudes and practices, such as the research conducted in this PhD (lower left quadrant), could inform the design of new technologies, interfaces or processes (lower right quadrant), which if built could make new structures available to have an impact on improving individual-provider relationships (upper-right quadrant). The changes to those relationships could then in turn lead to

individuals thinking and feeling differently (upper left quadrant), for example feeling more empowered or having greater awareness of data practices. The diagram at the start of each of the following subsections 9.3 to 9.6 illustrate that approach's trajectories of change through the ToC quadrants.

9.2 Additional Background on Personal Data Stores: Databox, Solid and the Hub-of-All-Things

Before explaining the four approaches, and in particular Approach 2 ['Building the Human-centric Future'–9.4] it is important to acknowledge some additional developments in the personal data economy space beyond those outlined in 2.3.4. Building upon common ideals of personal data ecosystems to control one's personal data and influence how and when it is used, three notable initiatives have emerged and developed over the last decade in academic and entrepreneurial contexts:

- The *Hub-of-All Things (HAT)* project, which aims to provide an online 'Personal Data Account' PDS into which individuals can store their personal data.
- The *Databox* project, which is based around a PDS software that could store users' personal data at home for secure processing and personal use.
- The *Solid* project, which offers *Solid Pods*, open-source personal data store software which allows users to keep and harness their personal data.

The importance of these efforts and others in this space such as Mydex and NextCloud, has been acknowledged by the European Union's Data Protection Supervisor, which names these (confusingly in light of the earlier use of this same term 2.2.2) as Personal Information Management Systems or PIMS (Attoresi and Moraes, 2020).

The Hub-of-All-Things project was founded by Professor Irine C.L. Ng of the University of Warwick (Ng, 2018) and collaborators. A proof of concept was built in 2016 and the initiative continues to offer an online service today at hubofallthings.com through a private-public collaboration between a commercial entity called Dataswyft and the non-profit HAT Community Foundation which acts as a regulator. The HAT project includes a consent and data sharing framework and an app ecosystem designed to encourage the emergence of

new services and apps based upon the platform. The HAT project has produced a set of 11 principles (Ng, 2018):

- Co-production access rights without lien [i.e. without debt or obligation];
- Alienable rights;
- Non-rivalrous consumption;
- Expansibility;
- Excludability;
- Data Derivatives;
- Data as Store of Value;
- Data as Medium of Exchange;
- Transparency;
- Trust Anchoring; and
- Market Design.

The Databox project was founded in 2015 by a consortium of researchers from the Universities of Cambridge and Nottingham as well as Imperial College London, including two of the authors of the original Human Data Interaction vision paper (Haddadi *et al.*, 2015). The premise was to pursue the HDI ideals of agency, legibility and negotiability by creating a physical box that could serve as a trusted platform from which individuals could control, manage, explore and query their personal data (Haddadi *et al.*, 2016). The design of the Databox emphasised privacy-preserving data processing, and secure data sharing using cryptographic techniques. Like HAT, it developed an ecosystem to encourage app development, and sought wider collaborations across academia and industry to drive adoption of the technology. BBC R&D ran a project as part of their 'Living Room of the Future' study, where they developed Databox into a physical consumer-friendly hardware device called the *BBC Box* (Forrester, Thompson and Jones, 2016). The Databox has also been positioned as a tangible embodiment of the data protection and accountability principles that the GDPR seeks to embody, including data portability and data access, that could be especially applied in the Internet of Things (IoT) space (Urquhart, Sailaja and McAuley, 2018; Urquhart, Lodge and Crabtree, 2019). In this IoT context the Databox

project has also been examined in terms of its potential to yield economic value for individuals who might wish to sell their own data (Perera *et al.*, 2017).

The Solid project was founded by Sir Tim Berners-Lee, creator of the World Wide Web, and is operationalised through his company Inrupt (Lohr, 2021; Berners-Lee, 2022). By establishing protocols and standards for personal datastores, known as Solid PODs (Personal Online Data Stores), users can be empowered to choose different providers of PDS hosting services. As with Databox and HAT, the goal is to create an ecosystem of apps, services and providers. The decentralised approach of Solid aims to create a healthy ecosystem where users have agency and negotiability to change data service providers and move data freely as their needs evolve (Mansour *et al.*, 2016; Sambra *et al.*, 2016; Solanki, 2021).

Taken together, these initiatives align well with the MyData ideology for a more human-centric personal data landscape, and can be seen as attempts to operationalise the call for 'an ecosystem of data trusts' working through public and private initiatives in order to rebalance digital power (Delacroix and Lawrence, 2019).

9.3 Approach 1 to Improving HDR: Discovery-Driven Activism

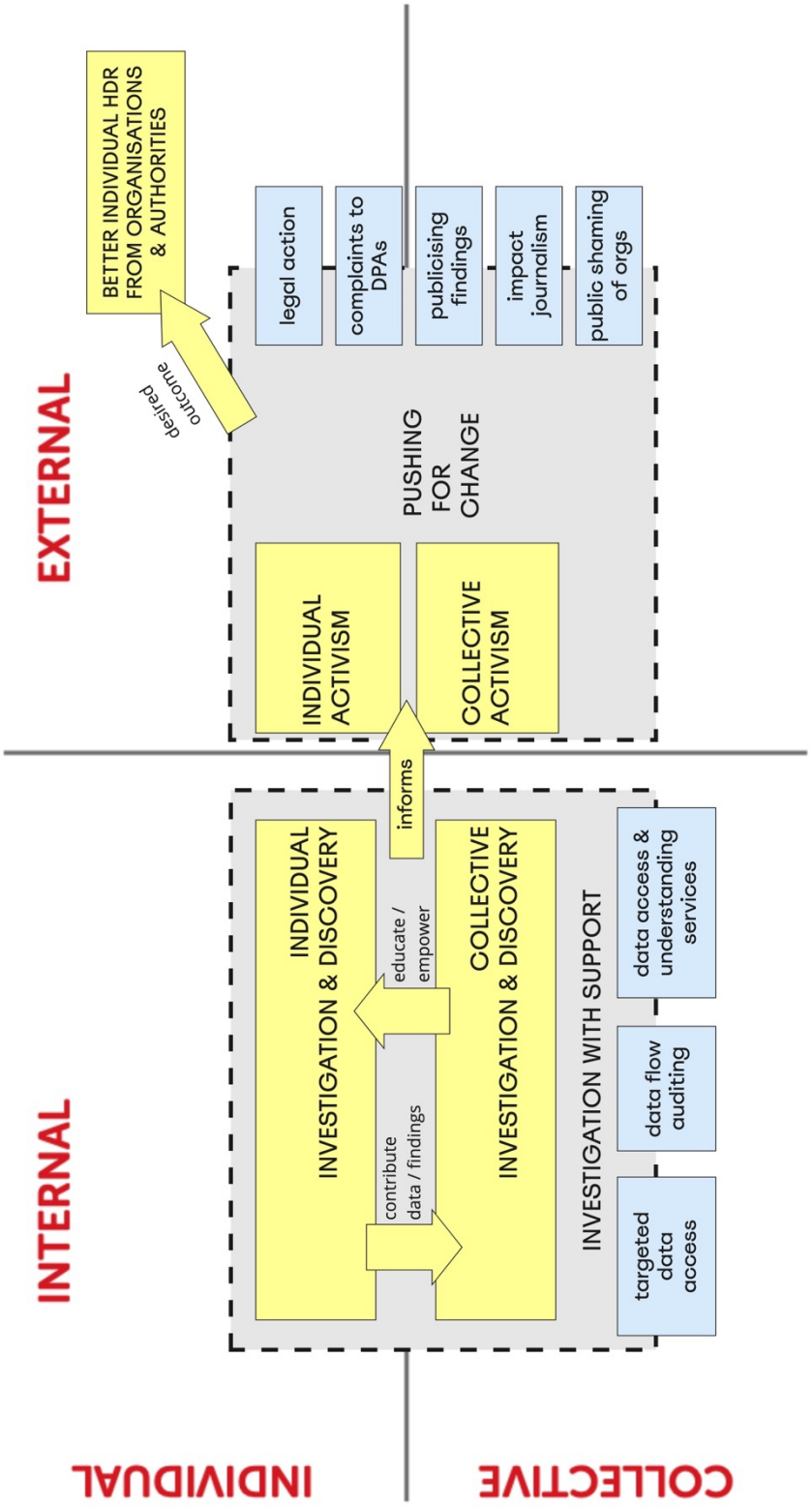


Figure 9.2: HDR Approach 1: Discovery-Driven Activism

The approach to HDR reform presented in this section, and illustrated in Figure 9.2 above, embraces the *activist* aspect of being a recursive public [7.6], as well as the individualist idea of reconfiguring one's world [3.1]. This approach focuses on the realities of the current data-centric provider ecosystem, and is focused upon deeply understanding it so that it can be challenged from a grounded position of strength. The approach, which would be applied typically to a single service provider, app or platform (which could also be a public sector service) is fourfold:

- **Gather Evidence**: Using all available means, gather personal data and information from the target organisation.
- **Determine Current Practice**: Analyse the gathered data and information, using targeted approaches to dig deeper where needed, to establish facts about how the target organisation stores and handles personal data.
- **Identify Desired Change**: By consideration of the specific practices observed from the analysis, specific problematic practices, gaps or areas where improvement or change is needed are identified.
- **Appeal for, or Compel, Others to Act**: Using all available means, directly or indirectly persuade or force the target organisation to execute the desired change.

I describe this approach as **Discovery-Driven Activism**. The discovery phase, which aims simply to establish facts about the past or current data practices of the target organisation, can be broad ('let's see what we find', as in the *digipower* project [see Section IV Introduction], or highly targeted, such as when *The Citizens* (a non-profit pressure group in the UK who *'use impact journalism to hold government and big tech to account'* ('The citizens - about us', 2020)) used Subject Access Requests to investigate a breach of personal data by the Labour Party in the UK (Colbert, 2022).

9.3.1 Obtaining Your Personal Data

Subject Access Requests and ***Data Portability Requests*** (Information Commissioner's Office, 2018) are two of the most powerful tools for this kind of investigation. Freedom of Information Requests have previously been used to obtain otherwise hidden data and information from governments and public sector organisations (BBC News, 2014). Similarly, these new data access rights are beginning to be used to force commercial

organisations to release personal data or information about data processing. There are challenges in non-compliance, as discussed in 5.5.2 and 5.6.1, but the ability for the individual to ask very broad or very precisely targeted questions and to be able to threaten a complaint to a Data Protection Authority (backed by a potential fine if their question is not answered) is a significant new power that HDR reformers can exploit. This new power is summarised by Principle 9 in 10.1.9: Individual GDPR requests can compel companies to change data practices.

Facebook's *Download Your Information (DYI)* tool, mentioned in Principle 9 [10.1.9], represents a useful class of tool in the arsenal of the activist HDR reformer. Along with Google Takeout, it is one of number of ***data download portals*** that allow users to download their own data. Since GDPR's introduction in 2018, an increasing number of large online platforms including Facebook, Google, Apple, Netflix, Twitter, Spotify, Uber, Instagram and Strava, faced with the need to reduce the cost impact of GDPR request handling for their large user bases, have developed and augmented online self-service portals available where users can download a copy of their personal data. This has some advantages over Subject Access Requests in that data can usually be obtained within minutes or hours rather than taking up to 30 days, but has some disadvantages in that the data returned is a voluntary offering by the company, that may not cover the data that the individual is seeking or entitled to and does not provide any ability to ask follow-up questions. This technique was sometimes used as a fallback means to obtain data in Case Study Two, and was used more strategically in the digipower project [See Section IV Introduction], which also examined its merits and limitations (Bowyer, Pidoux, *et al.*, 2022).

9.3.2 Data Flow Auditing

Both access requests and download portals rely on the organisation in question to be transparent, accurate and thorough in their provision of information, but an alternative technique of *data flow auditing* allows individuals to investigate and collect data on the *actual* behaviour of a target organisation. This was used effectively in the digipower project. Using an Android app called TrackerControl (Kollnig, 2021), a service provider's app can be monitored while the user is using it normally, to see which servers or domains that app is contacting (and one can imply, exchanging data with). Apple has recently introduced an equivalent function on iOS, known as *App Activity Reports* (Apple, 2022), providing iPhone

users with the same ability as part of the phone's operating system. Both tools have limitations, in that the content of the data exchanges is not known, but can be valuable to verify claims made in privacy policies or GDPR responses. These tools can also *generate questions* for further investigation, for example by identifying third parties such as data brokers with which the target organisation may be sharing personal data. This technique is described further in (Bowyer, Pidoux, *et al.*, 2022), along with a comparison of the different techniques of data flow auditing, data download portals and data access requests.

In general, what the discovery-driven activism approach highlights is that there is a role for pro-active citizens and citizen journalists to play in challenging the power of data-holding organisations by treating those organisations as a subject of investigation, both in research (Walby and Larsen, 2012) and in the pursuit of improving civic society (Schrock, 2016). As a specific example, I am currently participating in an investigation with The Citizens ('The citizens - about us', 2020) by making GDPR requests to help them explore the data-broking and voter profiling practices of political parties in the UK, and supported a similar earlier research effort by the Open Rights Group by sharing my political party GDPR request responses with them.

9.3.3 How Activists Pursue Change

Once information has been obtained, the activist HDR reformer can use a variety of means to try to bring about the desired change:

- If a target organisation fails to comply with a data access request, or a demand to erase or correct data, they can be *reported to the appropriate Data Protection Authority*. In some cases even the threat of this (which can carry a large fine) can be enough to compel the organisation to change.
- If a breach of law is found, the target organisation could be *taken to court*, as seen in the Schrems case above, which resulted in new legislation that Facebook had to comply with (Kuchler, 2018). As well as individual cases, this also often happens in the form of class action lawsuits, as with Facebook and Cambridge Analytica (Bowcott and Hern, 2018). Increasingly, unethical or illegal data practices are being challenged. In some cases, such extreme measures are not needed, however, as many

companies do want to improve their transparency to customers, even when they find it hard to do so.

- Simply *making data available to the public* can be empowering to society at large. This approach has been demonstrated by the UK website *TheyWorkForYou*, which increases democratic accountability of MPs by making MPs' votes and public statements more readily accessible (mySociety, 2004).
- As well as structured *impact journalism* such as that conducted by *The Citizens* as mentioned above, another technique available to individual activists is *public shaming of misbehaving organisations*, especially on Twitter. While the ethics of this are complex and it does not always succeed, the technique has been used effectively to force organisations to change, in order that they might avoid further bad publicity (Silver, 2014; Braw, 2022).

A key insight about how activists can work most effectively in the personal data / individual power space is captured in Principle 10 [10.1.10]: that by working together in *collectives*, individuals can pool and compare their data and learn more than they can alone.

9.3.4 Data Access and Ecosystem Understanding Services

Having identified that there is a trajectory where individuals and collectives can obtain data to empower them, it is clear that this complex work can be supported. We see the emergence of what I call *data access and ecosystem understanding services*, with entrepreneurs and activist enthusiasts:

- creating *tools to help people get their data back* from organisations, such as Tap My Data ('Tap my data', 2022), AccessMyInfo ('Access my info', 2022) and My Data Done Right ('Privacy salon & my data done right', 2022);
- creating *tools to help people understand their personal data*, such as Ethi (Jelly, 2021) and digipower.academy ('About digipower.academy', 2022); and
- helping collectives and journalists with *training, workshops and skills development* to take advantage of their data rights and *understand data ecosystems*, like Hestia.ai (Dehaye, 2019), *The Eyeballs* (Foucault-Dumas, 2022) and *PersonalData.io* (Pidoux, 2022).

Such emergent endeavours, given the support that enables them to flourish, could make other HDR reformers using the discovery-driven activism approach more successful by ensuring that a lack of legal, technical or investigative skill does not become a barrier to any HDR practitioner or citizen wanting to use this approach.

This approach shows that there is a role for independent actors and organisations to carry out discovery-driven activism–access requests, complaints, legal challenges, public campaigns and more. Discovery-driven activism can empower individuals and collectives to incrementally work towards building the world of better HDR that this thesis outlines.

9.4 Approach 2 to Improving HDR: Building the Human-centric Future

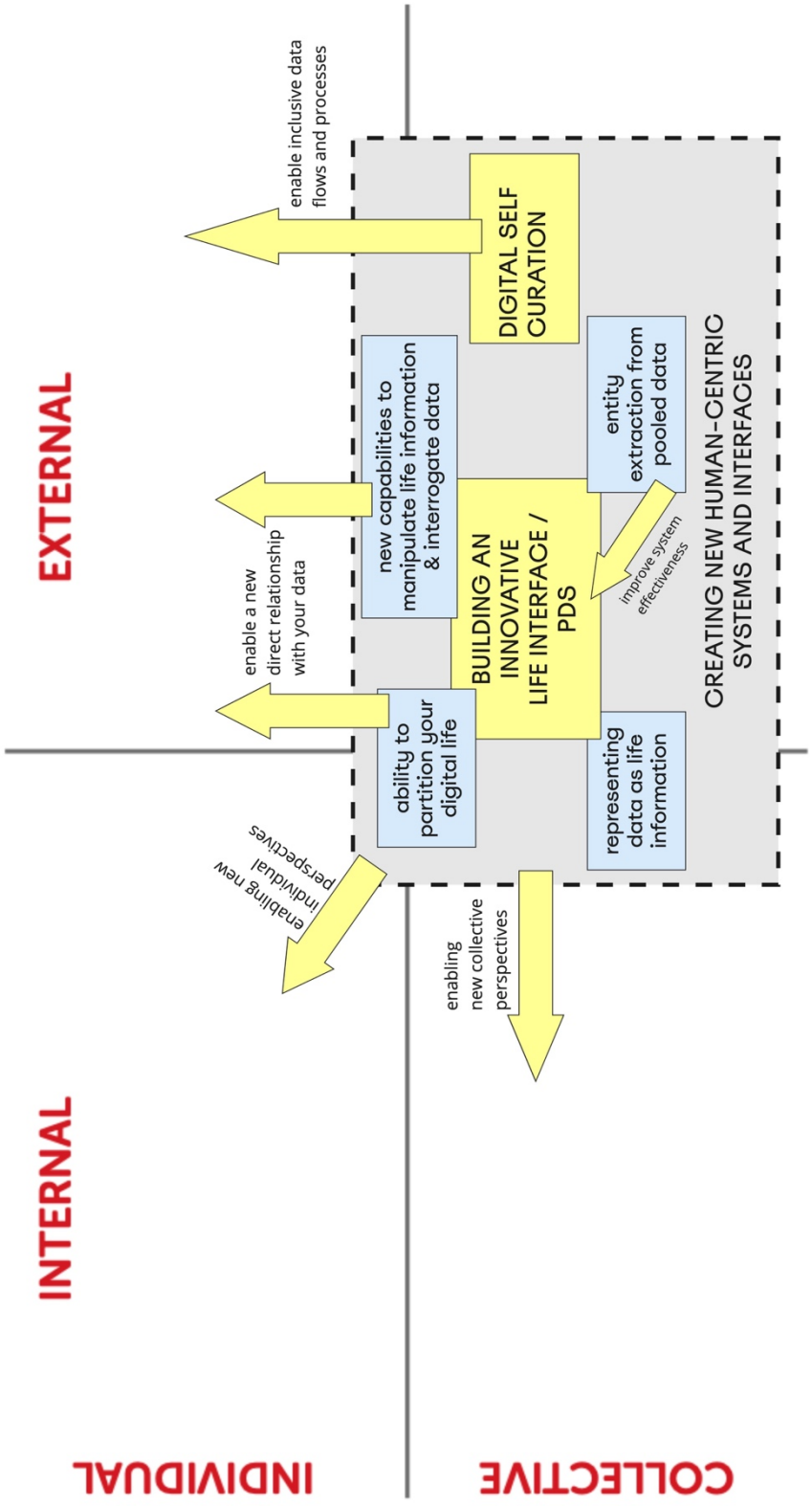


Figure 9.3: HDR Approach 2: Building the Human-centric Future

The approach to HDR reform presented in this section, depicted in Figure 9.3 above, focuses on the gaps in individual data interaction capability that exist today. The objective here is to design and build proofs of concept for **novel human-centric information systems** that can deliver people **new capabilities over their data**. In this approach, the focus is more introspective than Approach 1 [9.3]: it is about how the individual can improve their relationship *with* data in the context of their own digital life. The bulk of this section describes **specific design ideas** developed by myself and colleagues at BBC R&D during my 2020-2021 research internship on the Cornmarket project. As established in Principle 2 [10.1.2], one of the most promising models for giving people a new and improved relationship with their data is to create a place where one's scattered (Abiteboul, André and Kaplan, 2015) personal data can be stored and aggregated *in one place* (W. Jones, 2011a). Based on Principle 3 [10.1.3], that place should offer individuals the means to use their life and ecosystem information as a material they can explore, examine or repurpose as they see fit. This leads to the vision of a Personal Data Store [2.3.4] into which one can unify the data from the different parts of your digital life (as depicted in the AllOfMe vision video ('AllOfMe.com Teaser Clip', 2008)), and meet public demand for 'control over your data' (Teevan, 2001; Hartman *et al.*, 2020). The Cornmarket R&D project sought to develop a human-centred (i.e. non-commercial) PDS proof of concept, as shown in the conceptual model I developed for the BBC Cornmarket project depicted in Figure 9.4.

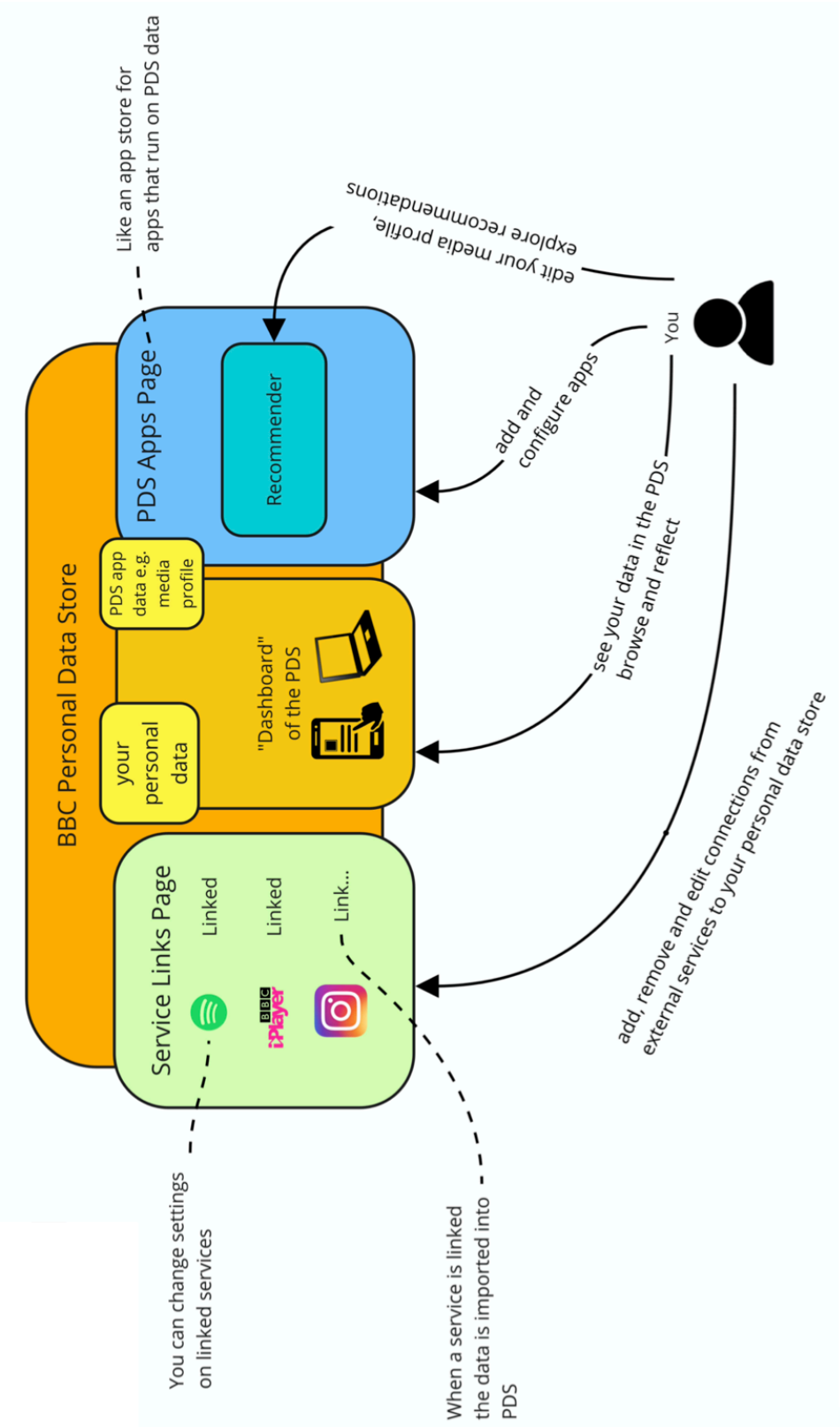


Figure 9.4: Conceptual Model for a Personal Data Store System

9.4.1 Life Interfaces

The first challenge in designing such as a system is to consider what data will be stored, and what data can be represented as *life information* [7.1; 7.4.1]. At a high level, I identified a number of different types of data that a user might wish to store in a PDS:

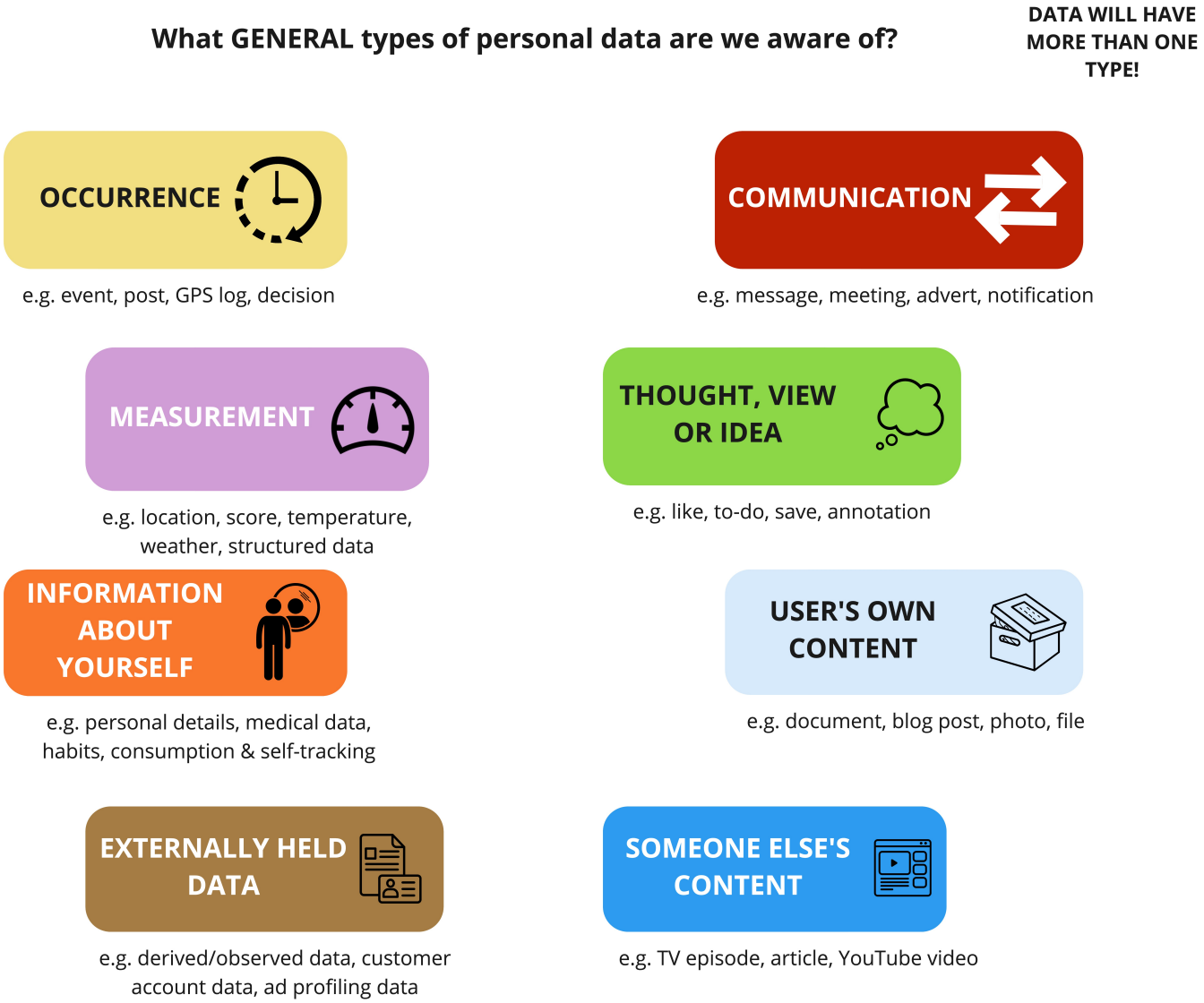


Figure 9.5: High Level Data Types

Then, considering the earlier observation by myself (Bowyer, 2011) and proponents of temporal PIM systems [2.2.2] that time can be used as a unifying concept for personal information, I considered how we might represent occurrences within personal data as *happenings*, anchored against a particular point in time. In effect, this would focus on different properties of the data, much like Karger's *lenses* [Karger *et al.* (2005); 2.2.2].

Figure 9.6: Life Information Modelled as Happenings

In modelling data as life information using the concepts shown in Figure 9.5 and Figure 9.6, it is important to come up with as simple a model as possible, so that the ***life interface*** is not overwhelmingly complex and unmanageable. Over several iterations, I was able to reduce the modelling of life information in a PDS down to the following four types - *activities*, *interactions*, *transactions* and *datapoints* [Figure 9.7]. These could correspond to four views or lenses within the user interface.

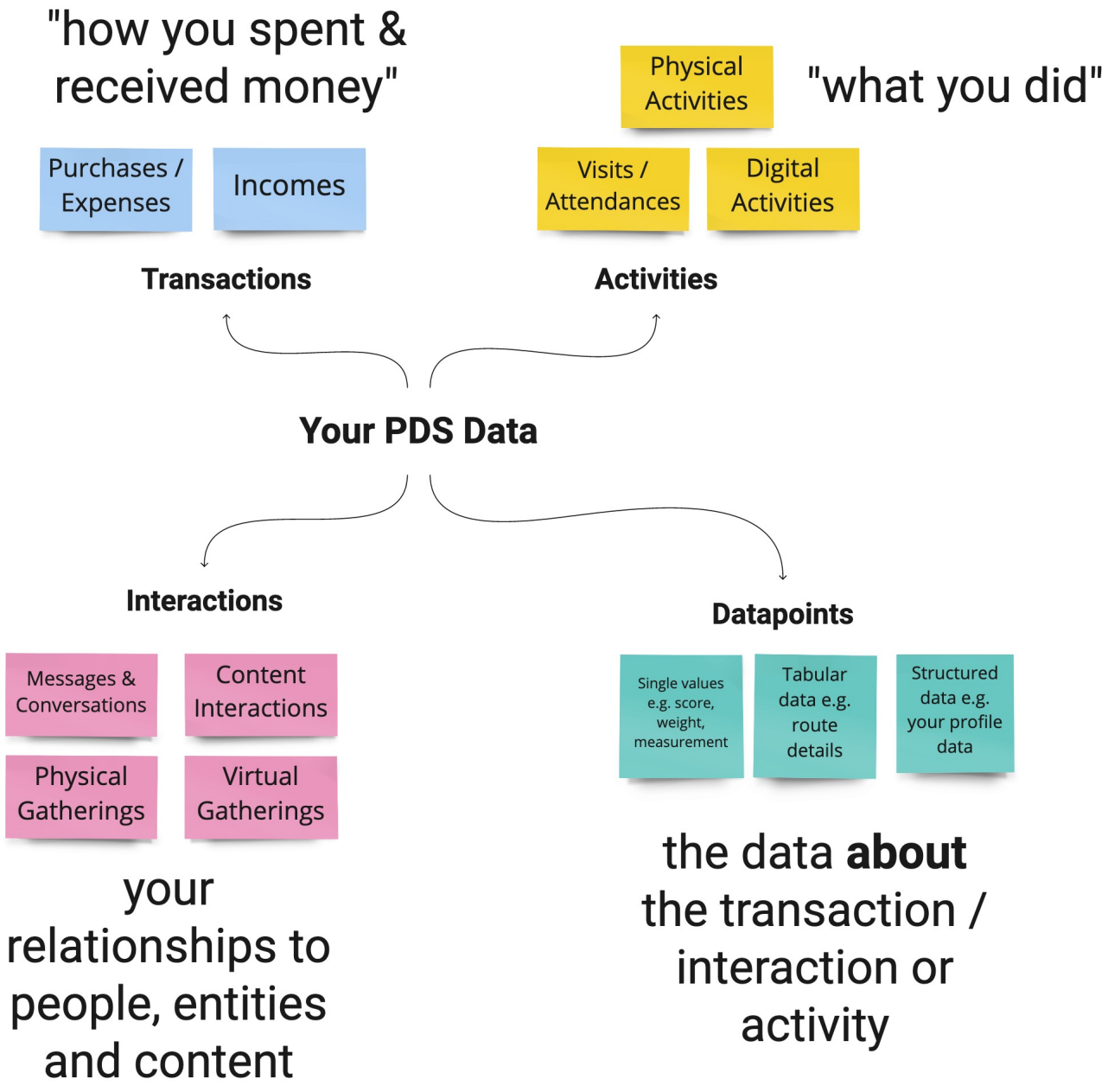


Figure 9.7: A Simple PDS Life Information Presentation Model

Having decided upon models for the information within a PDS, it becomes much easier to design visual presentations of that information. Figure 9.8 shows a design mock-up designed by Alex Ballantyne of BBC R&D of some of the views of life information that one could offer within a PDS:

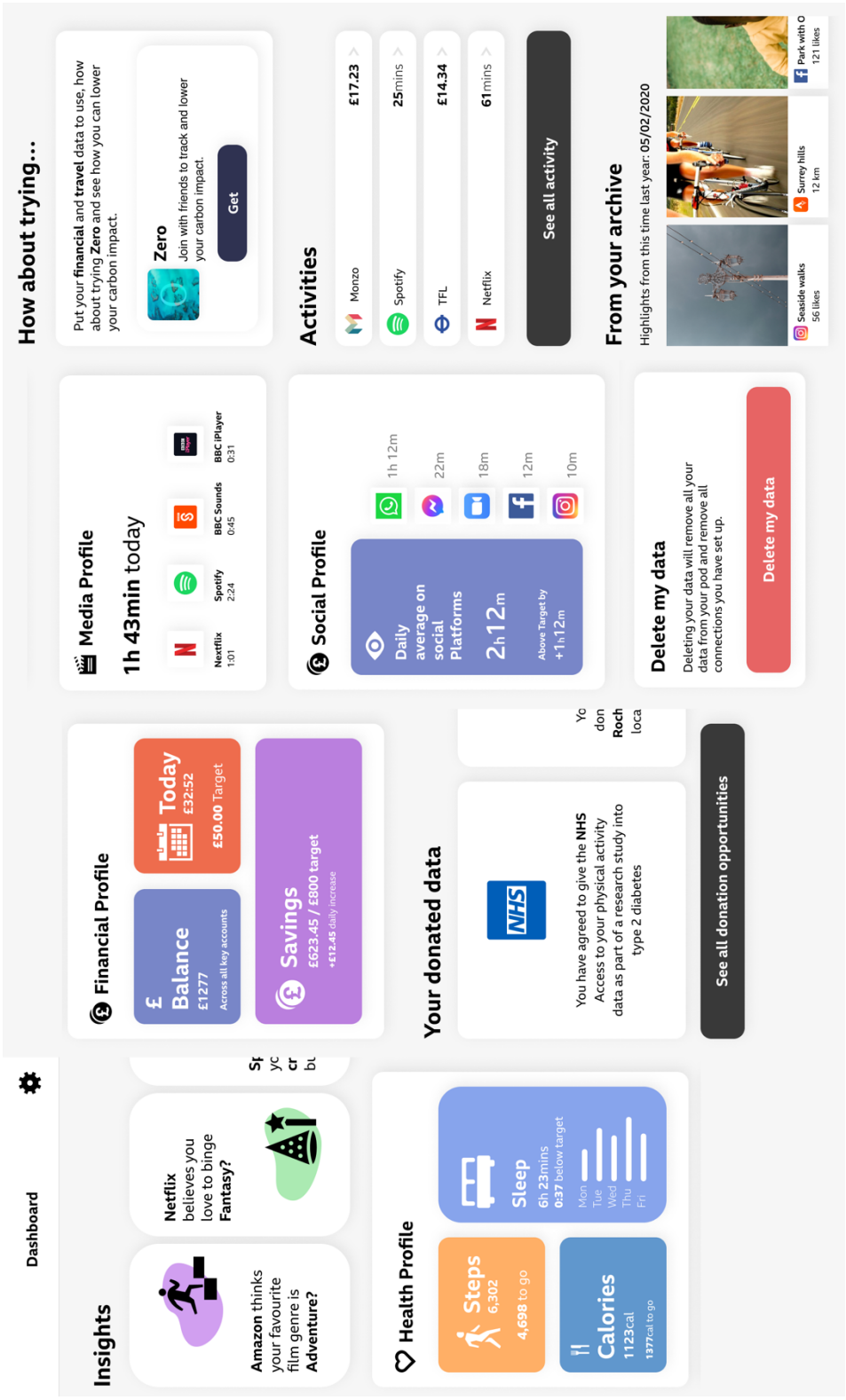


Figure 9.8: Mock-up of Life Information Presented in a PDS Interface

9.4.2 Life Partitioning

What has been shown in Figure 9.8 is the beginnings of imagining what a *life interface* might look like, a single holistic interface covering all aspects of one's digital life.

Previous HCI design work [2.3.1] has focused on, in the first wave, the functionality of the machine; in the second wave, on the common tasks of a work environment; and in the third wave, on classes of users and the commonalities of experiences in everyday life. But the challenge designers of life interfaces face is even more difficult. As Lindley noted, no single arrangement of information will suffice because in the same context two people may have different needs [Lindley *et al.* (2018); 2.2.2]. Because of the *subjectivity principles* [Bergman, Beyth-Marom and Nachmias (2003); 2.2.2], we are now faced with the need to build an interface that is suited to the individual, even though every individual has *different needs*. The design requirements for the interface are likely unique to the individual. Therefore, any life interface design must be able support *different mental models* - and as the life sketching exercise [Figure 5.2] in Case Study Two showed, people have many different ways of compartmentalising their life.

During a life interface design hack week at BBC R&D, colleague Jasmine Cox and I explored this problem. We considered that a key element of the interface design would need to be the versatility to **partition one's life in different ways** according to your own worldview, and then to be able to assign different collections of life information to the different partitions. As a visual illustration of one possible mental model for life partitioning, consider Figure 9.9, where, based on an idea I had had years earlier, I modified a Cluedo™ board to represent the 'rooms of your digital life':

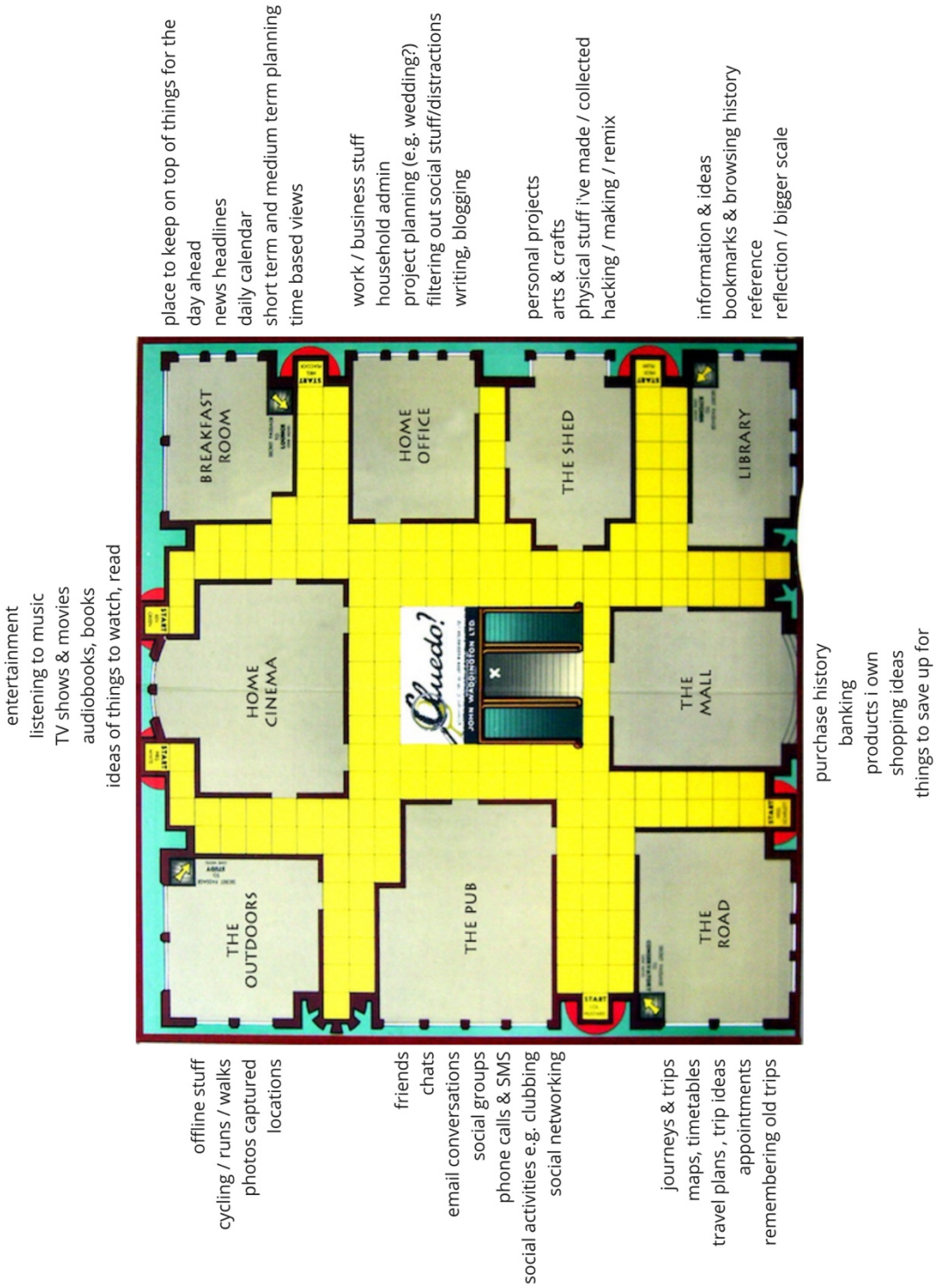


Figure 9.9: Life Partitioning Analogy using a Cluedo™ board[19].

[19] Cluedo board design is a copyright of Hasbro, Inc., fair use applies.

We also imagined other mental models, such as partitioning your life according to parts of the body (heart for relationships, body for fitness, brain for current projects, etc.) or a landscape with forest, lake and buildings representing different aspects of your life.

Whichever visual metaphor is chosen, the important thing is the functionality–being able to use these partitions to *filter* [2.1.4] your life information and focus on a particular *perspective* [Lansdale and Edmonds (1992); Krishnan and Jones (2005); 2.2.2]. Deciding to focus on a particular aspect of one's life is analogous to fixing a *conceptual anchor* [Teevan (2001); 2.2.2].

We then produced a mock-up app workflow for assigning different elements of your life (such as people, places or topics) to different partitions of your life, and viewing a visualisation of those different partitions in some subjectively meaningful form before filtering on a particular life partition and then being presented with a *timeline* [2.2.2] of life information pieces associated with that part of life. This is illustrated in Figure 9.10, with artwork by Jasmine Cox:

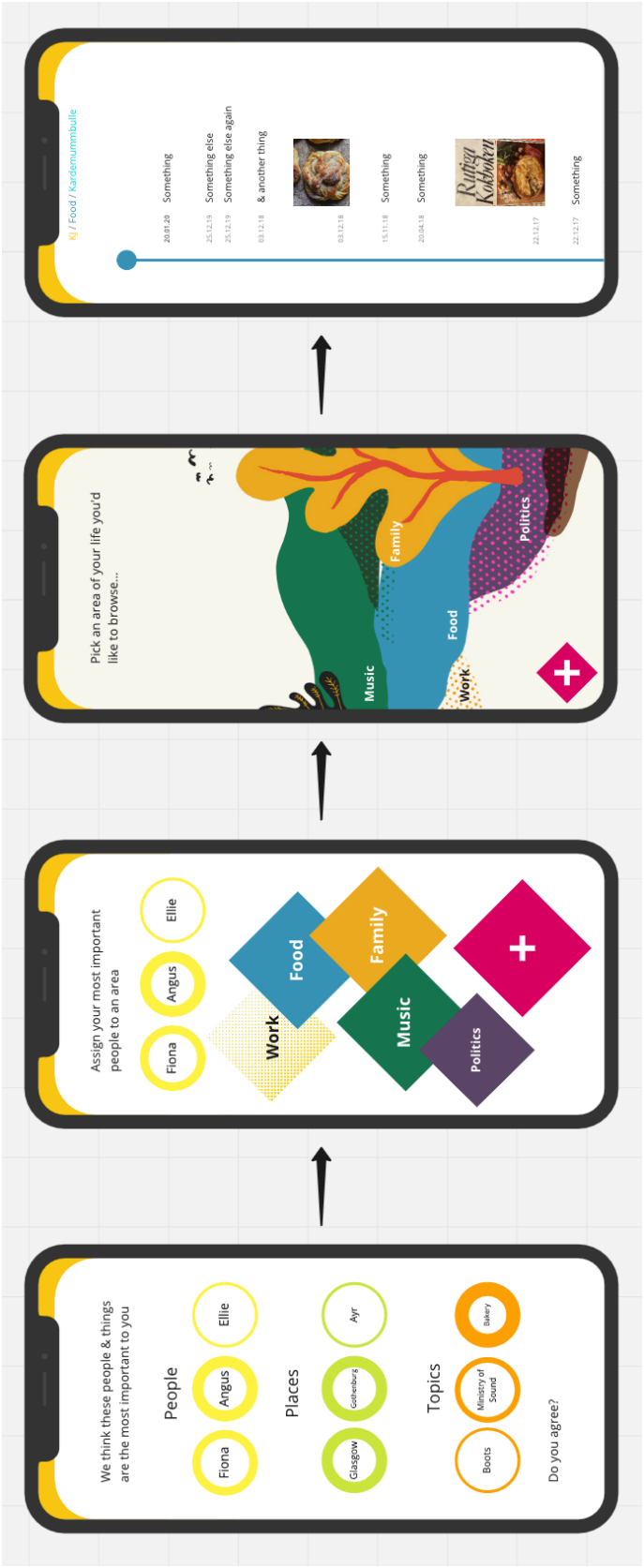


Figure 9.10: Mock-up: Browsing by Areas of Life

During a Cornmarket hack week, I coded a proof of concept for how users might be able to partition their digital life. This is detailed in ARI7.1 and shown in Figure ARI7.1.

9.4.3 Entity Extraction From Pooled Life Data

Principle 11 [10.1.11] offers a practical, theory-informed approach as to how we might start to build systems that are more able to understand the meaning of human information [Principle 8–10.1.8]: by building automatic entity identification systems. This is a technique that is already being used by corporations, both in products (such as semantic text analysis in the enterprise content analytics space) and for the interpretation of personal data. As an example of the latter, Facebook has developed an internal system called *world2vec*, shown in Figure 9.11, through which it mines information about the world from the data that people generate through social media posts, messages and interactions. This information is exploited for commercial benefit (better targeting of social network users with personalised ads). It is about time similar techniques were used for individual benefit and personal data empowerment.

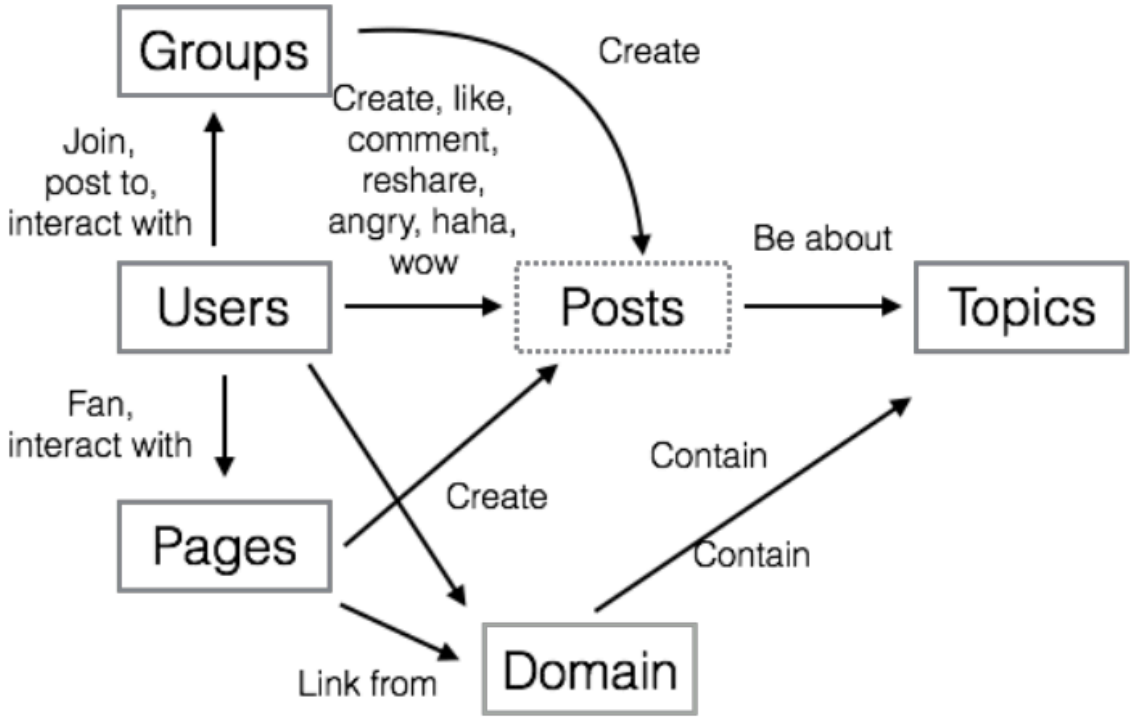


Figure 9.11: Facebook's World2vec Model, Semantically Modelling Human Information from Social Media Posts on Facebook (Peysakhovic and Wu, 2020)

Automatic ‘best guess’ attempts to arrange personal data in a PDS can go further than just association of related entities. In many cases, it is possible to programmatically determine the nature of a piece of data. Figure 9.12 shows a detailed approach for identifying and classifying data automatically in a PDS, drawing on the previously identified data types in Figure 9.5, the data attributes in Figure 9.14, and the entities identified in Figure 10.10. In Figure 9.13, a more abstracted approach is shown looking at how data could be classified according to the four different life information concepts in Figure 9.7.

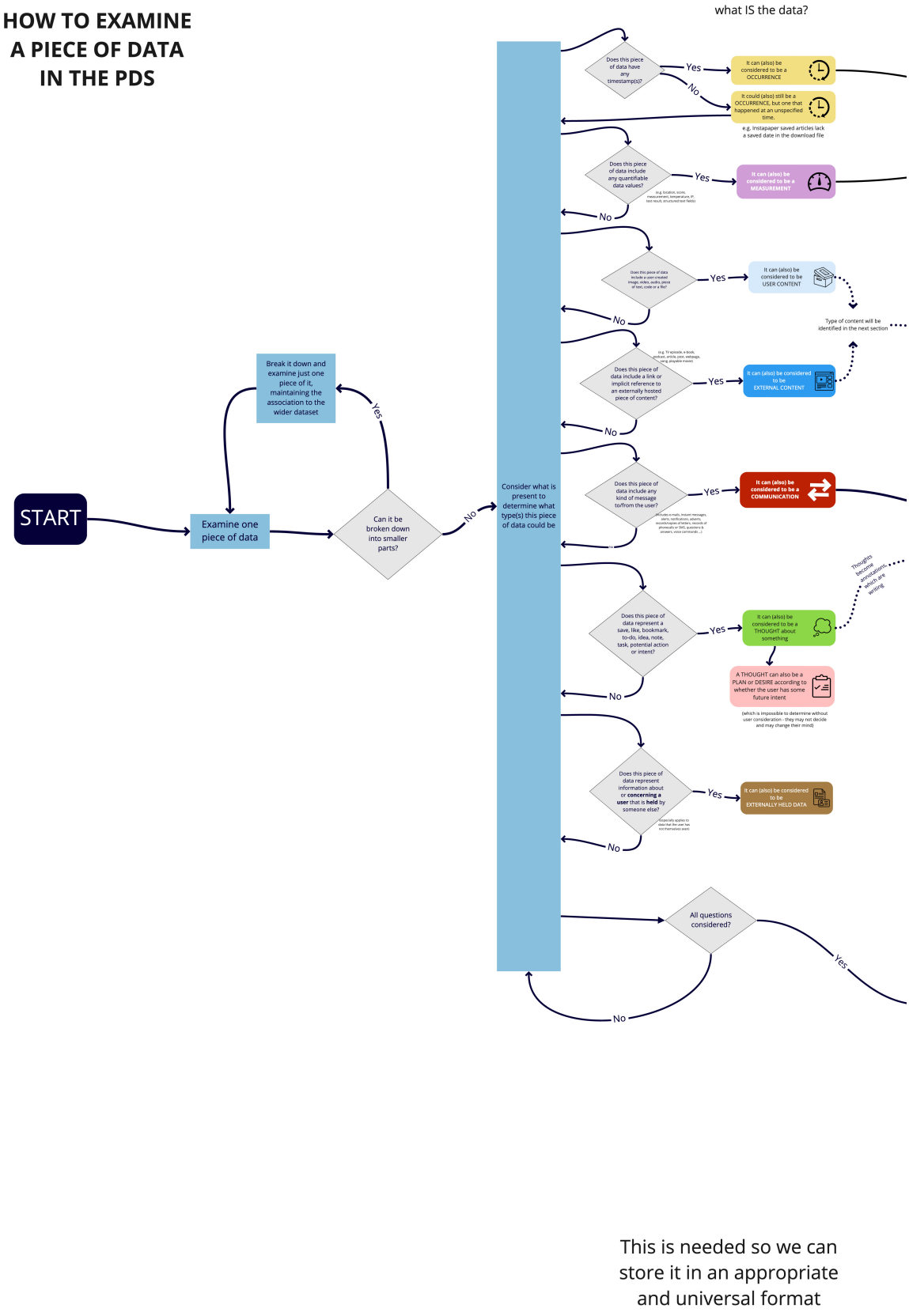


Figure 9.12: Identifying the Attributes of Data (left side)

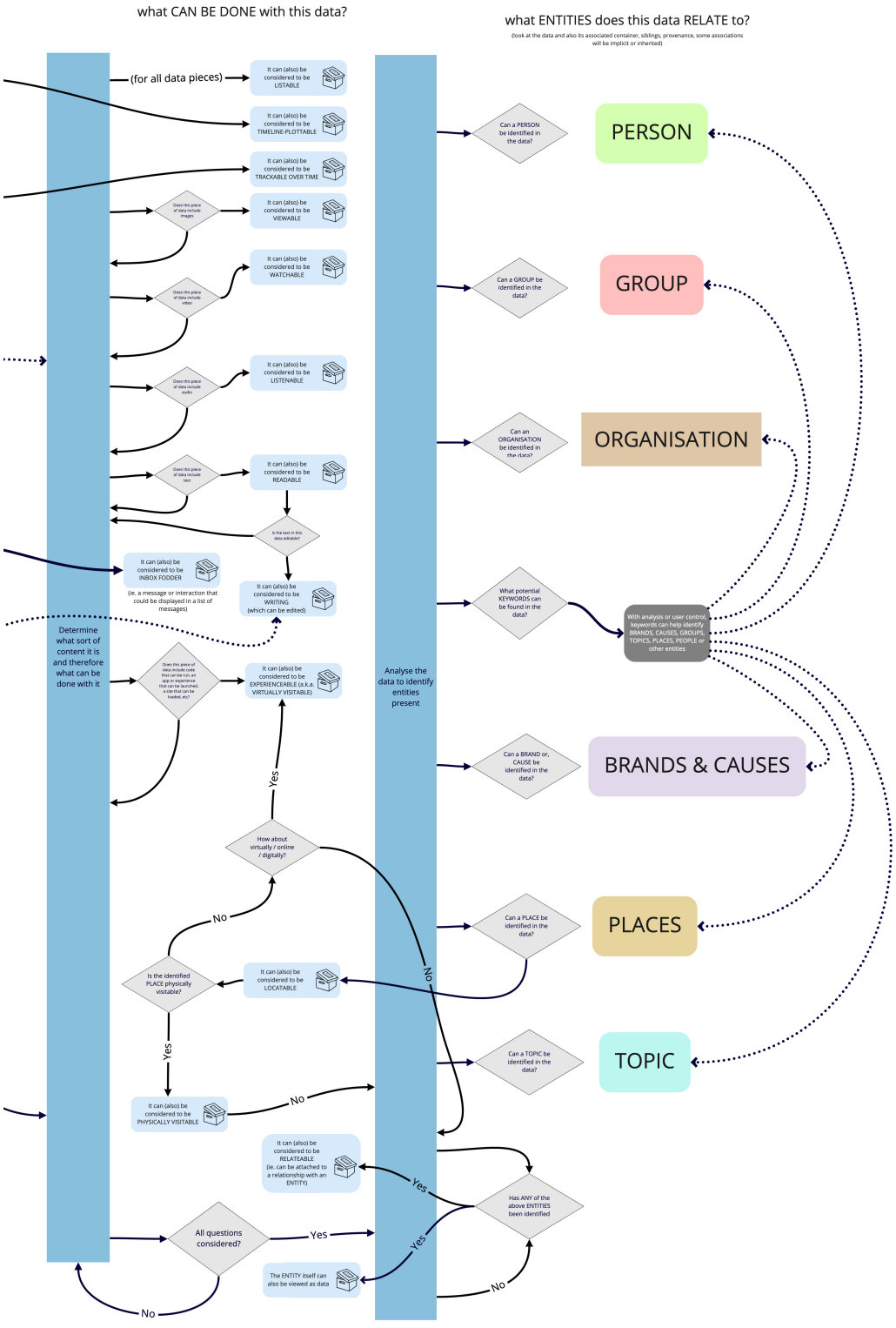


Figure 9.12 (continued): Identifying the Attributes of Data (right side)

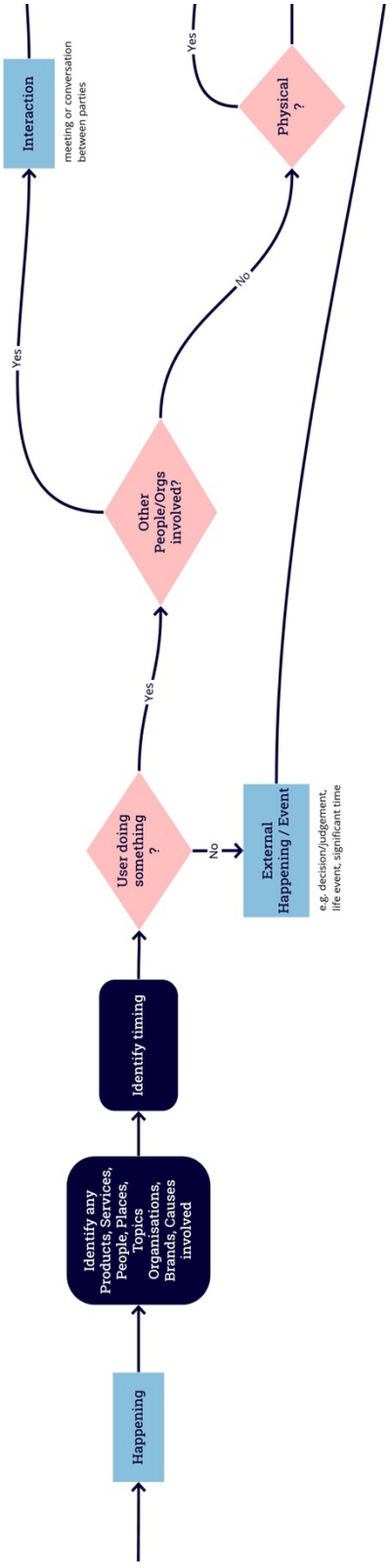


Figure 9.13: Determining the Nature of a Piece of Data (left side)

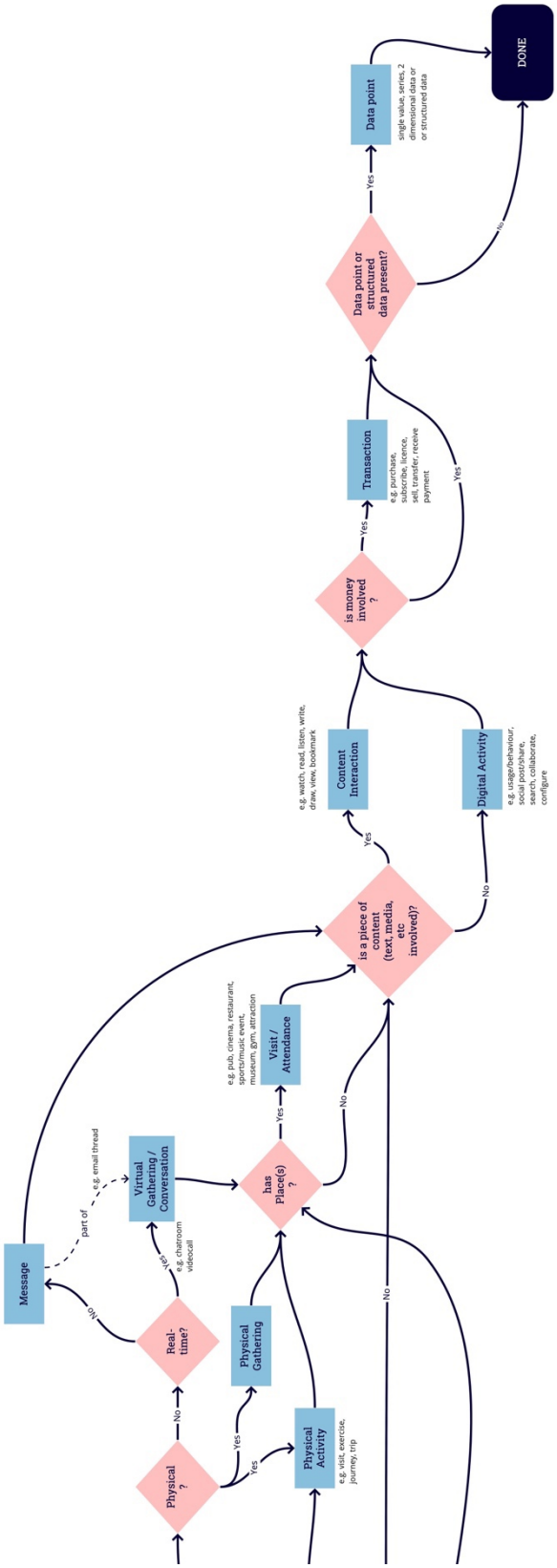


Figure 9.13 (continued): Determining the Nature of a Piece of Data (right side)

As part of a hack week while at BBC R&D, I built a working prototype performing basic entity identification within GDPR personal data returns. See ARI7.1 for further details.

9.4.4 Life Information Manipulation Capabilities

Moving beyond the initial challenge of classifying and arranging human information within a *life interface*, a core consideration in designing such an interface is to consider what *capabilities* the user might be given over the pieces of life information represented in the interface. It is through the provision of a wide range of useful operations upon information items that the information will start to feel like *a material* as described in Principle 3 [10.1.3] and deliver the *new capabilities* needed for success Principle 7 [10.1.7].

There are two ways to consider this problem - one starting with the data, and another starting with the user.

Starting with the data, and considering the different types of data that a user might be able to gather from a across their digital life, I identified that each datapoint or dataset is likely to have properties which determine *what can be done with it*. A number of possibilities are shown in Figure 9.14:

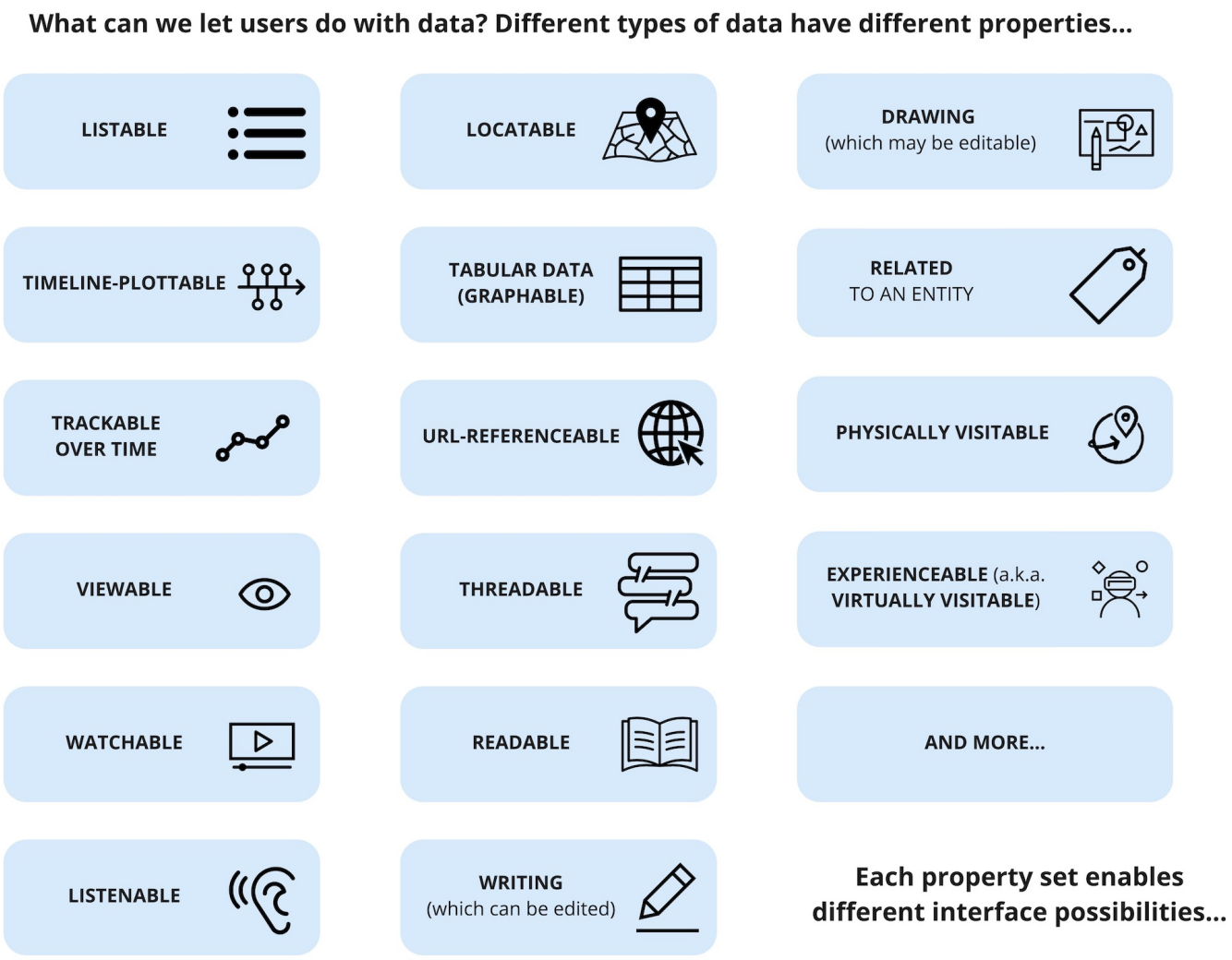


Figure 9.14: Attributes of Data

These attributes could then be used to determine which datapoints would be eligible for inclusion in different views (or *lenses* (Karger *et al.*, 2005)) of the data, such as tables, graphs, maps, lists, media thumbnail views, etc.

If we consider capabilities from the user's perspective, it is useful to recognise that visualisations alone (such as those in Figure 9.8) are not sufficient to meet the need for either interrogability or malleability [8.2]. To address the need for malleability, we need to determine what actions an individual might want to perform on a piece of life information. To explore this design question, we can draw on multiple sources of inspiration:

- actions identified in prior PIM research & design work [2.2.2] such as arranging, browsing, keeping, and associating;

- actions needed for SI [2.2.3] such as combining, history viewing, interpreting patterns;
- actions needed for effective access [2.1.4] such as filtering, zooming, getting detail on demand; and
- actions inspired by the eight lenses of how people think of data [Table 7.1] such as summarising, deriving or remixing.

Having developed a large set of initial design ideas for possible actions, I worked with my BBC colleague Chris Gameson to distil this information design thinking into a 12 *data card* designs that could encapsulate some of the most common *actions a user might want to perform* upon personal data in a PDS. These are shown in Figure 9.15. These had a similar purpose to the cards in my pilot study [1.3.1] and Case Study One [4.3.1], acting as *'things to think with'* (Bowyer *et al.*, 2018).

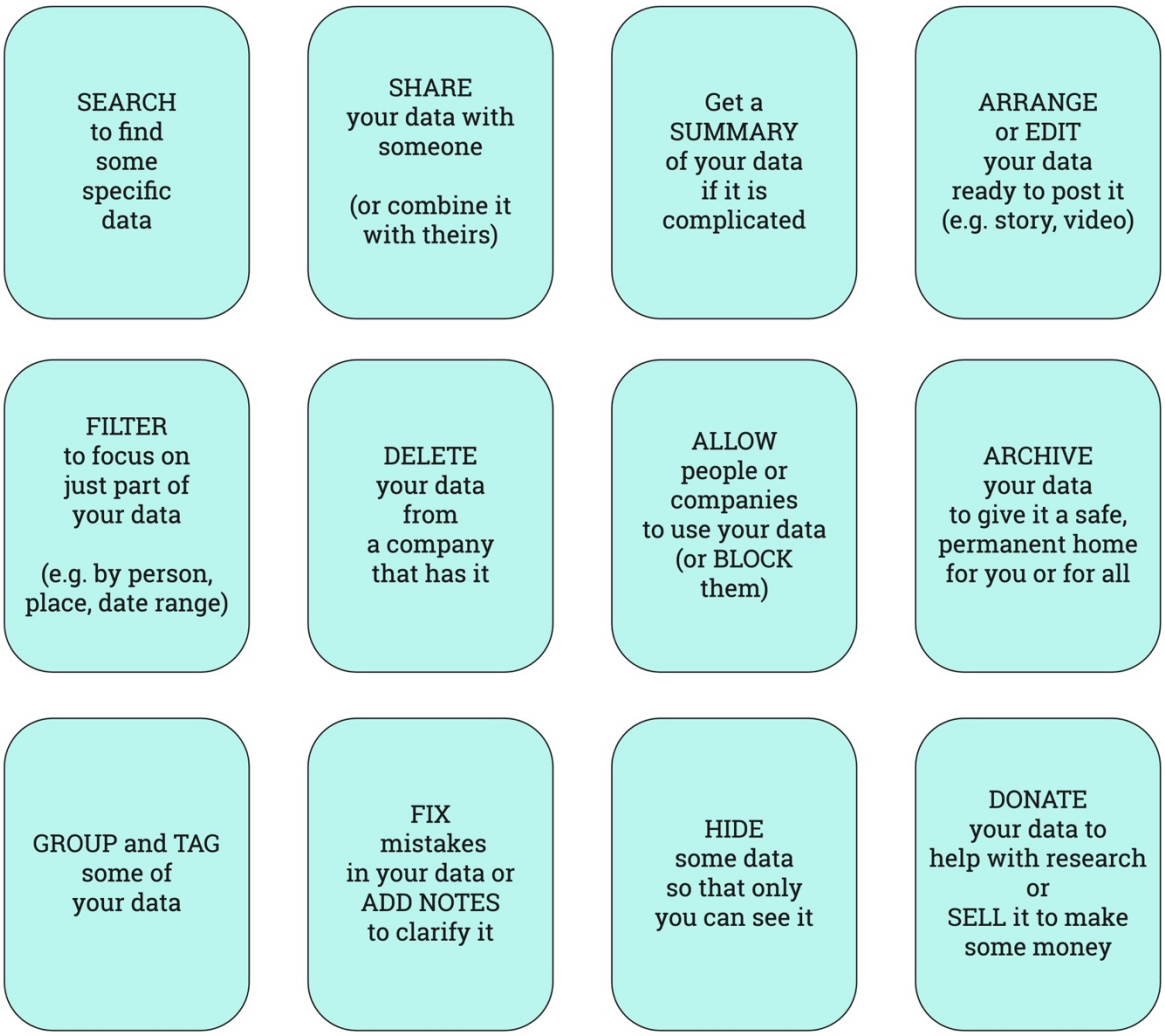


Figure 9.15: Actions One Might Perform on Life Information

As identified in Chapter 5 [5.4.3; 5.5.2] and explored further in 6.1.3, being able to ask questions of data is important to individuals. Addressing this need for interrogability in a similar approach, we produced 10 data card designs to encapsulate some of the different common questions that users might want to ask of their data, as shown in Figure 9.16.

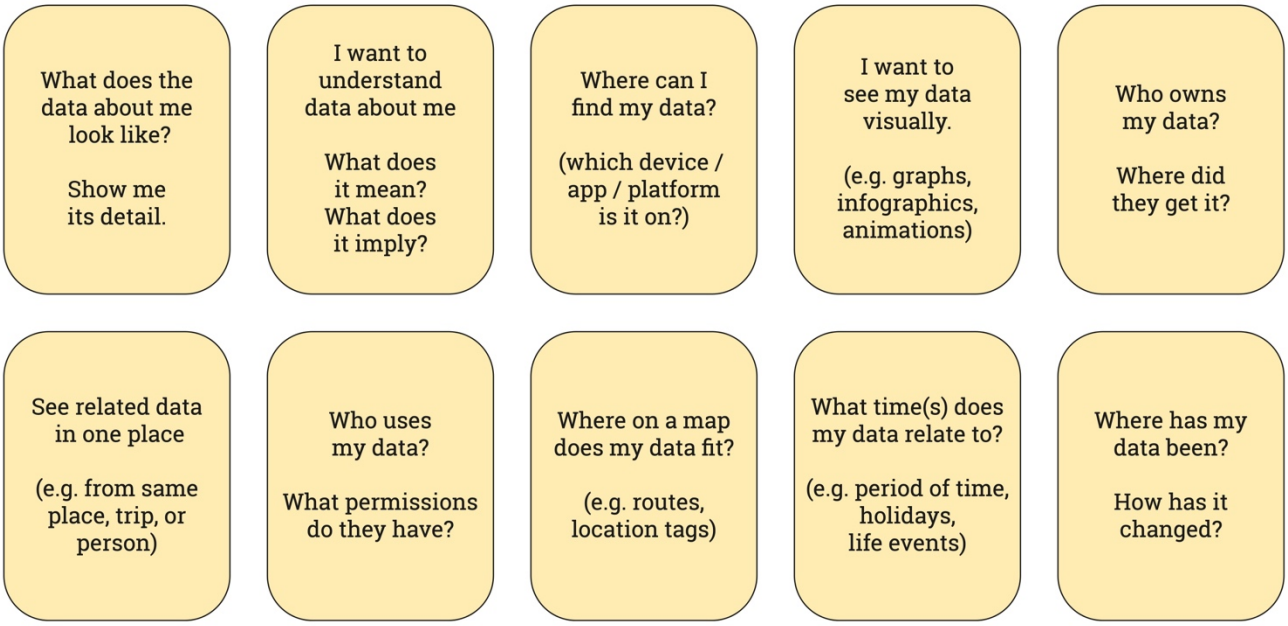


Figure 9.16: Questions One Might Ask of Life Information

These actions and questions act as a design building blocks, a precursor that can inform the design of features for a life interface that could provide the user with new capabilities to interrogate and manipulate the data in a much more interactive way. The cards were refined and subsequently used in a Cornmarket BBC research engagement with young adults.

Once information has been correctly identified semantically (using approaches such as those in Figure 9.12 and Figure 9.13), and associated to entities, it then becomes possible to enrich users' search and browse capabilities by enabling the use of ***facets*** (Televiciute, 2020) to allow users to precisely target a set of data points or a particular data item. An example of how this could be done is shown in Figure 9.17, another design artefact from my work at BBC R&D. This is not a user interface mock-up but rather a three-stage *strategy*, with examples, for the different types of facets that one could offer a user to help them explore their data in powerful ways:

Figure 9.17: Example Taxonomies for Life Information Navigation

9.4.5 Digital Self Curation & Inclusive Data Flows

This final element of Approach 2, unlike the previous subsections, has a broader focus than just the creation of an effective life interface, because it has sociotechnical/business process design implications too. It aims to address the problem identified in 8.4.3–the intractable data self. Drawing upon the philosophy behind VRM [2.3.4], we can imagine (as implied by both Case Studies [4.4.3; 5.6.2]) that were individuals able to **create or contribute to the digital representation of themselves in data**, that this would be more accurate. Furthermore, this would provide an opportunity to address the identified want for *involvement* [6.2.3]. In the BBC Cornmarket project, this was explored through the concept of profile-based recommendations. Functional interfaces were built which allowed users to construct and edit a representation of their own musical/media tastes known as a *'media profile'*, using personal data such as viewing/listening history imported from multiple sources as inputs. This profile would then be used to recommend new content, which would be likely to be better suited to the individual's tastes since it would be based on a more accurate representation of self. This aspect of the Cornmarket project (Orphanides, 2021) is not only a practical manifestation of the ideas of pull-centric/VRM approaches, but is also illustrative of a new way of thinking about personal data, where it can be **directly generated (or edited/fixed) by the individual**, rather than just asserted on inferred by the service provider without verification.

Such thinking could be expanded further to accommodate the idea that data is not static. Once we make the mindset shift from data as a static bundle of facts to **data as an ever-changing flow or *stream*** [2.2.5; 4.4.3; 6.2.2], it becomes much easier to think about how processes and interfaces might be updated to give users a direct role in the evolution and shaping of that data. This also makes it easier to maintain knowledge of data's provenance [Principle 5–10.1.5]. Such ***inclusive data flows*** and processes could yield benefits for users in terms of increased ecosystem negotiability [8.4], but also for businesses, as Approach 4 [9.6] explores.

The future-centric designs and insights[20] I have presented as part of Approach 2 show that while developing interfaces and processes for a better HDR future is a challenging design problem (due to the variety in individuals' mental models and the breadth of data across an individual's digital life), it is certainly possible to tackle these challenges. It is likely that the reasons that few have ventured into this space are largely due to the lack of an obvious business model or incentive rather than any question of the merits of these approaches. It is possible to **empower individuals with new capabilities** over the life and ecosystem information encoded within their data, and even to design **new data interaction approaches** that give users an ongoing role in the curation of that data, and HDR reformers can help make a world of better HDR a reality by **building, or investing, in the building of life-centric technologies and processes**.

[20] The research and design work and insights from my time with BBC R&D's Cornmarket project are documented in more detail in the external publications detailed in 1.3.4.

9.5 Approach 3 to Improving HDR: Defending User Autonomy and Hacking the Information Landscape

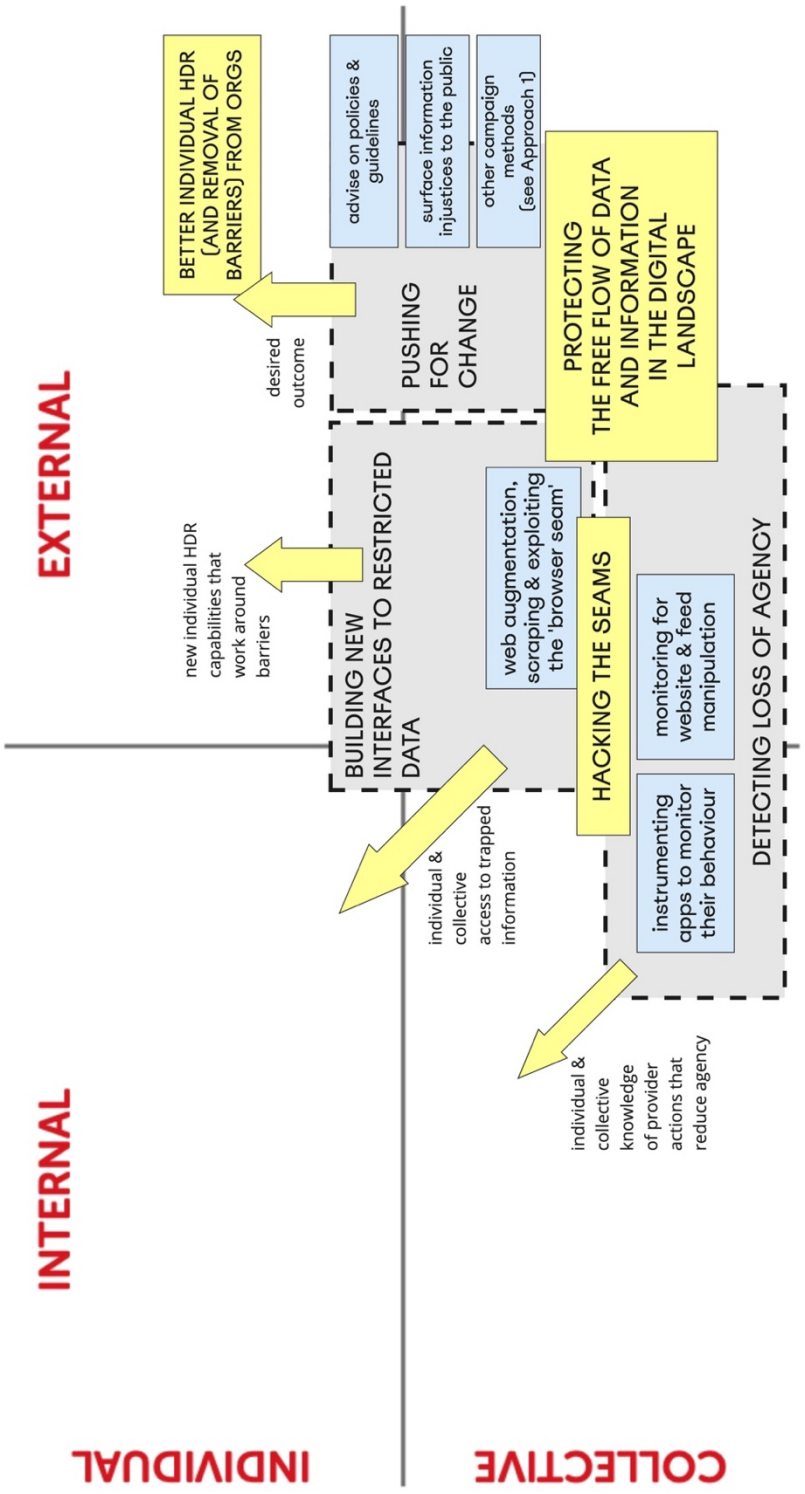


Figure 9.18: HDR Approach 3: Defending User Autonomy and Hacking the Information Landscape

9.5.1 A Deceptive Landscape

While Approach 2 [9.3] focuses on designing a better future, and Approach 1 [9.2] focuses on identifying and campaigning for positive changes in provider practices, there is a need for a more direct approach to tackle an urgent and evident problem head on: the active diminishing of users' agency [8.4]. In pursuit of mass market appeal and cost reduction, there is a trend towards the *dumbing down* of technology, which invariably leads to fewer features and less agency. Increasingly, companies encourage users to think of technology as a black box, which you cannot and should not look inside, as depicted in Figure 9.19, and discourage users from varying their usage, digging deeper or challenging the normal way of using a product:



Figure 9.19: The Modern 'Black Box' View of Technology

This disempowering trend is seen across hardware, software and service industries, where providers attempt to **restrict the ways in which technology can be used** even after it has been purchased, for example:

- cars that can only be serviced by official dealers,
- printers that can only use officially branded ink cartridges,
- smartphones that cannot be repaired without special tools and warranty invalidation,
- media files that cannot be copied or shared,
- online content that cannot be accessed in certain geographies,
- operating systems that block or discourage certain software from running, and
- banking applications that forbid screenshots.

Apple, for example, encourages users to consider technology as *magical*, rather than as tools to be harnessed and understood. This approach is highly problematic for user agency:

> *"Magical design prioritises pleasing and surprising a passive user who can only use the solution as authorised."*—Cristiano Storni (Storni, 2014)

In order to challenge this trend towards disempowered, compliant users and *protect individual autonomy* as well as our ability to make sound judgements and hold organisations to account, direct activism and grassroots civic action is called for. This forms the core objective of this approach, which is represented in Figure 9.18 above: to *identify* ways in which agency is being diminished, and to find ways to *circumvent* the new barriers in order to maintain existing levels of capability.

It is bad enough when a particular technology blocks you from doing what you want to do, but the trend is even worse when viewed at a broader level: As the *digipower* investigation highlighted [Principle 6–10.1.6], the information landscape that we inhabit as digital citizens is being manipulated to **change what we see, what we understand of the world, and what we can do, in order to manipulate our behaviour**, which can have harmful effects on democracy, agency and individual autonomy. Therefore a further goal here is to protect our ability to **experience unbiased and unmanipulated information landscapes**. A recent US Congress investigation showed how the big five use their dominant positions as device manufacturers, search engines or major Internet platforms to promote their own interests (Sisco, 2022). In a paper looking at the growing manipulation of search results, Shah and Bender considered how this landscape could be protected and accountability preserved. They described their approach as establishing ***'guardrails for the status quo'***, and I find this phrase to be an excellent summary of what Approach 3 is trying to do (Shah and Bender, 2022). In August 2022, this issue of the unreliability of the social media landscape again came to public attention, as former Twitter head of security Peiter 'Mudge' Zatko revealed his grave concerns for the security of our information infrastructure and landscape:

> *"Your whole perception of the world is made from what you are seeing, reading and consuming online, and if you don't have an understanding of what's real and what's not, that is pretty scary."*—O'Sullivan, Duffy and Fung (2022)

9.5.2 The Battle at The Seams

We can also take this idea further through the idea of *adversarial design* (DiSalvo, 2012) which advocates using design processes to challenge the status quo. This is not to say that to pursue Human Data Relations requires adversarial design, but rather that in recognition of the sizable force of commercial data-centric norms working against individual interests, that designing in ways that subvert or challenge the status quo is an important tool at the HDR reformer's disposal. People need to be able to adapt their products and their data in ways that better meet their individuals needs, and sometimes that cannot be achieved through convential routes such as asking a service provider to improve or adapt their product. People need to be able to embrace *design after design* (Storni, 2014). They need to be able to *appropriate* technology to their own needs [Dix, 2007; 10.1.12]. Arguably, people have a right to demand technology that treats them more fairly–for example a right to better decision making with data, or a right to fair programs (Edwards and Veale, 2018; Kollnig, Datta and Kleek, 2021). An example of such an adversarial design approach is seen in research undertaken with colleagues from the DERC healthy eating project [see Section IV introduction], where researchers produced a detailed user interface redesign for the Just Eat takeaway interface that could, using web augmentation, be built without reliance upon the provider, in order to better meet users' health needs (Goffe *et al.*, 2021, 2022).

However, such adversarial approaches rely heavily on the availability of ***seams***–places where products can be modified, where data can be observed and interfaces reconfigured. This concept is explained in Principle 12 [10.1.12]. The activity of adaptation at the seam can take many forms, including but not limited to: using APIs, building browser extensions, using screen-scraping techniques, carrying out packet sniffing, or reverse engineering.

But these seams are under threat. As identified in 8.4.2, users' agency is being actively diminished by the closure of APIs and other seams and the removal of features that enable adaptation. Therefore, as Principle 12 [10.1.12] will illustrate, **an unseen battle for the free flow of information is underway at the seams of today's digital products**. HDR reformers are attempting to change the digital landscape to better meet their needs, and service providers are fighting back. The seams of digital services need to be identified, understood, protected and held open, and exploited by individuals. If such work is not undertaken in using adversarial design to challenge the status quo and hold providers to

account, the diminishing of user agency is likely to accelerate, closing doors to HDR reformers wishing to build better Human Data Relations.

9.5.3 Regulating to Free the Information Landscape

The reason that seams are so important, is that because they are a point of friction, a point of resistance to the idea that ***data should be separable from services*** [8.5.4]. One of the goals of HDR, and particularly this approach, is to wrest control of data away from service providers, so that it can be truly free-flowing, as I envision in (Bowyer, 2018b). In this approach, I have identified some of the key avenues for progressing this goal–through web extensions, accessibility tags, API exploitation and reverse engineering. But in the face of such extreme power, it appears that such **grassroots advocacy will not be enough**. For example, web extensions are a powerful approach but are limited in that only those *power users* who install the extension will gain the benefits. They do not directly help the layperson with their diminishing agency. Therefore, as outlined in [5.6.1], HDR reformers must also work to **educate and persuade policymakers of the need for change**. As an example of this, in my Case Study Two paper (Bowyer, Holt, *et al.*, 2022) I called for better guidelines from the EU towards data holders. The European Data Protection Board launched a public consultation on a new set of guidelines they had constructed around GDPR (European Data Protection Board, 2022a) and I contributed a detailed set of recommendations to that consultation, emphasising some of the important insights I have identified in this thesis, including a focus on human information, delivering ongoing understanding, establishing standards, and viewing data as an ongoing flow over time (Bowyer, 2022a). This sort of engagement by HDR reformers can help shape future policies and guidance, contributing expertise to help the civil servants making those rules. As HDR reformers, we must continue to push for better regulation. It is the only force that can significantly change data-holding organisations' practices. Better regulation is needed not only to help individual user autonomy, but to combat online extremism (Arthur, 2017). In this context, it is important to note that the European Union is–at the time of writing in summer 2022–developing a wide range of new laws that could help to improve the agency of individuals and the integrity of the information landscape:

- The Data Act (European Data Protection Board, 2022b)
- The Data Governance Act (European Commission, 2022a)

- The Digital Markets Act (European Commission, 2022b)
- The Digital Services Act (European Commission, 2022c)
- The Artificial Intelligence Act (Brakel, Uuk and Boine, 2022)

It is too early to evaluate exactly what the impact of these laws will be, but HDR reformers should observe them carefully. They will almost certainly play as pivotal a role in protecting the information landscape as the GDPR has in opening up data access to individuals.

In this approach, I have highlighted how important it is for HDR reformers to **seize and harness the powers we are given, and fight to hold onto them**. Groups of HDR reformers can combine development skills, innovation and disruptive design approaches to find and publicise new ways to circumvent providers' efforts to control and limit users' agency, and policymakers can prevent further erosion of individual agency through legislation and enforcement over the information landscape. As the popularised adage (based on the words of John Philpot Curran and Thomas Jefferson) goes, 'The price of freedom is eternal vigilance'.

9.6 Approach 4 to Improving HDR: Winning Hearts and Minds: Teaching, Championing and Selling the Vision

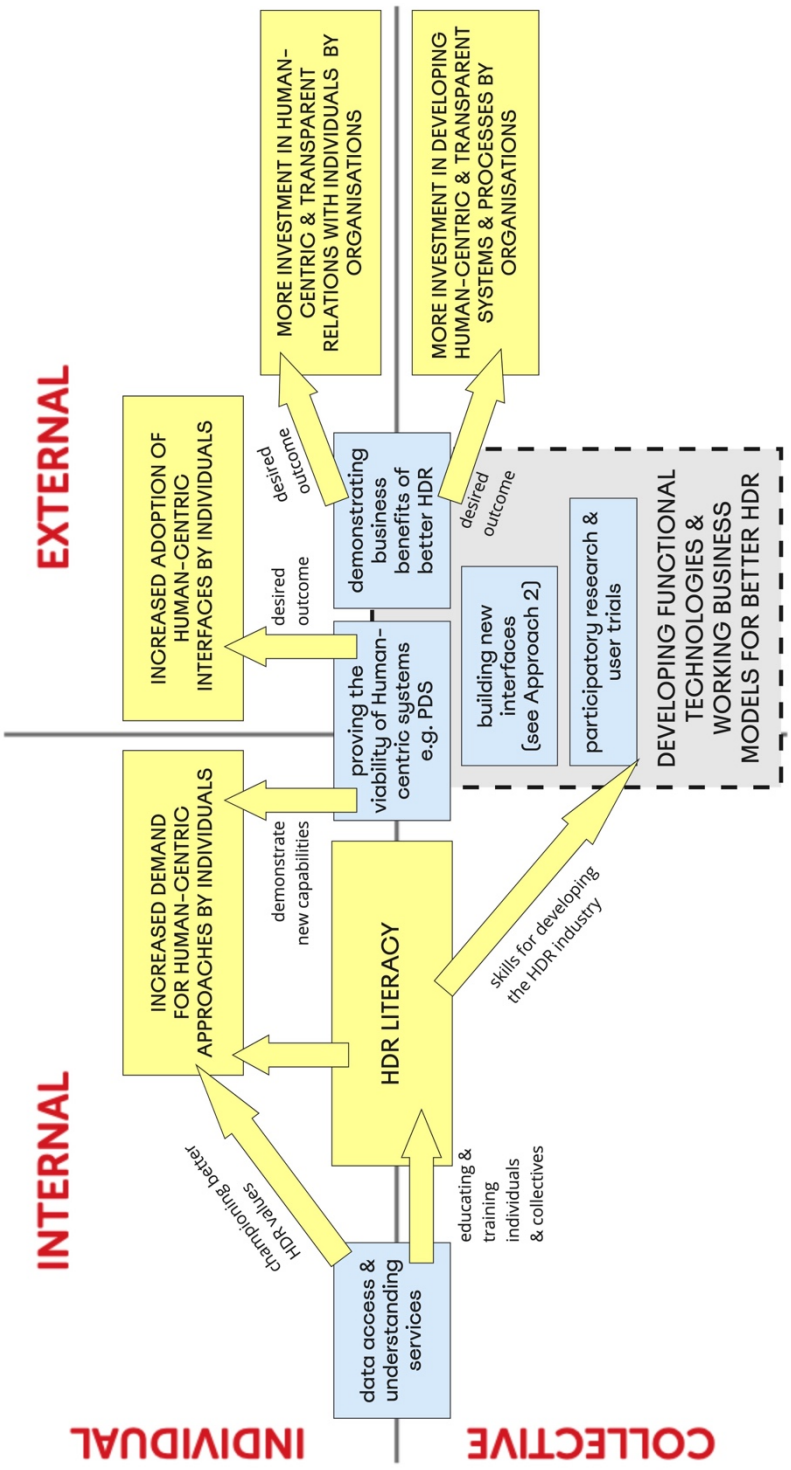


Figure 9.20: HDR Approach 4: Winning Hearts and Minds: Teaching, Championing and Selling the Vision

Approach 1, the investigative activist approach [9.3], and Approach 3, the digital freedom fighter approach [9.5], can help with users' ecosystem understanding [8.3] and ecosystem negotiability [8.4]. Meanwhile, Approach 2, the disruptive interaction designer approach [9.3], can help with user's direct understanding [8.1] and capabilities [8.2] over their data. But what of the fifth objective [8.5]? These approaches may not have sufficient impact until and unless the goals of HDR reform become widely accepted among business leaders, citizens, journalists and politicians. And for that to be truly accepted, **better HDR must be proven to work**.

The call for better Human Data Relations is a call for a radical reconfiguration of today's data world. As 8.5 outlines, where new systems are needed, system builders must invest in and *see the value* of HDR ideals (not just to individuals but to their organisations). Where new policies are needed, politicians must be persuaded that HDR's ideals are worthwhile and have public support. And most importantly for any change to occur, there must be a demand for change, and an engagement and appreciation of new HDR approaches once they become available. These are the goals of Approach 4, as illustrated in Figure 9.20 above.

Therefore, there a complementary parallel trajectory of HDR reformer effort is needed if the disruptive potential of HDR is to be realised; Across society, we must find ways to **demonstrate**, **persuade** and **prove the value** of better HDR. Collectively, we could call this **motivational work** in support of HDR reform. I have identified three aspects to this motivational work, which are explored below:

- Education & Data Literacy [9.6.1]
- Demonstrating Business Value [9.6.2]
- Proving the Viability [9.6.3]

9.6.1 HDR Education & HDR Literacy

Interactions with participants in the pilot study [1.3.1], Case Study One and Two have shown me that people do not feel comfortable and confident when it comes to matters of accessing and using their own data. This impression is empirically sound: A research study which surveyed over 1,500 members of the public about attitudes to personal data conducted at BBC R&D in early 2021 highlighted a lack of understanding and confidence

around personal data as one of four key findings. Other key takeaways from this research included feelings of helplessness and needing to know the basics (Sharp, 2021). A major part of any effort to overcome the lack of demand for HDR described in 8.5 therefore must begin with **educating people about data**–and more specifically, to educate people about data, life information [7.4.1], and personal data ecosystems [2.3.4; 7.4.2] from a human-centric, forward-looking HDR reform perspective.

Data literacy is already a strong education and skills focus area across public and private sector, for both children and adults. However, I identify some inadequacy in this concept when viewed through an HDR lens. Given the broad and varying perspectives of data [2.1.1], there are also varying ideas about what data literacy is. To some, data literacy is about the technical skills of number crunching, spreadsheets and data analysis (Precisely Editor, 2022). To others, it is a more high-level ability to read, understand and argue with data, and to exercise critical thinking or identify bias (Knight, 2016). A third perspective is that of technical prowess, as outlined by Gurstein as one of the needs for effective data access–the literal ability to interpret and visualise the information within your data (Gurstein, 2011). While all of these are clearly important aspects, something broader is needed to encompass ***HDR literacy***. This would additionally encompass such as aspects as:

- being able to **appreciate the intrinsic value** of your personal data as containing stored facts and records about your life [5.6.3];
- being able to understand the implications of **how your personal data is used and shared** by organisations you interact with [8.3];
- having an appreciation of the importance of establishing **portable data and separation** between data and platforms/services [9.2];
- knowing how to **exercise your rights** to obtain your own data, being able to recognise incomplete data returns and knowing how to demand corrections or better responses [5.1]; and
- understanding the need for a **free and fair information landscape** and being able to identify when individual agency is being diminished [8.4; 9.3].

These skills should become part of school curricula, but also need to be taught to adults both in the private and public sector–both in their roles as citizens as well as in their roles within the organisations they work for.

As examples of the sort of educational work that can be beneficial, we need to look at organisations whose remit includes the delivery of education and training. For example:

- The BBC's primary duties are to *entertain*, *inform* and *educate*. In line with the latter two objectives, the efforts begun in the Cornmarket project [see Section IV Introduction] continued from behind-the-scenes research into public engagement. The Cornmarket team produced explanatory videos and blogs to convey important parts of the HDR reform/PDS messaging to the public (Sharp, 2021; BBC Cornmarket Team, 2021). A new BBC R&D prototype called 'BBC Together+ DataPod' was produced and released to the public in 2021 which combined Cornmarket PDS technology with existing 'BBC Together' watch party capability. This allowed groups of friends to have fun watching BBC content together, while also gaining an understanding of what data is generated while they do and how that information might be utilised to better understand themselves. This conveyed to users some of the basic elements of a personal data ecosystem under their own control.
- In Switzerland, Hestia.ai is pioneering a new approach in offering interactive data-exploration workshops to businesses, researchers, educational establishments and activist collectives around Europe. This takes the approach used with EU politicians in the digipower investigation (Härkönen and Vänskä, 2021) and broadens it into general training provision known as *digipower academy* ('About digipower.academy', 2022). Workshops will enable participants to upload their data, explore data flows and aggregate information insights, and learn about the important issues of the personal data economy.

To support this approach and reach wider audiences, efforts such as those of BBC R&D and Hestia.ai need to be invested in, replicated, scaled up and offered to all age groups and all levels of society. This shows that there is a second, perhaps more significant, reason to support the growth of *data access & ecosystem understanding services* in addition to the investigative angle in Approach 1, namely to **raise HDR literacy across society**.

9.6.2 Demonstrating the Business Value of Better HDR

There are two aspects to the motivational problem of generating demand. One is motivating the end users - for which I have shown an approach running through Principle 7 [10.1.7] and Approach 2 [9.4]. The other, perhaps more challenging, is to demonstrate that the radical new approaches of HDR reform, while they will entail significant changes and new work, will be worthwhile for businesses and organisations. Some of the challenges of shifting to more inclusive and more human-centric ways of operating are explored in Case Study One [4.5.3]. Principle 13 [10.1.13] considers the the ways in which HDR reform such as shared data stewardship [4.3.4], inclusive data flows and individually-sourced data [9.4.5] might be beneficial to businesses, concluding that it would be possible and useful to demonstrate such business benefits. Principle 13 [10.1.13] identifies areas which need to be evaluated and explored through research or entrepreneurial investment, in order to produce data that could **persuade businesses to adopt reformed HDR approaches** to data handling and service user interaction.

9.6.3 Proving the Viability of Human-centric Approaches

The third aspect of the motivational work in Approach 4 is that work must be done not just to create new systems and technologies that meet the HDR objectives [7.5], but to **prove that human-centric HDR approaches work**. This involves both developing functional technical **proofs of concept** to test HDR design concepts such as those in Approach 2 [9.4] in practice, as well as starting businesses which can explore new business models to discover which forms of value in the PDE / HDR space can be sufficient to drive the space and its players–which after all will be the engines of change in pursuing HDR–forward.

As an example, development work I undertook with Stuart Wheater as part of my role in the SILVER project (Connected Health Cities, 2017) was successful in building a working system to extract citizens' health data from EMIS, the medical system used by NHS GPs, and make it available as understandable and explorable timeline-based information that could be understood by support workers, drawing on the ideas of temporal PIM systems [2.2.2]. The system was extremely difficult to construct due the closed and insular nature of the different parts of the health and social care sector [8.5.2], resistance to change, and the fragmented nature of public sector systems (Pollock, 2011). Nonetheless, it was

successfully trialled with support workers (using test data) and received positive feedback. It proved that silos can be broken down and life information presented in new ways. A screenshot of the interface I developed is shown in Figure 9.21, and example videos have been published online (Bowyer and Wheater, 2017).

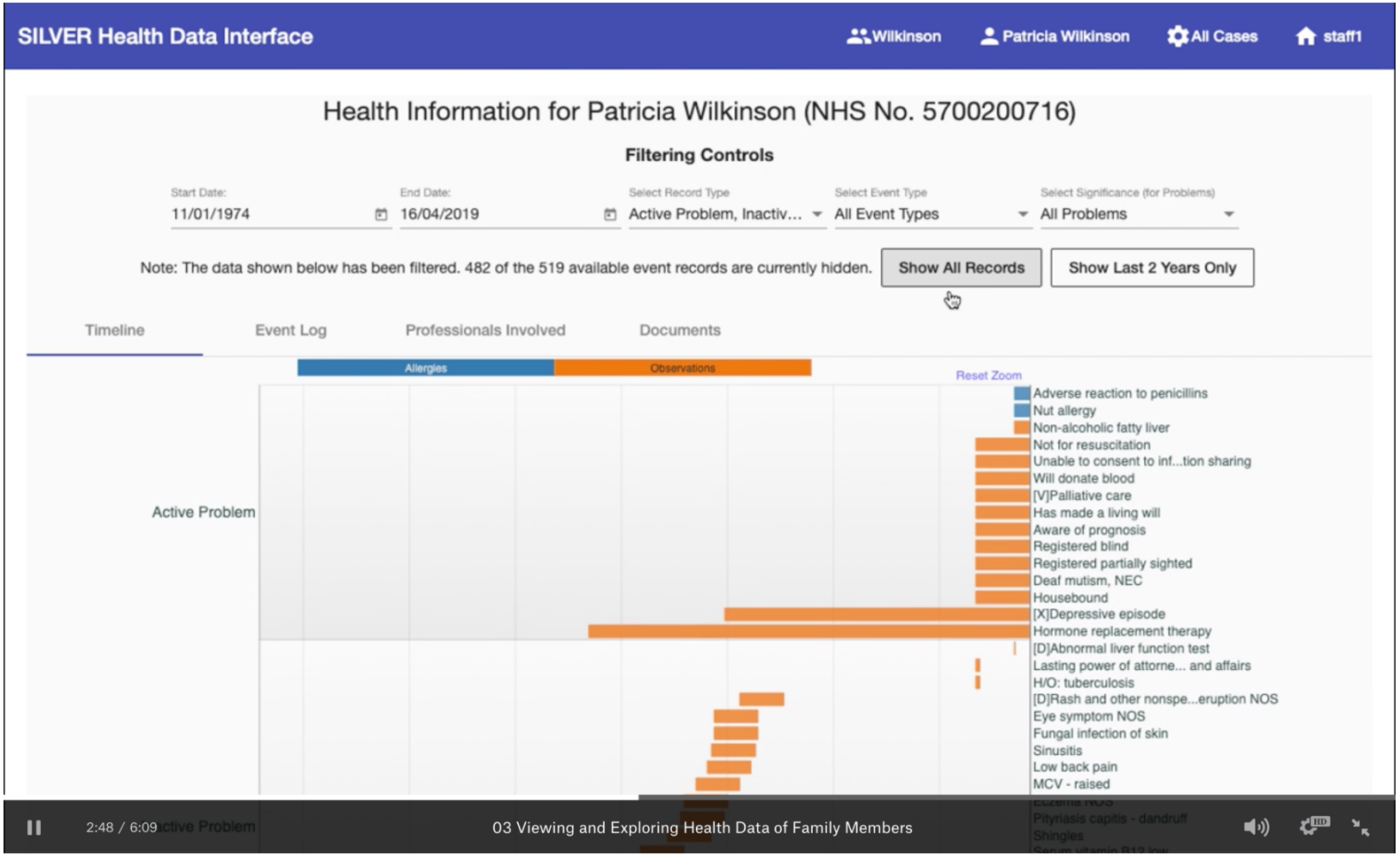


Figure 9.21: SILVER Health Data Viewing Interface

Across the *MyData* and PDE / HDR space, many small businesses have been established which have demonstrated successes in different elements of pursuing the visions of MyData and HDR. For example:

- Digi.me, along with their partners UBDI ('Whose data is it anyway?', 2019; 'Digi.me', no date), have demonstrated the technical and business viability of **human-centric solutions which allow users to securely share data**, through a personal data store-like library, with paying third parties, while maintaining individual oversight and information-viewing functionality. CitizenMe (CitizenMe, 2021) has also demonstrated success in this field, focusing on individual data generation, data donation and personal rewards.
- Tim Berners-Lee's company, Inrupt, has successfully produced an effective open source **PDS platform technology called *Solid***, which is in use in several organisations across the world [Olivo, 2020; 9.2]. The Solid community has also resulted in the first practical standard for semantically unifying human information from the personal data diaspora [8.1.2], as called for in 8.5, *data shapes* ('ShapeRepo: Make your apps interoperable', 2022), which is gaining adoption.
- BBC R&D's Cornmarket project has come up with a viable set of designs and mock-ups and a working R&D prototype. Built upon Solid technology, it connects to real data sources including Netflix, BBC and Spotify. If supplemented by some of the features and strategies described in 9.3, it could deliver a **real-world PDS-based life interface**. While not yet tested in the marketplace, the designs have received positive feedback in a 1,500-participant study, including specific appreciation for many ideas emerging from my research and design work, including life insights, goal setting, personal data unification, individual oversight, data exploration, the twin motivators of LIU and PDEC [7.4], and self-contributed data (Sharp, 2021).
- With several customers, including non-profit research organisations, universities, researchers, consumer organisations, local authorities and transport providers, Hestia.ai has proven that there is a **paying market for *data access and ecosystem understanding services*** (Dehaye, 2022), which are needed for the reasons outlined in Approach 1 [9.3] and Approach 3 [9.5].

Looking at Approach 4 as a whole, this section has shown that pursuing HDR is not purely a data interaction design problem, nor solely a political problem, nor solely a technical problem. It is all of these, but also, ultimately, even though research can be done in non-profit settings, **HDR remains a business problem**. In our capitalist society it is essential to find a path to better HDR that companies can get behind. And this approach too has shown that motivation and education go hand-in-hand with all of the above. It is not enough to build new systems. It is necessary to catalyse a cycle of constant feedback, of data-enabled design and action research or iterative software and business model development–**finding what works, championing it, communicating it and selling it**.

9.7 Summation of Chapter 9, and of Section IV

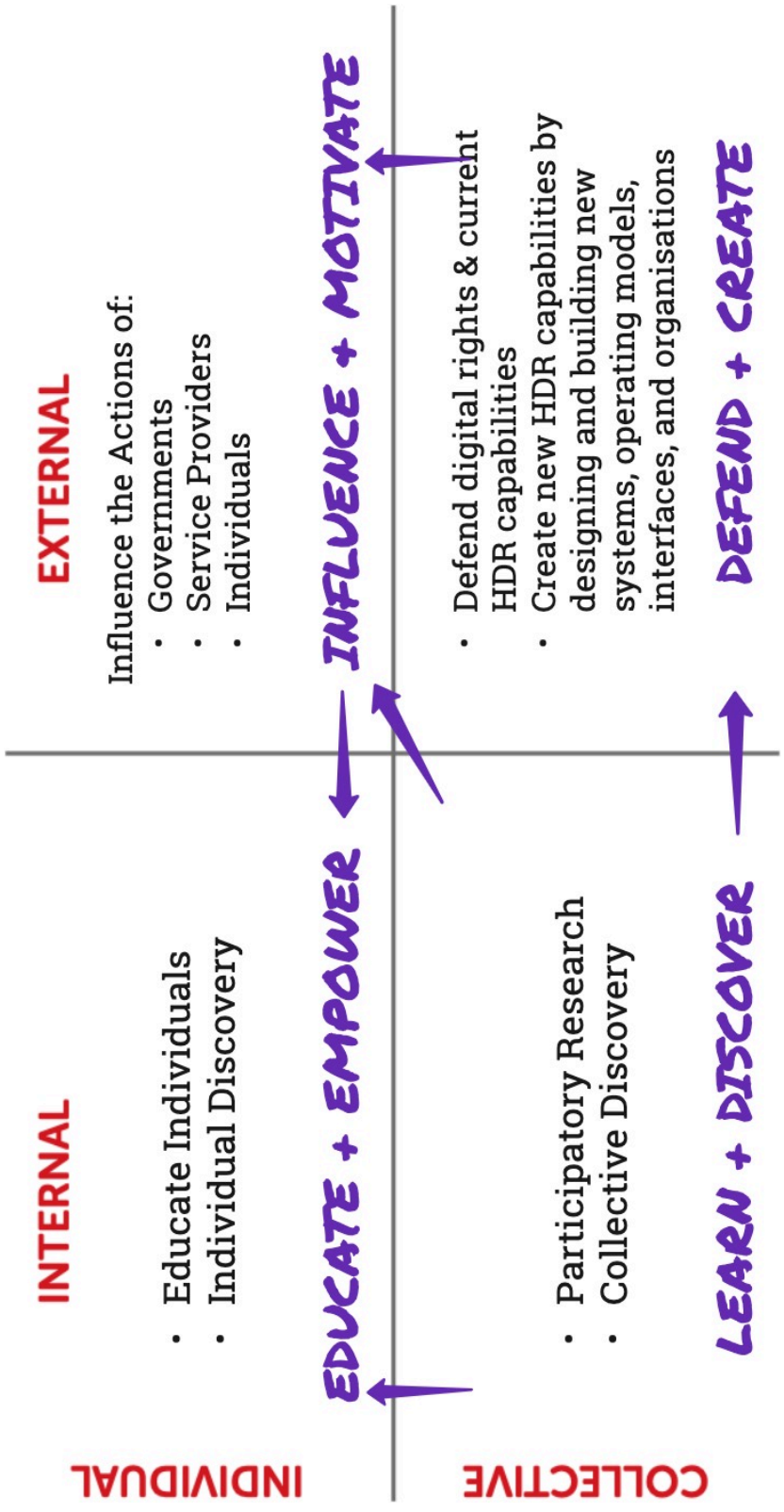


Figure 9.22: Summary of Generalised Change Strategies for Pursuing Better HDR, Using the ToC Model

This chapter examined the expanded research question [7.1] of how better Human Data Relations might be achieved in practice. Through detailed practical examples drawn from the embedded research settings [See Section IV Introduction] and elsewhere, and building upon the 13 Principles [10.1] inferred through the insights of Chapters 8 and 9, I have illustrated and described four distinct trajectories for improving HDR. These provide practical strategies for HDR reform that could bring about better HDR.

- **Discovery-Driven Activism** [9.2; Figure 9.2]
- **Building the Human-Centric Future** [9.3; Figure 9.3]
- **Defending User Autonomy and Hacking the Information Landscape** [9.4; Figure 9.18]
- **Teaching, Championing and Selling the HDR Vision** [9.5; Figure 9.20]

The common elements of these four approaches are summarised in abstract in Figure 9.22 above, which is best understood with reference to Figure 9.1. These common elements are positioned in the four ToC quadrants:

- ***Learn & Discover*** (CI): In this quadrant, individuals, researchers, activists and other stakeholders work in groups to understand data attitudes and user needs, and to gain collective knowledge of data collection and usage practices which are sometimes hidden.
- ***Defend & Create*** (CE): In this quadrant, activists work to ensure current HDR capabilities are not eroded, while researchers, designers, technologists and social innovators design and create new technologies, operating models, organisations and interface designs, the structures enabling a world with better HDR.
- ***Influence & Motivate*** (IE): In this quadrant, individuals' relationships with data and with data holders, as well as data holders and policymakers' relationships, can be improved. For us, as external actors seeking change, the task is to influence the many parties by showcasing and facilitating newly created structures and capabilities, and to harness our new collective knowledge to advocate the benefits of changing data-related behaviours.
- ***Educate & Empower*** (II): In this quadrant, individuals' ways of thinking about data and data holders grow and evolve towards a state we could call *feeling empowered*

[6.3]. This would constitute definitive progress in delivering the better HDR that I and the Case Studies' participants called for. This change is driven through education, improving HDR literacy, and through the experience of new capabilities and changed relationships with data and with data holders. Such changed relations would empower individuals to hold a more aware and equitable position in every aspect of their digital life.

Taken together, these multi-pronged efforts show **how HDR could be improved in practice**, even despite the identified obstacles. The HDR Index included at the back of this thesis provides a convenient way for future researchers, innovators, activities and other HDR reformers to quickly locate insights and designs of value within this the thesis, so that others can build on this research to pursue improved Human Data Relations. Those who wish to pursue these approaches need not be technical or experts. The HDR reform agenda, and in particular these four approaches, can be supported or pursued in many different ways, which may appeal to different readers:

- prototyping and creating proof-of-concepts;
- fundraising or investment;
- design activities;
- documentation and knowledge sharing;
- market research, participatory research or usability research;
- 'early adopter' testing and quality assurance of PDE / PDS offerings;
- promotion, advocacy and journalism of HDR issues;
- critical audits of provider practices;
- policy design;
- political pressure on governments and regulators;
- participation in open data, PDE, civic hacking or *MyData* communities; or even
- individual self-experimentation with HDR tools, rights and capabilities

This concludes Section IV of the thesis. In the final section, Section V, the core principles of Human Data Relations are explained, followed by detailed reflections upon the viability, legacy and contributions of the HDR research agenda and of this thesis.

SECTION V:

CONCLUSIONS & OUTLOOK

Introduction to Section V

The purpose of Section V is to reflect upon the learnings of this research and make them actionable and well-contextualised to future designers, researchers and innovators. Chapter 10 looks back upon the thesis's insights and findings to produce a set of clear and actionable principles for better Human Data Relations. It also includes pragmatic and personal reflections on this body of work and the agenda it advocates, and positions the PhD in terms of its potential legacy and limitations.

Chapter 10. Thesis Conclusion

"Our research should transform, not just inform, society."

—Kingsley ofosu-Ampong
(researcher & lecturer in digital transformation)

This Digital Civics PhD has explored, from an constructivist, individualist, and pragmatist point of view, the nature of the power imbalance between individuals and those who hold data about them. It has built upon literary bodies of work in multiple disciplines including information theory, data rights legislation, surveillance capitalism, Personal Information Management, Human-Computer Interaction, Human-Data Interaction, Personal Data Ecosystems and Human-centered Design. In doing so, it identified a research gap around the lived experience of today's data-centric world, recognising that there was a need to explore the role that personal data should play in people's lives, and how the current power imbalance over personal data [2.1.2] affects people's attitudes and capabilities.

Through the Case Studies in Section II it explored this research gap, reaching in Chapter 6 within Section III a clear answer to its research question [2.4] in terms of six wants that people have in their direct and indirect data relations; people want:

- visible, understandable and useable data, as well as
- process transparency, individual oversight and involvement in decision making.

These six wants provided a basis for envisioning a world where people become empowered digital citizens with mastery of their own personal data.

Through its novel two-track approach of real-world project placements taking place alongside the participatory research, as described in Section IV, this evolving vision for better relations with and through data was tested, refined and scrutinised through a practical, sociotechnical lens that began consideration of a further question, of how the world might be changed towards this new vision. The result of this hybrid learning was presented as a new research agenda, *Human Data Relations (HDR)*. HDR was defined in Chapter 7 within Section III, culminating in the expression of four clear objectives:

- data awareness and understanding;
- data useability[11];
- data ecosystem awareness and understanding; and
- data ecosystem negotiability.

Through Section IV, this new research agenda was given flesh, by mapping out the understood obstacles to progress when pursuing these objectives, and providing four strategic trajectories for pursuing societal change towards better HDR:

- discovery-driven activism,
- building human-centric life interfaces,
- defending and exploiting the seams of the information landscape, and
- championing and teaching the HDR vision.

What remains for this chapter to address, is to distil the learnings of what people want from data in their lives (and how their wants might be met) into a set of principles from which future researchers, designers and activists might be able to generate new research projects, social interventions, policies, or other initiatives [10.1], before reflecting critically on the research agenda [10.2] and upon this body of research in terms of its limitations [10.3], personal impact upon the researcher [10.4] and potential legacy [10.5].

10.1 Principles for Generative Action

In line with this researcher's objectives [1.1.1], this PhD set out to have an impact upon the world with regard to tackling the power imbalance over personal data. Indeed, through publications, methodological contributions and industry adoption [1.2; 1.3] it already has. In order to maximise potential impact, this thesis will conclude not by providing a prescriptive roadmap for HDR, but by sharing a versatile set of principles that can be applied by all who wish to pursue this research agenda.

The principles are framed as principles for *generative action*, which refers to the idea of producing solutions that themselves stimulate further solutions, in line with the concept of *generativity*, that in psychology refers to the creation of something novel, valuable and meaningful (McAdams, 1996), and in sociology refers to a stage of adult development focused on productivity, creativity and contributing to society (Slater, 2003).

In recognition of the fact that the power imbalance over personal data is a sociotechnical problem, these principles are broader than just design insights for influencing UI design or software system specifications; they could also be applied at a societal level to influence funding and policy decisions, social interventions, business strategy, grassroots activism and more. As such, they are a tangible and actionable output of this research, routed in human-centric philosophy, participatory findings and sociotechnical reality.

10.1.1 Principle 1: Life Information makes Data Relatable

In the pilot study and Case Study One, data cards were used to represent civic data [Figure 3.7]. In Case Study Two and in Hestia.ai's *digipower* investigation [See Section IV Introduction], a categorisation of provider-held data was displayed [Figure 3.8]. In my BBC research report (Bowyer, 2020a), the use of **relatable examples** was identified as an important way to help people understand what a piece of data represents. Recalling that to make data meaningful, we must be able to interpret it as information [2.1.1], this can be refined further:

To make data meaningful,it needs to be expressed as information about your life.

Spreadsheets and 'big data' sound dry and (to many) dauntingly technical, but once those same datapoints are expressed as 'facts about your life', the hurdle of relatability is overcome [4.3.1]. The effectiveness of applying this principle is evident in successful online services like Netflix, Spotify and Strava, and in social media platforms like Facebook: these interfaces show understandable everyday concepts like Friends, Events, Movies and Playlists, not files, records, folders or database rows. They have successfully *'pushed the technology into the background'*, in line with Weiser's vision (Weiser, 1991) and Rogers' *calm computing*. While exploring this idea of representing **life concepts** further at BBC R&D, I produced Figure 10.1, which shows a near-exhaustive overview of the many different informational concepts in an individual's life that providers might hold as data:

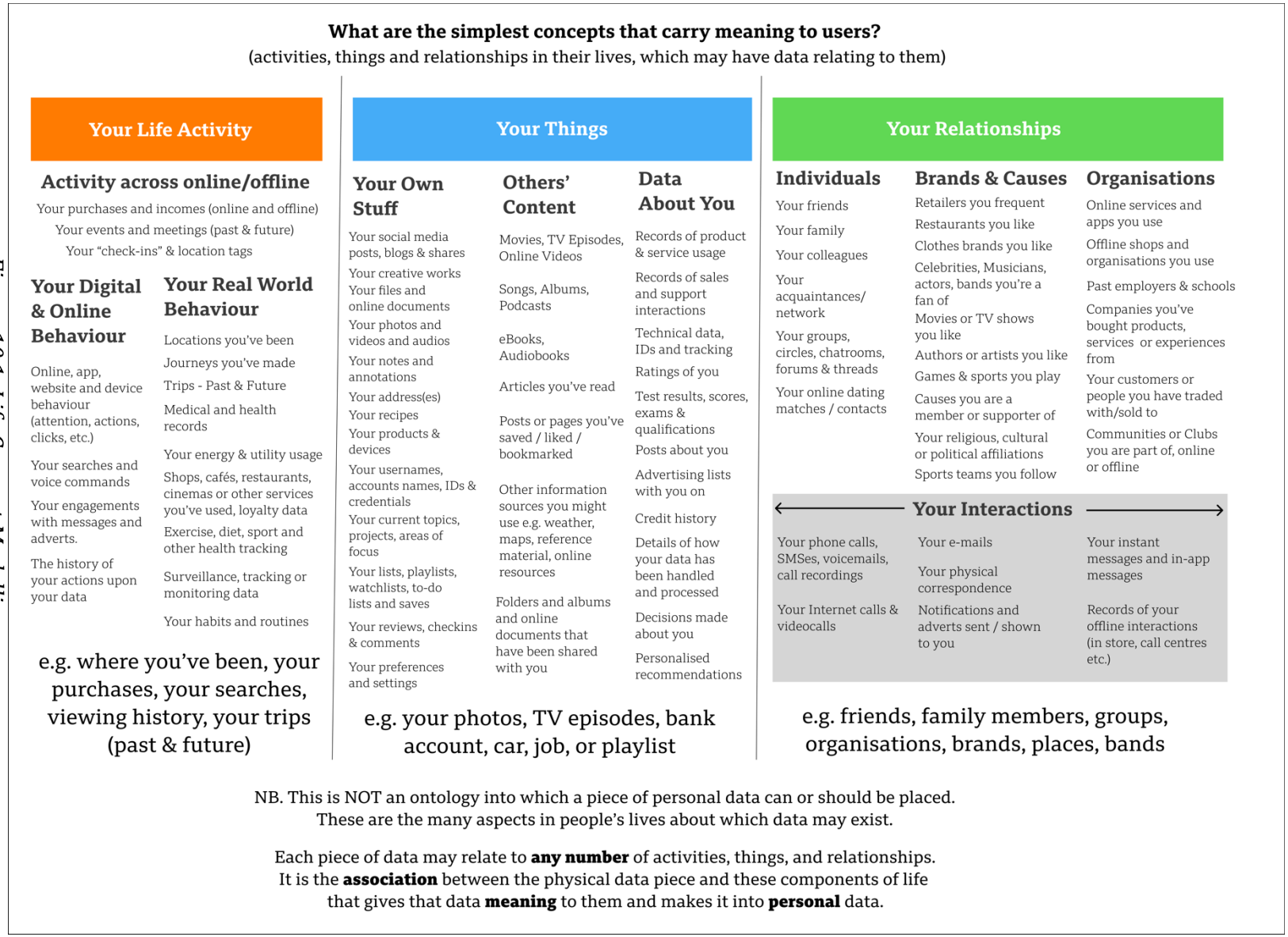


Figure 10.1: Life Concept Modelling

This diagram shows how most common personal data types handled today can be mapped to more relatable *life information* [7.1] concepts. These life concepts (exemplified where possible) can make data meaningful to individuals, and can help them find value in their data [5.5.3].

10.1.2 Principle 2: Personal Data Needs to be United and Unified

It is clear that better HDR involves recognising the splintered and scattered reality of where our data is [Lemley (2021); [8.2.1]] and moving beyond it. To make data useable[11] for individuals, the diaspora must be united. This means that data from different sources must first be **united**—brought together—and then **unified**, which means making it into a collection of data about the individual and their life, rather than scattered slices of company data that may have secondary value to the individual. This is a multi-faceted sociotechnical problem of access, interpretation and *integration* [Li, Forlizzi and Dey (2010); 2.2.3]]. Negotiability remains important; we can only unite data that we can access, and only data holders can fully explain it [see 8.3 and 8.4]. Setting that aspect aside, the pragmatic way forward begins with creating a space where data can be held, combined, controlled and **owned** by the individual - *'a place for your personal data'* [W. Jones (2011a); [2.2.4]]. This can form the seed of their new human-centric personal data ecosystem. This follows Bergman's *subjective classification principle*, as mentioned in 2.2.2:

> *'All related items should be classified together regardless of technological format'*—
> Bergman, Beyth-Marom and Nachmias (2003)

We could add: *'regardless of where they are held'*. This vision is embodied in the **Personal Data Stores** (PDS) concept [2.3.4]. A PDS can brings together personal data from multiple sources that has never co-existed before. This enables the provision of new capabilities over one's digital life. The BBC R&D Cornmarket project, as well as the projects described in 9.2, examine how to build PDSs. In 9.4, I explored possible design approaches. At this stage, only the *concept* is important. Once data is united and unified, PDSs enable the creation of new views of data that were not previously possible, because code can execute across data that was previously dispersed. For example, today each separate TV app, device or streaming service maintains separate records of what you have watched. Once unified in a PDS, it would be possible to present you with a unified view of all the past content you had viewed, across all channels, as this mock-up I made during my BBC internship shows:

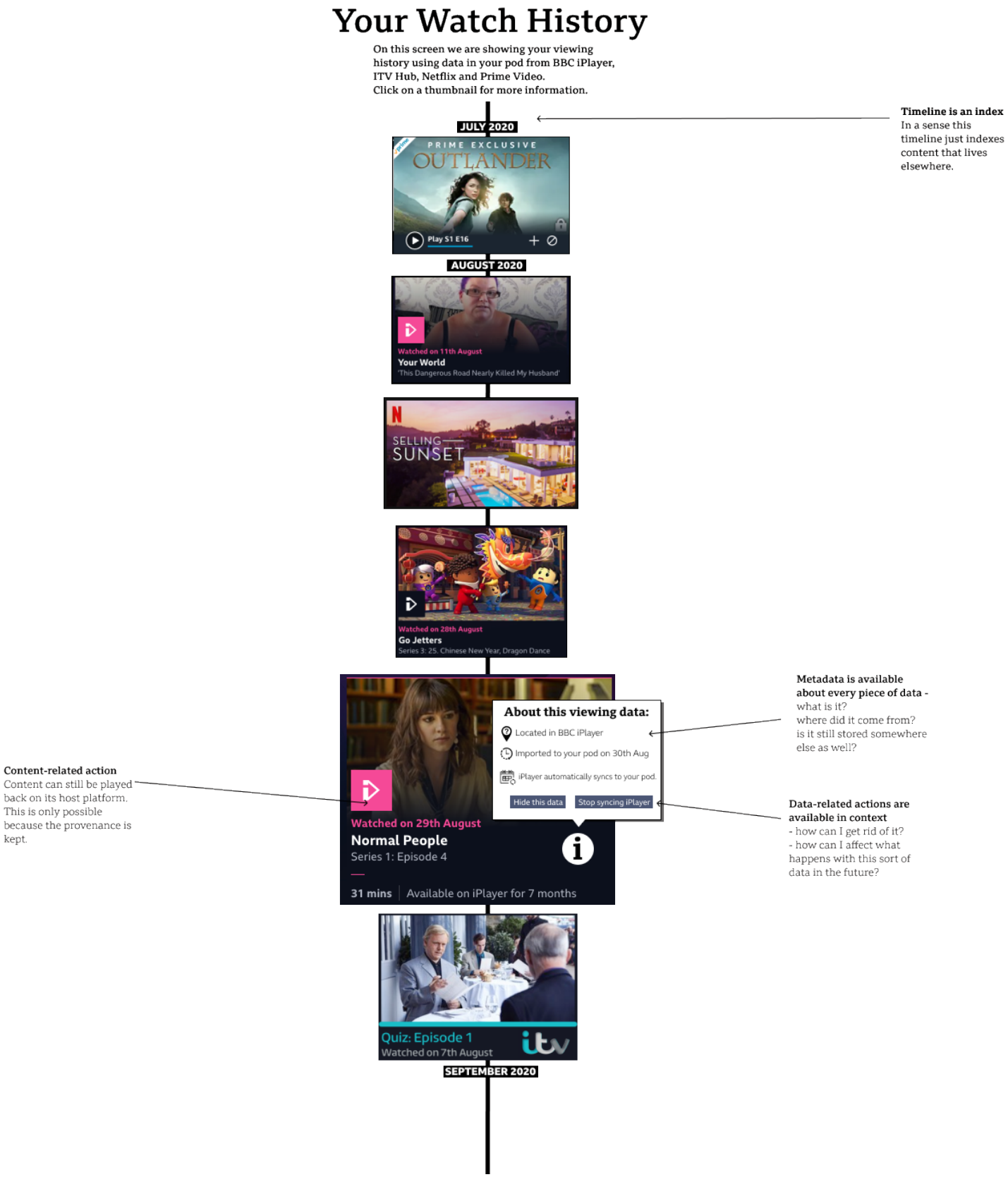


Figure 10.2: Mock-up of a Unified TV Viewing History Interface

10.1.3 Principle 3: Data must be Transformed into a Versatile Material

In Case Study Two [Table 5.4; Bowyer, Holt, *et al.* (2022), supplemental materials], participants expressed diverse goals for personal data, including reflection, pattern-finding, goal-tracking, and creative use. In the PIM space [2.2.2] relevant innovations include associative exploration, spatial arrangement, and embodied interaction for different contexts. Drawing on all of these allows me to infer that unified data must be transformed into a **versatile material**. Individuals need to be able to use data—represented as facts or assertions about one's life–by performing manipulations such as:

- creating,
- deleting,
- moving,
- grouping,
- annotating,
- copying,
- sharing,
- modifying,
- labelling,
- organising, and
- separating.

Data as material will be new to most except data scientists. This is novel not just for end users but for designers too. Eva Deckers, in her work on *data-enabled design*, an approach to design which also calls for data to become a material, notes (and we could expand this to laypeople too):

> *"Designers are in general not trained and prepared to work with data. They're not equipped with the right tools. Data manipulation is not part of the schools' curriculum and designers are rarely interested in understanding data."*—(Deckers, 2018).

Her work with colleagues on the 'connected baby bottle' illustrates how treating data as a design material creates a novel iterative user-centred development of new capabilities (Bogers *et al.*, 2016). The HDR principle of data as versatile material is distinct from prior

conceptualisations of *data materialisation* as the making of physical objects that manifest data (Beghelli, Huerta-Cánepa and Segal, 2019), but rather refers to the realisation of human information of a pliable digital substance which users can manipulate for themselves in order to gain new insights and exercise new capabilities. In HDR terms, I theorise that what the material should *be* is *human information* - life information and ecosystem information [7.4]. To achieve data useability [6.1.3] therefore implies a call for the creation of systems that enable **human information to be treated as a material**.

10.1.4 Principle 4: Ecosystem Information is an Antidote to Digital Life Complexity

In 2014, Microsoft official Craig Mundie said:

> "*Today, there is simply so much data being collected, in so many ways, that it is practically impossible to give people a meaningful way to keep track of all the information about them that exists out there, much less to consent to its collection in the first place."* – Craig Mundie, writing in Foreign Affairs magazine (Mundie, 2014)

Based on what I have learned in the formation of the Human Data Relations agenda, such an attitude is defeatist. It is certainly possible to give people an understanding of a large part of what is happening in their digital lives. And I believe that the concept of *ecosystem information* [7.1] is central to this. My research showed that acquiring ecosystem information and understanding is a key motivator for many people—encompassing 74% of participant goals in Case Study Two [Table 5.4]—and is essential for better HDR. This suggests two distinct goals for system builders: **ecosystem detection** and **ecosystem information display** as ingredients to help overcome the obstacle. As a representative example let us examine a recent app called SubsCrab [Figure 10.3]:

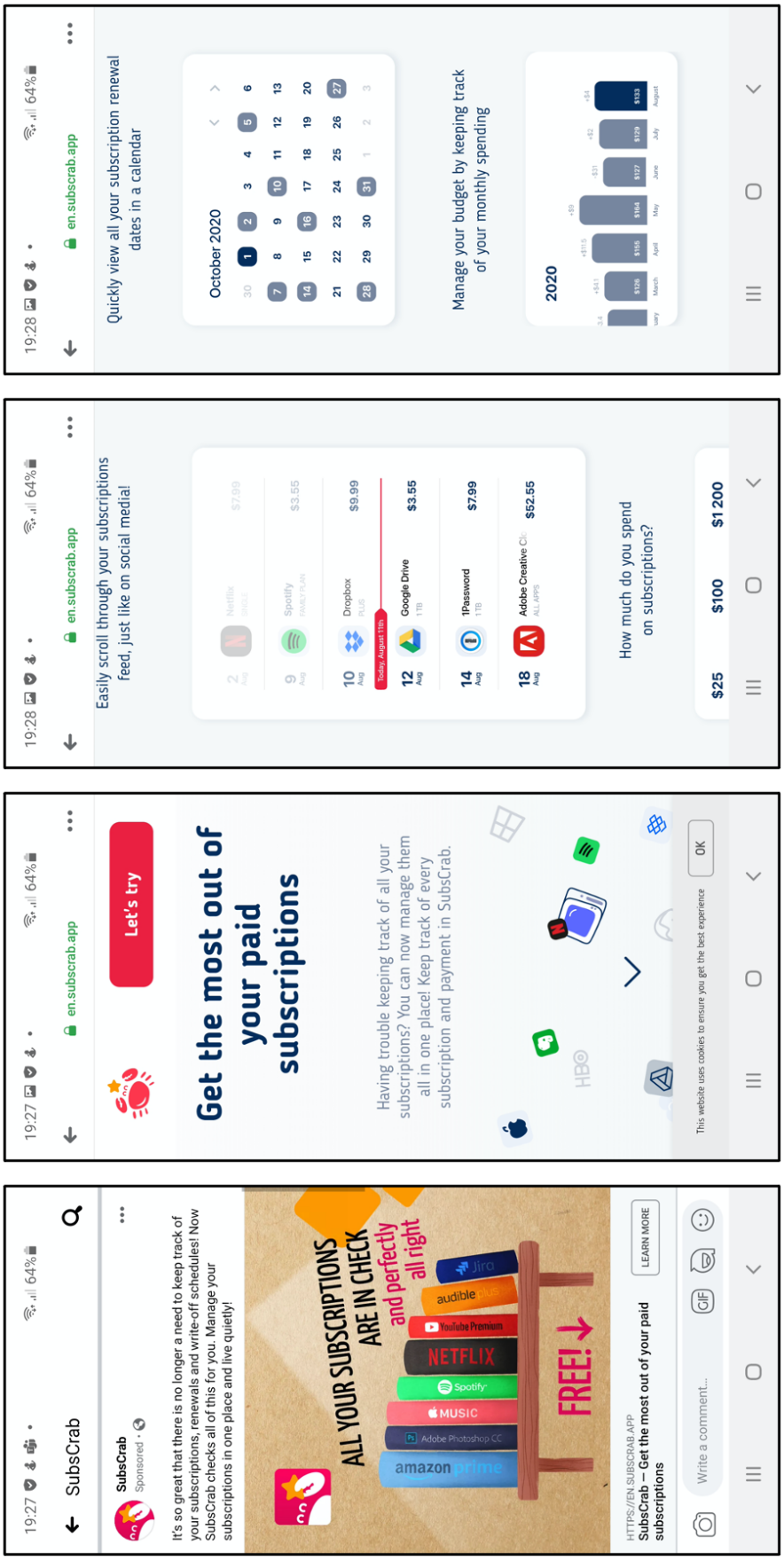


Figure 10.3: SubsCrab: An Example Application for Ecosystem Detection and Visualisation

This app connects to the user's e-mail account, and searches it and monitors it for e-mails from service providers such as Netflix, Spotify, Dropbox, or Google with which the user has monthly or annual subscriptions. In doing so, it is detecting part of the user's ecosystem. It is identifying which companies they have a payment relationship with. It parses found e-mails to identify billing dates and payment amounts. It then provides additional representations of that ecosystem information to the user, so that they might get on top of their subscriptions, see what they need to pay (or cancel), and feel more 'in control' [Teevan (2001); 2.2.2] of this aspect of their digital life. From this example, it is easy to imagine other types of ecosystem detectors, which could detecting relationships with free services and websites, identify account numbers and e-mail addresses, password resets, address book syncs, OAuth logins, family identities and more. Alistair Croll and I explored possibilities for *inbox scanning* in 2009 (Croll and Bowyer, 2009), and while there has been some innovation in this space, it has largely been done with a commercial focus (Braun, 2018). New ecosystem detectors could power new interfaces, contributing to the simplification of the user's digital life. This would give people more visibility and control over their previously unmanageable data ecosystem.

A secondary consideration in achieving the required sea change in approaches to HDR, is that current PDS and SI approaches are very life-information-centric. It is implicitly assumed that the only way to unite data is to collect it. The difficulty in such an approach is that you can only collect that which you can extract. To address this, I draw inspiration from a computer programming concept known as *pass by reference* (as opposed to *pass by value*) (Ananya, 2020) where data is 'pointed to' rather than moved. Productivity guru David Allen recommends the use of *'placeholders'* (Allen, 2015) to keep track of tasks you cannot otherwise bring into your planning. To build a complete map of a user's ecosystem we must be able to keep track of accounts and data that are remote, much like a search engine points to information on different pages around the web. We can create **proxy representations** of service-provider-held or otherwise immobile data (e.g. data which is offline or restricted). These representations become part of the manipulable material in the user interface, and could be augmented with links to visit the remote source.

10.1.5 Principle 5: We Must Know Data's Provenance

Metadata is what gives information *context*. Context is critical to *sensemaking* [2.2.3] and enables good experience-centred design [2.3.2; 2.3.3]. Without context, data loses meaning [5.5.3]. Collecting historical data about the individual is important for reflection [2.2.3] and considered valuable [4.4.3], but knowing the **history of a piece of data** allows its context to be understood. Data is not neutral, and is inherently biased, since it was created for a specific purpose with a specific agenda in mind (Gitelman, 2013; Neff, 2013). To combat this bias, more context is needed. Significant research in this space has been undertaken by Professors Mike Martin and Rob Wilson at Northumbria University, formerly Newcastle University, who promote the idea of **data with provenance**; in other words:

Data must carry with it the details of why it exists, how it came to be, and what has happened to it since its inception.

Provenance should be communicated alongside any visualisation of the data, in order for it to be fairly assessed in context. Provenance is essential for data to be trusted, argues Martin, and should be quite granular: a piece of data should be attributed not just to an individual or organisation, but to the relationship between role-holding individuals in a specific context. Greater insights can be gained when considering all actions upon data as motivated communications from one party to another; only by capturing this information in-situ can the data's meaning be fully appreciated (Martin, 2022). This framing essentially advances the concept of history tracking [2.2.3] into the sociotechnical, ecosystem-aware problem space. While everyday system designs have not approached this level of granularity, the importance of data provenance has been recognised in the PIM space. Temporal PIM systems [2.2.2] from Lifestreams (Freeman and Gelernter, 1996) to *activity streams* (Hart-Davidson, Zachry and Spinuzzi, 2012) rely upon data provenance in some form. A study by Jensen *et al.* concluded that provenance tracking can be valuable for identifying related documents, a critical part of knowledge work today (Jensen *et al.*, 2010). Lindley *et al.* proposed the idea of *file biographies*, which view the lifetime of a file as something that should remain connected, so it could be traversed in order to understand the context of the file different moments of interaction (Lindley *et al.*, 2018). This comes close to Martin's vision but does not capture the motivation for each interaction. While provenance capture is not a solution in its own right to the understanding of data and of

ecosystems, it is clear that data with provenance is very likely to be a valuable part of any design that aims to provide understanding of complex and invisible personal data ecosystems.

10.1.6 Principle 6: Data Holders use Four Levers of Infrastructural Power

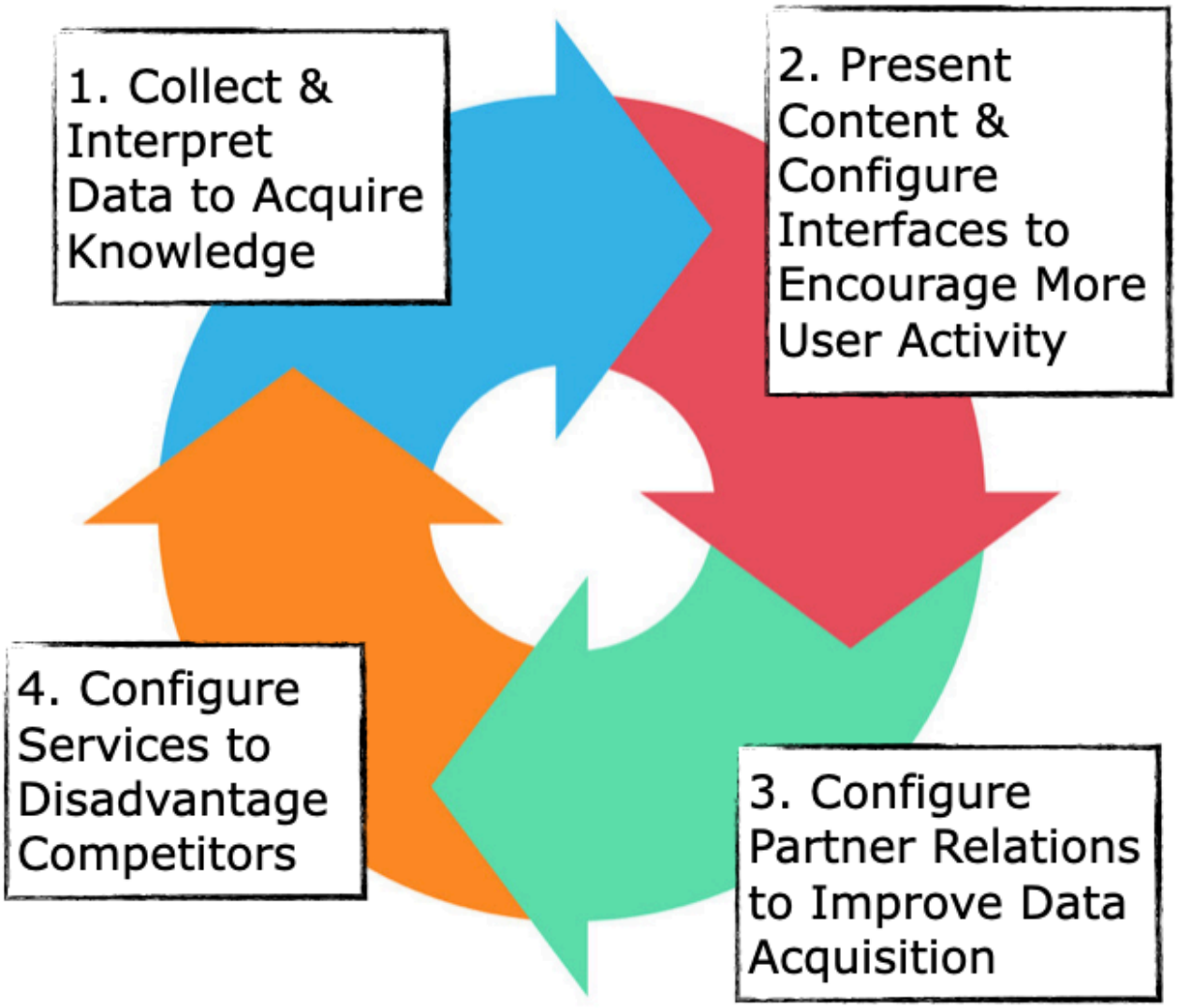


Figure 10.4: The Four Levers of Infrastructural Power

At Hestia.ai [See Section IV Introduction] we produced produced a model to explain the mechanisms by which technology companies gain power and use it to shape today's digital landscape. In this model, *infrastructural power* comes from three things:

- *technical ability*,
- *organisational ability*, and
- *the acquisition of data about individuals and populations*.

As organisations (especially platforms such as Google or Facebook) collect more data, and grow in market influence or technical capability, they gain power over individuals and over other organisations. They exert this power using four 'levers'. Simplified and expressed in the terms of this thesis, and shown in Figure 10.4, these are:

- **Collect & Interpret Data to Acquire Knowledge**: Data and *signals* are collected from individuals and interpreted in order to infer their intents and interests. For example, Google collects raw GPS and wi-fi hotspot data from mobile phones, which it then statistically analyses to infer which shops or venues you visited and what forms of transport you used, increasing Google's knowledge about individuals and populations.
- **Present Content and Configure Interfaces to Encourage More User Activity**: Knowledge of individual intents and interests is exploited within user interfaces to influence desired individual actions. For example, Facebook or presents a user with a product relevant to their interests, which they are motivated to click upon, generating ad revenue. Another example would be Twitter manipulating the content of the user's feed to show more tweets from conversation topics where they can show promoted tweets, increasing ad revenue.
- **Configure Partner Relations to Improve Knowledge Acquisition**: A provider uses its dominant position to force other organisations to improve the provider's ability to acquire knowledge. For example, Google provides free analytics tools to web developers, but requires the end users of those client websites to supply visitors' data back to Google, increasing their ability to acquire knowledge about individuals and populations.
- **Configure Services to Disadvantage Competitors**: Certain providers (typically of operating systems or popular devices) can configure the structural relationships

between other parties. For example, a smartphone manufacturer could limit data exchange between other apps, while still extensively collecting data signals themselves, such as when Google was found to be collecting call history from Android's dialer app.

The precise mechanisms and techniques employed when exerting infrastructural power, as well as the social and market consequences of these practices, are explored in detail in Hestia.ai's digipower technical reports, of which I was a co-author (Bowyer, Pidoux, *et al.*, 2022; Pidoux *et al.*, 2022).

The research highlights that platform providers' power is far greater than many realise. Unlike in the physical realm, providers of popular online platforms can **reconfigure the landscape to change the way that individuals perceive reality**, in line with the powers of interpretative influence, behavioural influence and socially shaped power described above (Bowyer, Pidoux, *et al.*, 2022). Providers control the extent to which (if at all) data stored behind the scenes, and internal processes that use that data, are visible, and how data and processes are represented. Common to the last three levers is an implicit, if not malicious, dishonesty; all three rely upon changing other actors' perceptions and understanding of the landscape, such that they might be more likely to act in the platform provider's interest.

The model shows that the accumulation of data (and hence, information) is implicitly and objectively a form of power, consistent with participants' observations in 5.5.4. As long as current service providers are free to collect so much personal information, the information landscape is likely to remain imbalanced and individuals will not be able to acquire ecosystem negotiability.

This insights shows that the most powerful data holders exert huge influence over the digital landscape, in terms of what is *knowable* and what is *doable*. HDR reformers' abilities to balance the landscape are hindered by the fact that they are operating in a landscape that the incumbent platform and service providers effectively control.

10.1.7 Principle 7: Human-centred Information Systems must serve Human Values, Relieve Pain and Deliver New Life Capabilities

This principle aims to combat the challenges outlined in 8.5, that there is a lack of individual demand and a question over the commercial viability of human-centred information systems. Given the earlier observation that life information is what makes data relatable, the insight I offer here is that the way to make people care about their data is to **use it to help them in their life**. The way people find value in data is to connect it their lives. The more that people see relatable life information and can imagine ways to harness that information in their everyday life, the more motivated they are likely to be. By starting with a focus on a user's world, one can then focus in on their life, and then the data that represents elements of that life. Then, the individual has a vested interest; the abstract becomes meaningful and the theoretical becomes possible. Systems and features should be designed from this life-centric perspective, identifying what people want to achieve and designing around these needs. This is known as *value-centred design* (Reber and Duffy, 2005) and it has been argued that this should become the guiding design philosophy in HCI (Cockton, 2004). And to offer true individual value, all human-centric system designs must also consider *context* [2.3.2], *environment* (Abowd, 2012) and *experience* [3.2.1].

However, a key consideration for design in this space is that any technology that handles or represents personal data is designed in a way that is *value-sensitive*. This is distinct from value-centred design, which focuses on shaping a technology design directly around the needs of the user it serves. As Batya Friedman explains (Friedman and Hendry, 2019), *value-sensitive design* is about ensuring that human values are incorporated into the design thinking behind **every** feature of a system.

BBC R&D explored how to better connect people with their data, and in doing so gained clarity on what those values might be. A research project was carried out (Forrester, 2021) that identified fourteen specific Human Values that people seek to satisfy in their lives, which are shown in in Figure 10.5. These are, at the most abstract, goals that people care about in their daily existence.



Figure 10.5: Human Values, as Identified in BBC R&D Research Funded by Nesta

Applying the principles of value-sensitive design, it will be critical to keep these desires in mind when desigining human-centred information systems. If a design does not help with one or more of these values, it will be less likely to succeed. These values may offer designers a way to combat the indifference and apathy of users described in 8.5.1.

Moving from design philosophy to design practicalities, how can we begin to design to a human-centric information system that people want to use? To help with this design challenge, I look to business modelling, where a tool is called the *value proposition canvas* is used. This identifies three ways of conceptualising value: *gain creators*, *pain relievers* and *jobs-to-be-done*. Against this framing, we can design better human-centric functionality that relieves an individual's pain points, helps them complete their tasks, or offers them some gain over the status quo. In the HDR space, given the lack of existing tools for digital life management, we have the opportunity to create quite a unique type of gain: **new capabilities over your digital life** that you have never had before. This ability to do new things has been identified as key ingredient of user empowerment (Meschtscherjakov, Wilfinger and Tscheligi, 2014; Schneider *et al.*, 2018). As 2.1.4 and 2.2.2 showed, a range of novel capabilities are needed for effective PIM.

Here is an example of what this value-centric approach might look like in the HDR space: Myself and BBC R&D colleague Jasmine Cox imagined focusing on address books and

contact lists as a strong relatable starting point to generate demand for a human-centric interface. This could provide people with new life capabilities while also relieving pains. Many people have address and contact information scattered far and wide, and face a complexity they cannot easily manage when it comes to the automated syncing and sharing of potentially sensitive contact information between devices, apps and providers. Developing human-centric personal information management capabilities to bring that messy situation under control would offer a clear and tangible benefit to users. Figure 10.6 shows a possible strategic path built around this need that we mapped out, beginning with detecting ecosystem and life information from the individual's calendar and e-mail inbox, through to building up to more holistic life-level PDS capabilities.

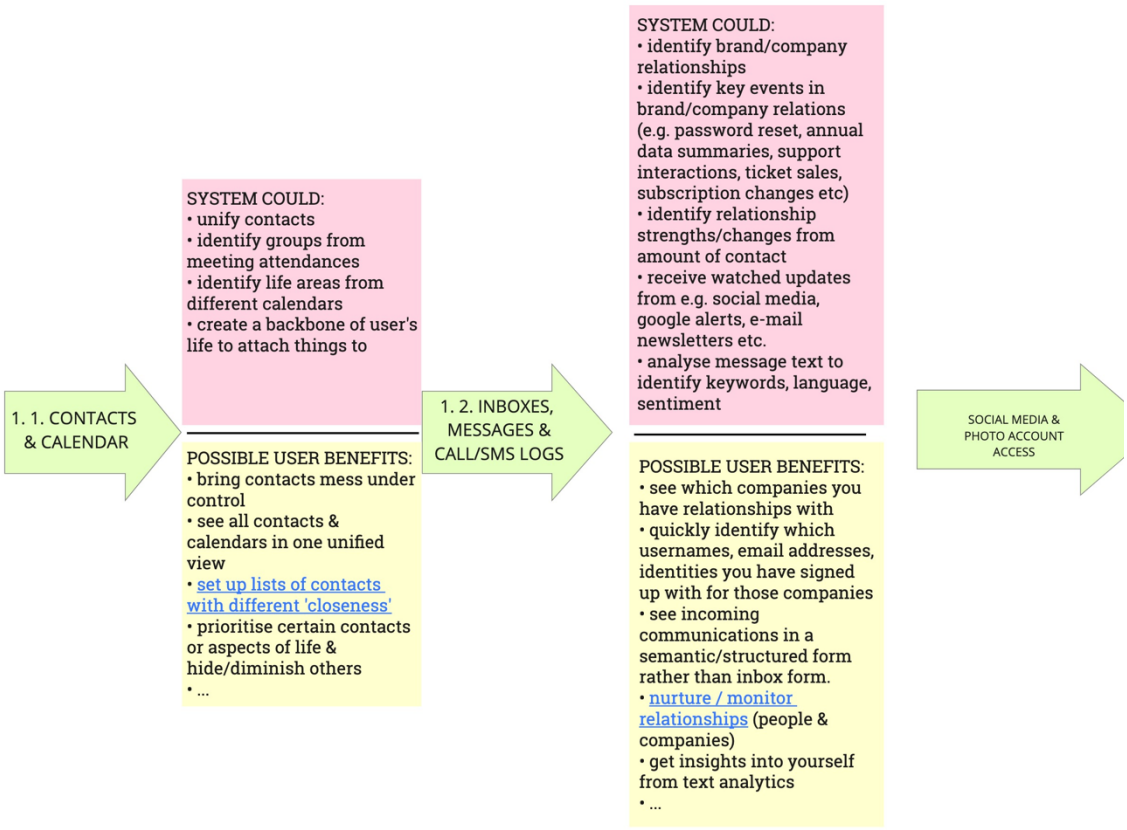


Figure 10.6: A Contact-and-Calendar-centric PDS Approach (left side)

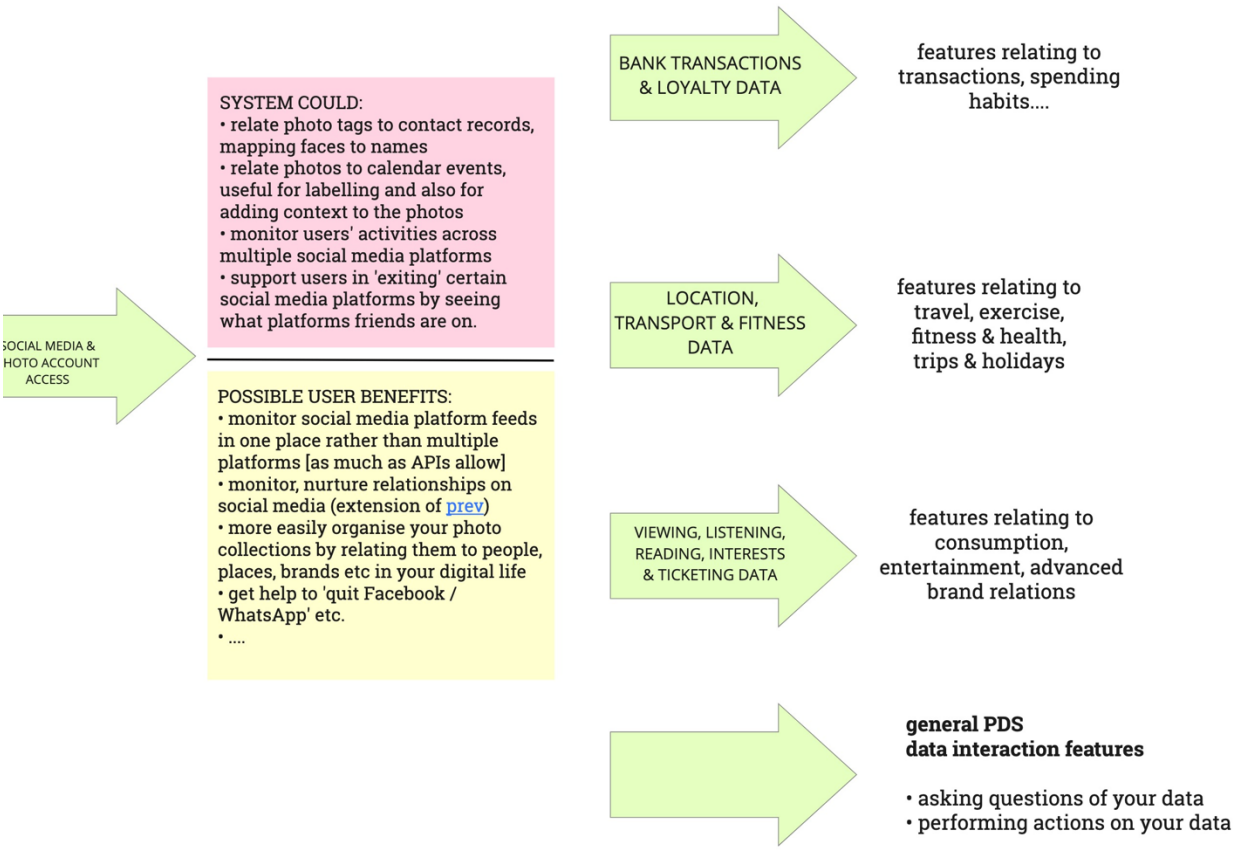


Figure 10.6 (continued): A Contact-and-Calendar-centric PDS Approach (right side)

A helpful example is that of a vacation from my 2011 article (Bowyer, 2011) and shown in Figure 10.7. Today, all the information around such a holiday is scattered into multiple systems - emails, online provider bookings, chat logs, cloud synced photos, web browser bookmarks, smartphone location logs, etc. It is not hard to imagine that a system that was able to bring all related information about that vacation together in one central interface (mock-up in Figure 10.8) could deliver huge value to users and be very compelling. Such context-targeted human-centric offerings can have a much greater chance of generating interest and impact than offerings that merely allow you to 'organise your data' or some other abstract phrasing not rooted in everyday life.

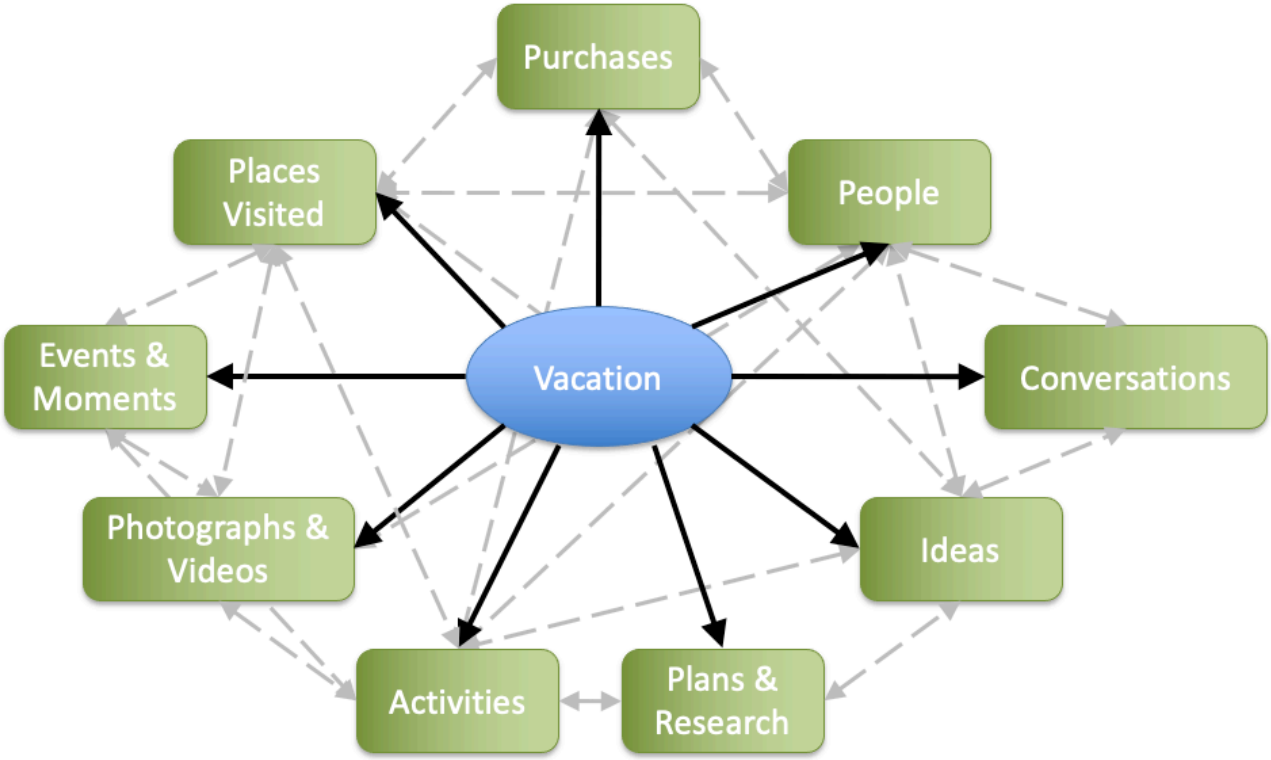


Figure 10.7: The Scattered Data Relating to a Vacation

Figure 10.8: Mock-up of a Unified Interface for a Vacation

10.1.8 Principle 8: We Need to Teach Computers To Understand Human Information

In order to move towards standardised ways to store and unify personal data from multiple sources, computer systems must be taught to understand the information within the data, and how it relates to an individual and the world. This moves beyond just capturing data provenance: put simply, **computers need to understand human information**. They need to move beyond files (Bowyer, 2011) and databases, and begin to perform operations on human informational concepts, and to associate those concepts according to what they mean - i.e. ***semantically***. This is a preliminary step that will enable the building of systems and interfaces that are able to deal in human concepts and represent the elements of everyday life.

We need to store **semantic context and semantic associations**, i.e. the meaning of things, not just raw bundles of data. This is advocated by the Web's inventor Tim Berners-Lee in his vision of a *Semantic Web* (Berners-Lee, Hendler and Lassila, 2001) and by other proponents of *networked* and *semantic* PIM systems [2.2.2]. There is a need to develop standard ways to digitally model facts and assertions about users' lives, so that those disparate pieces of data can be unified, connected, correlated and compared. Some standards are emerging, such as *data shapes* ('ShapeRepo: Make your apps interoperable', 2022). The extraction of meaning from data has a business domain all of its own. Sizable industries have built up around Content Analytics and Enterprise Content Management. But to consider the problem at its simplest level, I offer this insight: Through **the capture of metadata** at the point of data recording, and through **subsequent programmatic analysis** of stored data, as illustrated in Figure 10.9, we can begin to teach computers what the data we store represents.

Systems that store data without any understanding of what it represents in the physical/human world will be less powerful and versatile than those that we can train to understand our world. The emergence of *Large Language Models (LLMs)* like *ChatGPT* in late 2022 show a promising trajectory here, but also carry great societal risk as they only understand human language, not mathematics, logic, science, or how to distinguish truth from opinion, all of which will be critical if accurate, reliable and trustworthy human-centric life interfaces are to be developed.

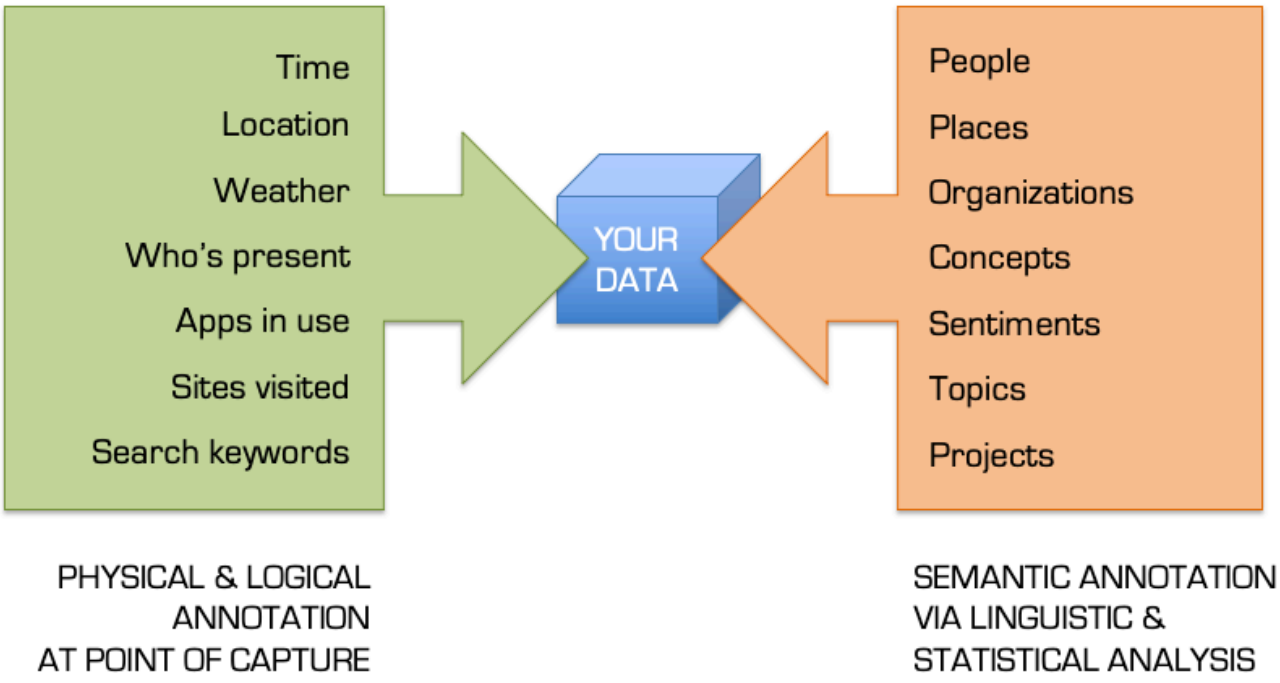


Figure 10.9: Annotating Data with Semantic Context

Machine learning technologies and Artificial Intelligence have pushed machine understanding of human words, images and content to impressive levels in recent years and such technologies can certainly be helpful, but in fact at the core what we are talking about here is something much simpler than AI; It is simply about automatically labelling datapoints in as many different ways as possible (using a similar principle to *lifelogging* [2.2.3]) so that those datapoints can be associatively retrieved from many different angles, and providing humans with ways to amend incorrect labels and to reclassify data or apply new semantic associations. Such approaches are in their infancy, and have not yet been adopted extensively in commercial settings. Issues of interoperability for PDS systems are being actively explored and developed in the *Solid* community [Berners-Lee (2022); Bansal (2018); 9.2] in pursuit of a decentralised web (Verborgh, 2017).

10.1.9 Principle 9: Individual GDPR requests can Compel Companies to Change Data Practices.

In this principle, which again looks at a more sociotechnical design perspective, I will explain how one person can apply the discovery-driven activist approach to **compel** a multi-billion-dollar international data-centric organisation to improve their HDR.

As an avid user for several years of the music streaming service Spotify, I have built up a large library of playlists. I was interest to build an app using my listening data, so made a GDPR request to get a copy of my personal data. When I received that data, I was disappointed to find it was not suitable for programmatic use, because the tracks in my listening history were identified not by any unique identifiers such as `spotify:track:4cOdK2wGLETKBW3PvgPWqT` which I could use to construct clickable song links, just by freeform text strings. Through a long and complicated saga, explained in detail in ARI9.1, which involved much persistence and sending over 30 e-mails in an eight-month period, I was ultimately successful in getting Spotify to improve the format of their GDPR data returns, **not just for me but for all customers who make GDPR requests in future**. I had proven that one individual can use their GDPR rights to exert power over a corporation, with persistence.

A larger scale example of individuals forcing giant corporations to change is seen in the case of Facebook. In the early 2010s, Austrian lawyer Max Schrems began to pressure Facebook to disclose more personal data to their users. He created a tool to enable people to make their own data access requests, which over 40,000 people used. Faced with an overwhelming volume of work and massive liability of future data access requests, Facebook was forced to launch the self-service *Download Your Information (DYI)* download tool, increasing transparency for all Facebook users worldwide (Solon, 2012). Facebook was forced to increase its transparency further when Paul-Olivier Dehaye (now CEO of Hestia.ai) made a GDPR request (later backed by legal action) to force Facebook to disclose more information about which advertisers Facebook had enabled to target him using the *Facebook Custom Audiences* feature. Apparently in order to avoid being embarrassed in court, Facebook updated *DYI* so that every user's downloaded information includes a list of advertisers who have added you to a *Custom Audience* (Dehaye, 2017). Dehaye and Schrems both continue to act as HDR reformers and civic hackers following the discovery-driven activism approach, through their organisations *Hestia.ai* [See Section IV Introduction] and privacy rights organisation *noyb.eu* ('none of your business') (Schrems, 2017) respectively.

10.1.10 Principle 10: Collectives can Compare and Unify their Data and use their Pooled Knowledge to Demand Change.

Continuing the sociotechnical design focus, it is important to recognise that groups of individuals may be able to achieve more than individuals can on their own. Increasingly, the Internet that each individual experiences is not the same as that experienced by anyone else. Thanks to recommendations, targeted ads and social media feeds personalised to your interests, no two people will see the same digital reality. This makes it much more difficult for regulators or individuals to hold digital service providers to account than their analogue counterparts. In recent years, many activists have embraced **the power of collectives** to try and fill this regulatory gap, realising that together, they can discover far more than they can alone, and that through such collaboration there is an opportunity to improve awareness of digital providers' practices.

An example of this is the *WhoTargetsMe* project, launched in 2017 (Jeffers and Webb, 2017). The objective of this project was to monitor political advertising in the UK. Recognising (as larger studies have shown (Bakshy, Messing and Adamic, 2015)) that everyone was seeing different advertisements, the goal was to have each individual report what adverts they see on Facebook, so that these can be pooled and compared with others. Over 50,000 people participated, building up an otherwise unavailable picture of the ways in which different political demographics were being targeted. This is a powerful mechanism available to collectives in this space: the ability to have **individuals obtain their own datapoints and then compare them**.

Another example is seen in the Worker Info Exchange ('Worker info exchange', 2022), a collective that helps gig economy workers such as *Uber* drivers and *Deliveroo* riders to make data requests. Using the pooled data, they conduct investigations to understand algorithmic inequalities and identify unfair treatment of worker by employers. They then help those workers to fight for better working conditions, much like a traditional trade union, but powered by collectively-sourced data. This resulted in Uber being taken to court, and some gains being made for drivers (Lomas, 2021; Foucault-Dumas, 2021).

As the aforementioned case with Max Schrems showed [Principle 9–10.1.9], collectives can be particularly powerful when exerting their data access rights *en masse*, and this can

improve HDR and force greater transparency. René Mahieu and Jef Ausloos have published an exhaustive list of collective actions taken using GDPR rights, addressing issues such as discrimination by US colleges, corporate surveillance of climate activists, identifying gaps in data disclosures, and manipulation of users on dating apps (R. Mahieu and Ausloos, 2020). The authors identify that the GDPR provides an *architecture of empowerment* and have called for better enforcement and for European authorities to provide better support for the ability for collectives to make data access requests together (R. L. P. Mahieu and Ausloos, 2020). Hestia.ai's *digipower* investigation [See Section IV Introduction] concluded that data-discovery driven collectives are a vital step on the road to a more digitally empowered society (Pidoux *et al.*, 2022, p. 70). It is clear that organised collectives exploiting data access rights represent a powerful vector for impactful discovery-driven activism.

10.1.11 Principle 11: Automated Entity Identification could Enhance Machine Understanding and Unburden Life Interface Users.

Having identified the need to assign every piece of a user's life information to a particular partition (or multiple partitions) of their life, it quickly becomes apparent that this would be *too much work* for the user to do alone. Systems that use manual categorisation and tagging to classify information work best with a large userbase to contribute effort to the classification operation (Golder and Huberman, 2006). As part of the explorations of PDS approaches at BBC R&D, I therefore also examined how this challenge might be addressed (considering also that effort could be a deterrent to adoption [8.5]). I identified an approach that could help with this problem: If the *entities* (for example, a person, a place, an event or a topic) associated with a piece of data can be programmatically identified, then a lot of the **assignment of data to life partitions can be handled automatically**. For example, association with your office location would indicate that any data associated to that location is likely to relate to the 'work' part of your life, and this could be done automatically, reducing the effort for the PDS user. The process of **identifying entities within data**, known as *entity extraction* or *named entity recognition (NER)* is a well-established technique, which relies on the trained recognition of proper nouns and keywords combined with the statistical analysis of sentence grammar (Marshall, 2019). This technique is used extensively in text-mining products within the Content Analytics industry such as those produced by my former employer, OpenText (formerly nStein) ('What is text mining and

content analytics?', 2022). However, in the context of a PDS, I propose that new techniques can be applied, making use of the data touchpoints into different parts of an individual's life to identify entities relevant to them personally (including, for example, names of friends or private projects that a standard NER solution would not detect). Data is full of references to entities that have personal relevance in your life. Finding these allows meaningful metadata to be attached to each datapoint. Figure 10.10 shows how a large number of entities could be detected from different parts of an individual's data once it has been imported into a PDS environment:

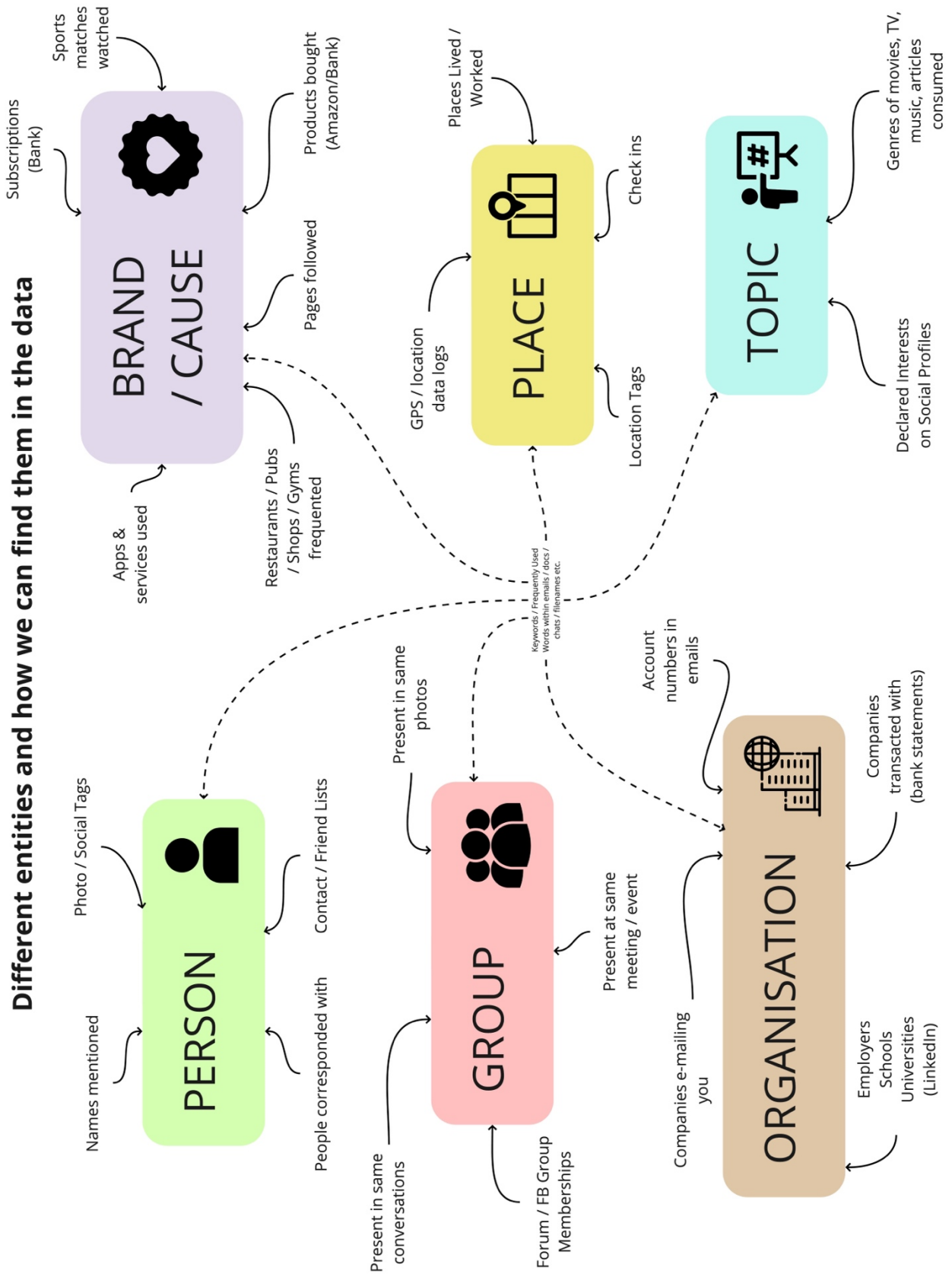


Figure 10.10: Identifying Entity Associations in Data

This sort of approach could be quite powerful in reducing the effort for life interface users. By scanning the data, the most prevalent entities could be identified, and the user need only assign the entities to different parts of their life, as illustrated in the first two frames of Figure 9.10. This would then allow hundreds of associated data points which had been programmatically associated to that entity, to be assigned to the correct 'bucket' or life partition. I was able to prototype this technique successfully to prove the concept, as described in [ARI7.1].

While such an approach would not be perfect, and there would need to be some corrections made by the user, this is far preferable to them having to provide all the classifications and is likely to motivate greater engagement. I have observed in user experience design and consideration of productivity systems that users are more motivated to correct errors, than to fill in a blank page.

Philosophically, we are moving here towards a *learning system*, a system that can be told when it is right and when it is wrong, and get better at classifying things correctly, analogous to the way an executive might train an assistant to anticipate his/her needs better, a sort of *digital life assistant* (Bowyer, 2018a). Bayesian classification techniques could also be used to help with the learning here (Various Authors, 2022). This approach is also useful for *ecosystem detection*–as outlined in Principle 4 [10.1.4]–as identification of relationships with external entities is a key first step to mapping a user's ecosystem.

10.1.12 Principle 12: The 'Seams' of Digital Services need to be Identified, Exploited and Protected.

As identified in 8.4.1, product design (be it hardware or software) is political. Designers pass some power to the user through their design, but also, users should be able to take some power on their own terms. This is the case made by Cristiano Storni in his 2014 paper on *the politics of seams*, Cristiano Storni identified the idea of *empowerment-in-use* which advocates the idea that people need to **appropriate** their technologies to different uses that the designers may not have foreseen (Storni, 2014). This is blocked by current black box, limit-what-the-user-can-do thinking. Central to this capability is the concept of *seams* - those **exposed areas which the user is free to change**. This concept was proposed by Mark Weiser and developed by Chalmers and others (Weiser, 1994; Weiser and Brown,

1997; Chalmers, MacColl and Bell, 2003). Changes such as closures of APIs or removal of ports [8.4.2] can be seem as the removal of seams. As Storni highlights, the availability of design seams is a critical determiner of user power. Companies gain power and reduce agency when they remove or restrict activity at seams. It follows that by identifying, exploiting and protecting the seams of digital services and devices, user autonomy and the viability of data-unification efforts can be protected.

An unseen battle is for the free flow of information is underway at the seams of today's digital products.

Hackers, civic activists and makers seek to repurpose and exploit the edges of products for their own means, while digital service providers and platforms try to block such activity. For example:

- A successful tool called Findings allowed people to clip and share their favourite quotes from Kindle books. Amazon blocked and banned this tool, and the company shut down (Owen, 2012; Maldre, 2012).
- Louis Barclay created a tool called Unfollow Everything, which allowed Facebook users to automatically unfollow all friends and pages, in order to give them greater control of their News Feed reading experience and avoid being manipulated into reading more than they want to. He was banned for life from Facebook and threatened with legal action should he ever build any tools that manipulated the Facebook experience (Barclay, 2021).
- Various activist groups have for several years been fighting to give individuals the legal ***right to repair*** their own products (Miller, 2021), which has often been blocked through planned obsolence, inaccessible seams or restricting access to parts. The problem has been described as ***device tenancy***, the idea that our relationship with our technology products is more like a tenant, where a landlord retains overall control and permits us to perform certain activities (Tufecki, 2019). New laws have been introduced in the EU (Tett, 2022), forcing companies to support individuals to repair their devices. Apple has subsequently released self-service repair kits, though these themselves force parts to be paired with particular phones, limiting the utility of self-repair (Moore-Colyer, 2022).

- As detailed in my co-authored paper with Louis Goffe and colleagues (Goffe *et al.*, 2021, 2022), ***web extensions*** and ***web augmentation*** offer a powerful technique for modifying web experiences and repurposing user interfaces. This is because once a website is loaded into your browser, it is no longer under the control of the remote site, and by creating a *web extension* to run code within your local web browser that loaded website can be edited, *scraped* (P., 2021), or otherwise repurposed. This has been successfully used to stop clickbait, dispute fake news, combat addiction, filter explicit words and more. However, in order to re-assert control over these customisations, Google has announced changes to the way Chrome extensions will work, which could 'stifle innovation' and limit what developers can do within the web extension (Miagkov, Gillula and Cyphers, 2019).
- An example from 2016 shows how seams can be exploited to obtain information and increase transparency. By brute force querying of a Facebook API, researchers were able to identify a complete list of 282,000 interests on Facebook and identify the relative popularity of each interest. (Havlak and Abelson, 2016).
- A number of HDR reformers, myself included, had identified a new seam for subverting some of Facebook's control over how its content is consumed [8.4]: *accessibility tags* or *ARIA tags* (Various Authors, 2022). These are specially marked-up tags in HTML web pages used by screenreaders to display or read content in a more accessible way for partially-sighted or blind people. Because these show page content in a standard format (whereas the HTML of most web pages varies widely and often changes), they present a reliable way to more easily *scrape* content from the loaded web page within a web extension. In experiments at Open Lab, posts were successfully scraped from friends' feeds (which Facebook do not make available anywhere except the News Feed) so that they could be consumed separately in a more human-focused user experience. This technique has been used successfully to monitor Facebook ads by NYU's *Ad Observatory* (Watzman, 2021), and was used by *WhoTargetsMe* [Principle 10–10.1.10]. In 2021, Facebook was found to have deliberately obfuscated content within ARIA tags to prevent such investigations, resulting in visually impaired users being *unable to differentiate* ads from posts, and *hearing junk characters* read aloud. This can be taken as an adversarial stance against researchers, activists and HDR reformers, and shows that companies like

Facebook will go to extraordinary lengths to assert their dominance and reduce user agency (Faife, 2021).

- One reason why many companies and services have produced apps is because these are much more locked-down and controllable than the web browser environment; there are fewer seams. However, adopting the same philosophy as using web extensions to modify web-based experiences, and drawing on data flow auditing technologies like TrackerControl [9.3] researchers at Oxford University have now developed techniques by which mobile apps can be reverse-engineered and modified to change user experiences to better meet users' needs, offering the promise of a *right to fair programs* (Kollnig, Datta and Kleek, 2021).

These examples make it quite clear that Storni was right: **product seams are the place where control can be asserted or regained**. They are the setting for an ongoing battle for the freedom and integrity of today's information landscape, and it is important for HDR reform that this space is specifically targeted. The role of the HDR reformer here is twofold:

1) To **surface information injustices**, especially the closures of seams. |
2) To **push or 'hack' the seams to gain transparency and re-assert control**, including gaining access to otherwise inaccessible data and to acquire new functionality.

In this context, the work of whistleblowers such as Frances Haugen (Horwitz *et al.*, 2021) and Edward Snowden (Macaskill *et al.*, 2013) is particularly validated and important. Whistleblowers can expose internal practices that harm the information landscape's integrity that are not otherwise visible. In order to hold online platforms to account, the public must be **aware** and **able to attribute** any restriction in freedom or information access to the correct source. They need to know that the information or functionality is being modified or restricted. These ideas are explored further in (Bowyer, 2017). Seams should be much more in the public consciousness than they are.

10.1.13 Principle 13: We could Demonstrate Business Benefits of Transparency and Human-centricity

As outlined in 8.5 and in this section, it is not just possible but essential that work is done to persuade data-holding organisations of the benefits of moving towards the new paradigms

outlined in this thesis. The following avenues for possible future research and advocacy toward data holding organisations have been identified:

- **Trust & Reputation**: In line with the third *public relations*-like aspect of HDR [7.2] as well as the recommendations in 4.4.4, 4.5.1, 5.6.2 and 6.2.1, displaying a more inclusive, open and supportive attitude to data handling could strengthen the service relationship and increase customer loyalty and trust. Organisations that are seen to have good Human Data Relations are preferred.
- **Consent**: In the wake of the GDPR, ensuring consent is becoming an increasing concern to organisations, and the risks of legal consequences for mistakes are high. It makes sense that a more dynamic [Bowyer *et al.* (2018); 4.5.1; 5.6.2; 6.2.2] consent approach that involves individuals [6.2.3] and keeps them in the loop would enable them to speak up much earlier and express consent wishes that might otherwise go undetected.
- **Accuracy**: The best-placed person to spot errors in data's accuracy or fairness is the individual about whom the data is concerned. Therefore, increasing their involvement is likely to improve the quality of the data, especially if additional data is contributed or curated by the service user [4.4.3, 6.2.3]
- **Liability**: In an increasingly litigious society, storage of personal data, especially health or financial data, is a significant liability for businesses, especially if something goes wrong. Investment in human-centred personal ecosystems would outsource the storage of sensitive data to data trusts or PDS providers, reducing liability for the service business. By ensuring that data is accessed only in ways that are centralised outside of the business and remaining in the user's control—such as PDS company digi.me's Private Sharing model (digi.me, 2019; Bowyer, 2020a)—organisations can ensure that have negligible risk of mishandling customer data.
- **Better Customer Targeting:** The most radical, but perhaps the most persuasive, business model relating to better HDR, is the VRM approach [2.3.4], where individuals express their own service or product desires explicitly, which vendors then respond to. This turns traditional models inside out, and would empower users more, but due to the inherently improved accuracy of a self-declared interest, might also give businesses a greater confidence that their investment in converting those

customers to a sale would be worthwhile. It is important to remember that the current drive towards collecting more data that drives the platformisation trend is in order to improve ad targeting, so that businesses can get a better return on their investment. A VRM approach, or any other approach where the individual contributes improved data to their data self, is in line with that current business objective.

The 13 Principles that have been asserted in this section are by no means the only principles that guide the HDR agenda, nor are they necessarily the most important. However, they are the ones that I have been discovered, understood and refined through the pursuit of the six human needs [Chapter 6] through my embedded involvement with the four projects [See Section IV Introduction] as well as self-experimentation and prior design thinking. They are located here within the conclusion of the thesis because they constitute the most tangible and actionable learnings of the thesis; the first steps that will guide an HDR reformer as they embark on the pursuit of the HDR agenda in their design, research, transformation or innovation work. The remainder of this chapter will reflect upon the challenges of pursuing that agenda and the legacy, impact and limitations of this body of research.

10.2 Pragmatic Reflections on Pursuing Better HDR

This section will reflect upon the proposed objectives and approaches towards better HDR laid out in Sections III and IV, from a critically pragmatic and broad perspective. One of the first things to note is that while Chapter 8 identified many of the clearest obstacles to the HDR objectives, it should by no means be considered exhaustive or complete. It is important to recognise the uphill struggle that faces anyone pursuing an HDR agenda, and that there are many forces working against the HDR reformer, including but not limited to commerce, resistance to change, insufficiently effective regulation, insufficent funding, technical challenges, disinterest by almost all parties, and the normalising effect of the status quo. HDR requires not just a halt of current data-centric trends but an about-turn and pivot to a completely different model of organisational personal data handling. In the Principles, insights and Approaches presented in Sections IV and V through which I attempt to operationalise the HDR agenda, I lay out a direction of travel, highlighting the actions that appear to be the best chances to succeed. The HDR agenda is about **asking the right**

questions, not about providing a complete or definitive answer. The HDR agenda is explored both from within the system and from outside the system, and it is possible that even with parties pursuing both conformist and activist approaches in parallel that this may not be enough to bring about the desired change. Nonetheless, the cause of better human-data relations is clearly a worthwhile one for society.

The challenges of better Human Data Relations are intrinsically sociotechnical. They cannot be solved by technological design, by law or by social change alone. This was one of the reasons why it was important to lay out a variety of approaches in the recommendations from the Case Studies in Chapters 4 and 5, and in the HDR approaches in Chapter 9. One benefit of this, however, is that practitioners from many different fields ranging from technologists and designers to journalists, activists and civil servants should be able to find something in the Principles or Approaches that they can build upon. One of the underlying needs identified in the findings of Case Study Two [5.6.1] but also where Chapter 9's approaches hope to influence governments, is that changes in law are required. Like other jurisdictions, both the EU [9.5] and the UK are, at the time or writing in 2023, developing new or changed laws that will affect the use of personal data online; but their interests are not only to protect their citizens, but to help businesses too, not to mention any political interests of those in power. As such, we cannot expect that all legal changes will align with the HDR agenda; in fact, in the UK there are serious concerns that new data protection laws set to replace GDPR following the UK's exit from the EU could *'seriously weaken data protection rights'* and cause harm to marginalised communities (Cowburn, 2023). Even the existing GDPR laws, which are designed from a more individual-centric perspective, fail to deliver sufficient benefits in practice, as Case Study Two showed [5.5.1]. So, it is important to acknowledge that while regulation is one of the most promising routes to achieving better HDR, it may not happen and it may not be sufficient even if it does happen. It would also require a more paternalistic attitude to protecting individuals, which goes against the libertarian trends of the early 2020s political landscape in the West.

In 1.2.3 and 9.6.1, I highlight the need for better education and data literacy around Human Data Relations: the need for better understandings about the intrinsic personal value of data as a source of life information, and the need for better awareness and capability when it comes to exercising data rights and recognising exploitations of power or diminishments

of individual agency. This should not be taken as a criticism of the public for not knowing enough, nor does it imply that it is the responsibility of 'the people' to demand fairer HDR from providers. We need better data literacy at *all* levels of society, and perhaps particularly amongst policymakers, public and civic service providers, and politicians, as well as journalists and public figures of all kinds. By targeting those individuals who shape future legislation and have the greatest capacity to influence business and societal norms as well as public opinions, we stand the greatest chance of creating impact. So far, the problems of disempowerment over personal data represent a systemic failure to recognise the negative impacts upon individuals and act in the interests of individual liberty across all aspects of society—including technologists, lawmakers, and media. If the issues are properly understood and pursued by those in positions of power and influence, which is not easy given their complexity, then it should not become a burden upon citizens to become aware of injustice and demand better HDR, but rather the system would already be changing to meet their needs through the actions of those who shape the systems and processes of society. This is why projects such as digipower [See Section IV Introduction], which aimed specifically to raise awareness of digital power issues among influential European VIPs, and participation in policy consultations by experts such as my contribution to the GDPR Guidelines consultation [9.5.3] are vital. HDR reformers will need to continue to educate and influence those who are best placed to change society.

The issue of individual consent to data sharing has been covered at length in this thesis. In the pilot study [Bowyer *et al.* (2018); Appendix A] and Case Study One [4.5.1], the framing of consent is used to consider whether or not individuals feel sufficiently consulted in matters pertaining to the use of their data, with the conclusion being that dynamic consent is required, and that informed consent produces a 'point of severance' beyond which individuals lose control. However, it is important to recognise that consent is just one of the lawful bases for data processing ('A guide to lawful basis', 2022), and that data holders may decide to rely on a basis such as 'legitimate interests' or 'fulfilling a contract' which reduce the requirements upon the data holder. In particular, the former removes any ability for the individual to object to processing or to have data erased. There is a trend to move away from consent, as TikTok recently attempted to do, and as was rumoured by a local authority in the Early Help context [8.4.2]. It is important to recognise, therefore, that this is a possible vector through which data holders might act against HDR and reduce individual

data rights over the personal data they hold in future. Just as with the proposed changes to UK data protection law mentioned above, this would represent a serious diminishing of individual agency, and the ability of activist and research groups to investigate their personal data [9.3.1] would be crippled by such a change.

The reliance on personal data rights is also not a universal right that covers all cases in which data is held about individuals. Personal data is a tightly-defined term relating to data that does or can identify individuals [ARI2.1]. Holders of anonymised, pseudonymised or aggregated data, or of data which from which individuals are not easily identified (for example crowd CCTV recordings, traffic flow data or Large Language Models (LLMs)) commonly argue that rights such as GDPR do not apply to their datasets. Such datasets fall into legal, technical and moral grey areas. People still feel a sense of ownership over such data, and perhaps rightly so. For example, Apple—lauded for its improved approaches to privacy—states in its own privacy policy that it may keep information extracted from Siri voice recordings but only once they have been de-identified and 'attached to a random identifier'. This means that content of that voice command is held by Apple and may still be revealing or personal, but is now no longer accessible or erasable by the individual nor subject to any data protection law. Several researchers have shown that anonymous data can be 're-identified', in one study by using as few as 15 points of data (Bradbury, 2019). As such, we can see there is a gap in individual rights and protections over some types of data. Morally, individuals still deserve a relationship to that data and the information within it. Whether this moral claim can be actualised in law or technical capability remains to be seen. The findings of Case Study Two would suggest that these 'rights removing' approaches, once understood by the individual service user, would harm their trust in their provider. Furthermore it could have the potential to generate feelings of having no choice [4.4.3] or of powerlessness [5.5.4].

The principle of unifying and uniting data and hence life information [Principle 2–10.1.2] would clearly be of significant value to individuals in their quest for better HDR, allowing patterns and behaviours across devices, platforms and services to be visualised and reflected upon. However, it may be in some cases unachievable due to technical, legal or corporate policy reasons. An example of this was experienced in the SILVER project's attempts to build a health data interface for Early Help support workers [see Section IV

Introduction]. Prior to the building of the prototype, support workers reported they experienced the sense of a 'wall' between different public sector agencies, especially between healthcare providers and social care providers. Employees of healthcare providers felt they would be violating internal policies and potentially law if they made personal data available, even if it would be in the individual's interest to do so. And similar barriers were encountered when trying to build technical channels for data-sharing. Even once access to the requisite APIs was acquired, local authority and local health authority legal departments debated and negotiated at great length over how to establish suitable data-sharing agreements. This made development particularly difficult, as we had to build user interfaces for data that we could not access, but could not access the data until the system was built. Extrapolating this problem more generally, it is clear that (particularly in the public sector) various agencies consider (rightly so) that they have duties of care not to share individuals' personal data with anyone under any circumstances, even if the access would be programmatic and then limited to staff who shared a duty of care to the individual or had a need to know. Clearly there is a question of service provider liability when it comes to certain categories of public data. We could also imagine similar challenges in allowing any kind of access to commercially held data, even via API, in areas of sensitive data such as therapy, welfare services, dating sites or legal or professional personal services. These challenges make it especially hard for third parties and developers to help individuals to unify and unite their own data, because they cannot (even programmatically) access personal data on the individuals' behalf. Better data standards [8.5.4] and proxy representations [Principle 4–10.1.4] may help technically, but we can expect much resistance from organisations with strict policies about access to user data (which, in the current litigious landscape, are quite common). Additionally, it is important to recognise that the uniting and unifying of personal data may present additional risks, for example where linking data enables identification that was not previously possible, or breaks expectations of privacy between individuals in a multi-party data set such as that for a family or household. Revealing the provenance of information may also break expectations of privacy in some contexts. Any systems and processes that unify and unite data or reveal its origin must therefore be design with the utmost attention to both individual and intrapersonal privacy, in line with *privacy by design* (Cavoukian, 2010) thinking.

The pursuit of better HDR is fundamentally a design problem. The designs it demands, however, may not always be implementable in practice. For example, some companies simply do not provide APIs to access the data that people want to access, nor are they compelled to do so. As experiences with Cornmarket's PDS prototype showed, building user experiences that enabled people to obtain data via GDPR requests and download portals and then upload them into the PDS proved not only technically difficult and hard to design, but these user experiences were very complex and unsatisfactory to users, and worse still would need to be repeated whenever an up-to-date view of data was required. As Case Study Two suggests [5.6.1], something that would help greatly here is a recognition by policymakers and data holders, that one-off, manual access to data is inadequate, and that really what people need (both programmatically as data and in human-readable form as information), is a live view or feed of their data. This could be integrated into Human-centric life interfaces and ecosystem viewers, which would become easier to design, build and use. This would represent a far less cumbersome user experience. Interestingly the desire for a feed of data changes also surfaced in Case Study One [4.4.3] as something that would be valuable to both parties, validating the need to view data as dynamic and changing rather than static. It is clear that to build workable solutions in these areas that meet user needs will require multiple specialisations: in data rights law, in user experience design, and in specific domains such as commercial data holding or social care, and it is hard to see how these changes could be co-ordinated by any one party. It will be especially difficult for activist researchers, especially when efforts to build alternative ways to access platform-held data feeds can carry costly personal penalties, as in Louis Barclay's case [Principle 12-10.1.12; Barclay (2021)].

The technical and design challenge of building unified human information interfaces is a 'grand challenge' (Lufkin, 2017) level of problem. This goes further than the organisational barriers and resistance to change described in 4.5.3. To build what is in essence an operating system that deals in life information and ecosystem information concepts rather than files and data records is, to put it mildly, non-trivial. That does not mean that is not necessary or worthwhile to engage in this endeavour though. Human information modelling, such as that described in 9.4, could help, as could semantic information standards [8.5.4] and automatic entity identification [9.4.3]. Since a key part of such interfaces relate to both human and machine having a greater understanding of information

with data, past guidelines on legibility [2.3.2] and effective access [2.1.4], as well as all the lessons from the PIM space [2.2.2] about how people need to interact with personal information, are also important. It may also be possible to take lessons from the explainable AI space, where a report has identified four principles: *Explanation*, *Meaningful*, *Explanation Accuracy* and *Knowledge Limits* (Phillips *et al.*, 2020). However, the understanding of such explanations should **not** equate to the acquisition of technical expertise. As with explainable AI, being able to explain a highly complex technological construct is not useful to an individual if it requires them to gain expertise and interest in the technology itself, because this creates issues around technical skills and whether or not the individual is inclined to take an interest in technology. As Mark Weiser envisioned (Weiser and Brown, 1996), technology should ideally recede into the background of our lives. Thus, explainability and understandability should also be very much be anchored in the human side of human data relations, rather than in the technology and the data. The focus should be on bringing the technology into the human mental landscape, not bringing the human into the world of the computer (and even when learning about ecosystems, which might seem to go against this, explanations can be presented primarily in terms of companies that the individual has relations with, and third party entities, rather than servers and databases etc.).

Perhaps the most challenging practicality of shifting towards a human-centric paradigm for data and information relations, relates to the amount of effort that is currently required for people to take charge of their digital life. Even a motivated individual such as the author of this thesis this finds it very difficult to achieve a meaningful awareness and control over their digital life. A further problem is seen that in many cases, the motivation is not there, both in terms of apathy or powerlessness as mentioned above [4.4.4; 5.5.4], but also in that seeking better Human Data Relations is currently hard work that few are inclined to do. It is easier to use the system as a consumer than to participate in one's digital life as a citizen (Alexander, Conrad and Eno, 2022). There is work to be done in winning hearts and minds [9.6]. However, the objective of the Human Data Relations agenda is not to place an undue burden on the citizen. The identified need for improved data literacy [9.6.1] applies particularly to those with the power to influence policies, technologies, civic or commercial structures, processes or public opinions, so it is important that they be persuaded act upon this new understanding in their professional capabilities not just through their role as users of digital technology. Another way of thinking about the the issues of digital power over

data is to compare this to another complex issue: personal finance. It is in every individual's interest to acquire utility and control over their personal finances, but this does not require every individual to take an interest in tax law or become an expert in accounting ledgers. Society has responded to this need by the production of specialist service providers–accountants and financial advisors–who can help individuals who do not wish to engage with the detail to manage these aspects of their lives. Self-service budgeting and tax software is now available as well for those with lower budgets. This is the model that a world of better Human Data Relations could also adopt. Already data access and understanding services are emerging [9.3.4], and this sector will inevitably grow as demand and investment increases. As with financial responsibility, or environmental impact, or personal safety, there is a certain amount that individuals can do for themselves, but the greater impact and moral responsibility to really bring about change lies with lawmakers, regulators and the organisations that create the infrastructure of society, be they in the private sector or public sector. The approaches presented in Chapter 9 are only first steps the beginning of possible trajectories to improving society towards a more human-centric and individually-empowered one. Of all the quadrants for change [Figure 9.1] it is the external ones that are most critical. The structures of society must change (bottom-right quadrant), and only when this occurs can people be offered a meaningful choice. Only when people are able, and inclined, to take that choice, between the data-centric status quo and the human-centric future, can our human data relations–the way we live our lives in a data-centric world–actually change in a meaningful way (top-right quadrant).

10.3 Limitations and Future Work

This section will examine the limitations and future potential not of the HDR agenda, but of this thesis and this body of research itself.

Methodologically, this research achieved significant success in helping participants engage with the world of data. The use of data cards [Figure 3.7] across activities in the Early Help context was particularly successful, because it converted dry data concepts into relatable life information concepts, but also because it gave participants boundary objects and things-to-think-with [Bowyer *et al.* (2018); Brandt and Messeter (2004); 3.4.2; 4.5.2]. The use of storyboarding cards [Figure 3.11; ARI4.3] to explore two-party interactions in place of more traditional UI-sketching approaches was also novel and effective. Such card-based

methods for engaging people with data could be developed and explored further, and indeed there is clearly still potential to do. This was not done within this PhD in part because the COVID-19 pandemic directed the research trajectory away from in-person work, and additionally the availability of the GDPR enabled more data-focused approaches. In fact, even though it did not use tangible data cards, the structured conversations around aspects of one's own life data used in Case Study Two and subsequently in the digipower project [See Section IV Introduction] proved effective and would also be worthy of further exploration as a methodology [1.2.2]. Any and all of these methods could be developed further to produce useful toolkits for educating people about data policy and law, for example in helping them understand proposed changes to UK data processing in health care with the confusingly-named GPDPR (Shemtob, 2021). Another methodological limitation can be observed with regard to the use of GDPR rights in the research; The methods employed in Case Study Two purposefully did not explore how well participants could carry out their GDPR rights without aid. The process was a guided one, and some participants freely admitted that without aid they might not have been able to commence or conclude the GDPR processes. The design of this study took a value judgement that there was greater benefit in educating individuals about GDPR practicalities, and in assessing companies against a level playing field, than in measuring GDPR education levels in the public. A study to explore this would be worthwhile and could highlight areas for companies and regulatory bodies to improve their public engagement around the use of data access rights. Despite it being a guided journey, great care was taken not to influence or bias participants' judgements around GDPR, which is why education and questioning focused on participant evaluation of the GDPR experience against privacy policy claims and GDPR rights, rather than against the ideals of the emerging research agenda.

As described in 3.2.2, this PhD used an unstructured process of feedback between Case Studies, individual experimentation, modelling and synthesis work, and the project placements. This means that (as mentioned in footnote 3) the study designs and overall research approach largely did not report findings back to participants to acquire additional feedback, unlike more formal Participatory Action Research approaches (Lewin, 1951; Chevalier and Buckles, 2008). Given further time and resource, additional work could have been carried out to include participant reflections upon findings or design critiques of HDR proposals into this body of research. However, this does not mean that findings of the

studies were not validated. In general, evolving findings and learnings were validated with participants in different groups or settings. As Chapter 6 shows, there were strong correlations between the Case Studies; each helps to validate the findings of the other. Within Case Study One, the Sentence Rankings exercise [4.3.5] served to validate findings from earlier research by the author and by the SILVER project with participants. In the case of Case Study Two, the digipower project [see Section IV introduction] (which used a very similar methodology) helped to validate participant experiences from the study. The HDR agenda evolved through regular exposure to real-world settings in the BBC R&D and Hestia.ai placements, which added a further feedback and validation cycle of the discursive contributions of the thesis. Every element of this research has involved engagement with members of the public to inform the evolving designs, in line with UCD [3.2.1] and design justice (Costanza-Chock, 2020) philosophies. Another opportunity would be to conduct further workshop-type activities involving parties representing individual, collective and establishment points of view. One way to do this would be to take the findings from the industrial track of the thesis back to the stakeholder groups from the participatory track. More generally, such multi-party investigatory research continues to take place in organisations such as BBC R&D, Sitra, Hestia.ai and The Citizens as well as many HDI-focused academic research labs. Where there is a significant opportunity for further work would be in engaging HDR-adjacent activist and grassroots groups and movements such as those listed in 7.6. It is important to acknowledge that this researcher has consciously engaged in activist movements around HDR reform through his engagements with Hestia.ai, The Citizens, and the Open Rights Group [9.3.2], not just to serve the research agenda of this thesis but also due to personal interest in better digital rights, which could invite accusations of bias. As a counter to this, we can note that such engagements have been balanced against the author's lifetime of experience 'within the system', working for commercial technology providers of various sizes including IBM, Open Text, InterRail and Logica. Furthermore within the timescale of the PhD, there were also multiple engagement with activities 'within the system' in such as through SILVER/Connected Health Cities and at the BBC. In the SILVER context, the collision of traditional data-centric approaches of the establishment and the researcher's human-centric agenda sparked significant debate and reflection among the research team. Established service-provider-centric values were inadvertently embedded into the project's solutionistic framing resulted in the production

of a health interface [see Section IV introduction] that favoured staff perspectives over supported family desires uncovered in the project's own research. This shows that even within this PhD, there has been extensive reflection and challenge to the ideas from different perspectives. Based on these reflections, every effort has been made to ensure that the approaches in Chapter 9 and the Principles can collectively support use within or outside of the establishment.

As laid out in 3.1, the research approach of this thesis is unashamedly aligned to an **individualist** philosophy of personal betterment. As such, it is important to recognise its gaps around exploration of collective sensemaking of data and the identification of data-related objectives that further group or community interests over individual ones. For example, Maskell's Spokespeople project explored how communities might share geolocated data from their cycling journeys to pressure local authorities to improve cycle lane provisions (Maskell *et al.*, 2018), and Elsden's Metadating project looked at how couples might understand and reflect upon their romantic lives through personal data (Elsden *et al.*, 2016). The issues of complex interplays of different individuals' personal data within families and households was present in participants' experiences in both Case Studies, but was not actively explored within this thesis. Indeed it could be argued that some of the HDR principles might create tension when viewed from a group perspective, for example the knowledge of provenance of data [Principle 5–10.1.5] could have implications for lost anonymity within households or communities. Activist groups such as Hestia.ai and noyb.eu have explored and continue to explore how collectives can work together to pursue their interests and rights through data, but this is limited by the individualist framing of the GDPR and its equivalents. As an example, there is no means for a community or group to request its collective data from a provider, and every individual must make their own request. While GDPR does provide mechanisms for third parties such as data understanding services or lawyers to make data requests on multiple individuals behalves, experiences at Hestia.ai showed this is rarely well-supported by data holders. Indeed in this regard the focus on GDPR as a data access methodology may be limiting, as it risks brushing over the fact that there is a gap in law around the protection of group data that needs to be confronted. Collective data management is not well accommodated for in system designs or policy designs today (Baker and Karasti, 2018). Another aspect of this limitation is that much data today is inherently social and interpersonal, such as social network graphs,

family media accounts or shared banking and retail accounts. There is a personal data (and thus human information) ecosystem within individuals' own lives as well as the external ecosystem held by providers. The matters of collective privacy (Taylor, Floridi and Van der Sloot, 2016) and collective harms through data such as through algorithmic profiling (Tisné, 2020), and the exploration of collective rights over data (Mühlhoff, 2023) are worthy of further study.

In this thesis, a deliberate decision was taken to refer to the distinct legal concepts of *data processor* and *data controller* collectively as *data holder* [ARI2.1]. This was useful for simplifying explanations and was easier for participants to understand. However, it has the consequence of brushing over some important distinctions. A data processor carries out processing of data on a data controller's behalf, and hence does not bear the same legal responsibilities for data protection (for example, they are not required to satisfy Subject Access Requests). As with the decision to use a legal basis for data processing other than consent, companies can use the concept of *joint controllership* to declare their data processing partners, or sometimes the data subject themselves (Edwards *et al.*, 2019), as a data controller. The digipower project found evidence of at least one retailer declaring hundreds of joint controller relationships with minor and unheard-of tracking and cookie companies, making an unmanageably complex picture for users (Bowyer, Pidoux, *et al.*, 2022). The declaration of joint controller relationships can result in a situation that where it is difficult to clearly delineate responsibility (Colcelli, 2019), and can create additional burdens for smaller companies (Loon, 2018). These complexities can make data rights become very difficult to apply, such as Internet-of-Things [IoT] environments (Mill and Quintais, 2022). The case of doorbell cameras has become especially complicated after court rulings have determined that owners of doorbell cameras can be considered data controllers and thus liable to large fines (Milmo, 2021) if they cannot satisfy GDPR requests correctly (something which a regular citizen would find very hard to do). This has raised serious questions over whether home IoT and home surveillance technologies are empowering and viable for their stated purpose (Kelly, 2023). The net result of all of these issues is to make a situation that is already complex and hard to understand for individuals become even more so. There is significant scope for future work to explore the complexities of data relations between organisations, especially with regard to the differing legal terms and responsibilities. The exchange of personal data between data brokers is extensive and

particularly poorly understood and under-regulated and deserves scrutiny (Keegan and Ng, 2022).

At an ideological level, the design thinking around how to improve data relations in line with identified wants [Chapter 6] are heavily influenced by human-centric idealism, as shared by PDE and MyData practitioners [2.3.4]. While human-centricity is a noble ideal (from an individual or social good perspective), it is important to acknowledge its limitations. Coulton and Lindley argue that a focus on the relations that a human has with his or her world is reductive, and that better *'More-than-Human-centric'* design could be achieved by focussing on a metaphor of *constellations* and a recognition that focusing only on the human perspective can cause aspects of the design context to be overlooked (Coulton and Lindley, 2019). Elisa Giacardi argues that a model where we consider the *co-performance* of humans and automated systems could unlock new design thinking (Giaccardi and Redström, 2020); indeed it is clear that with the emergence of LLM-powered chatbots such as ChatGPT, such a paradigm shift in HCI design thinking will be vital. Wakkary *et al.* consider that humans and technologies are fundamentally entangled in complex relations which require the use significantly more advanced philosophical lenses to be effectively designed for (Wakkary, Oogjes and Behzad, 2022). Pragmatically, however, simple human-centric visions such as the MyData principles (MyData, 2017) and the Human Data Relations agenda of this thesis, do seem to be useful mental models and sensible 'first steps' for research and design innovation as we move beyond data-centric and human-computer interaction thinking. The more advanced mental models of human-centricity's critics will surely be useful once some of the first steps toward human-centred personal data ecosystems have been realised.

In considering the possibilities for future work in improving people's relationships with and through data, the Human Data Relations agenda can be considered as a design challenge at multiple levels. It could be used by interface designers to consider how to better represent data as meaningful life or ecosystem information to users [7.4; 9.4.4; 9.4.5]. It could be used by user experience designers to rethink how existing interaction journeys can become more engaging, empowering and have a lasting value to users [4.4.2; 5.6.2]. It could be used in service design or business process design to consider how service relationships can be redesigned to be more involving and inclusive, or to achieve greater levels of data accuracy,

data use consent or improved customer loyalty [4.5.3; Principle 13–10.1.13]. It could be used in the design of law or policies to give people more effective digital rights [5.6.1]. It could be used by public service providers in the design of better public engagement programmes or social interventions 4.6. It could be used by activist groups to design campaigns or initiatives that might have greater impact [9.3; 9.5]. It could even be used by individuals to design better systems (in other words, to make different provider and app / workflow choices) for managing their own digital lives [5.6.3; 9.3; 9.6.1]. All of these areas and more are suitable for future design and research work building on the kernel presented here in the Human Data Relations agenda.

10.4 Personal Reflection

As an experienced software engineer, power user and technology blogger who had considered the loss of digital agency for many years [1.1], my journey into this research space was an unusual one; I arrived with already-formed ideas about the nature of the problem. This was not an ideal match for the traditionally participant-led approach of HCI, where ideas and insights normally arise solely from one's participants. However, through the discipline of the *Digital Civics* programme and the process of publishing peer-reviewed papers, I was able to approach the topic in a more objective fashion. Recognising that HDR issues would be unlikely to surface organically, I was able to use careful sensitisation [3.4.1], balanced and open questioning and neutrally-designed stimuli [3.4.2] in a way that elevated participant experience to be the primary source of data, to produce findings and discursive conclusions that are as much the participants as my own.

Along the way, I discovered vital areas of literature and existing work, most notably the foundational work of Weiser, Abowd, Crabtree and others [2.3.1; 2.3.3], the sub-discipline of *Human Data Interaction* [2.3.2], the emergent innovation around *Personal Data Ecosystems* and *MyData* [2.3.4], and the work of activist collectives and data understanding services like *Hestia.ai* [See Section IV Introduction]. Collectively through these discoveries, I solidified my existing understandings and was able to contextualise my evolving learning against the established research landscape.

As my understanding from the Case Studies coalesced into a clear, cross-validated understanding of what people want from data and from data holders [Chapter 6], this gave

me the confidence to grow and evolve as a researcher; moving from investigatory or theoretical research towards activist exploration of how to work towards delivering these new capabilities in practice, enabled by the models and ideas I developed. This ultimately gave me the confidence to recognise that, in this body of work, I had identified a newly emergent design philosophy, that deserved to be named, scoped, and explored—**the research agenda for better Human Data Relations**.

I was especially lucky to find external activities, especially with the BBC and Hestia.ai, that fitted so well alongside my research agenda. These activities slotted perfectly into the action research cycle [3.2.2; Figures 3.1 and 3.2] of my thesis, producing a powerful feedback loop where findings from the Case Studies became immediately applicable to real-world design activities, while experiences of the real-life barriers to pursuit of the HDR goals helped to challenge and evolve the theoretical models (such as shared data interaction) and designs emerging from the Case Studies. The project contexts provided a place where emerging qualitative findings and design ideas could be tested and iteratively improved through attempts to operationalise them, producing well-grounded and actionable learnings.

This dual research-and-practice approach has allowed me to push this thesis further than a traditional HCI study would allow, and underpins the two-track research approach of this thesis, where in the secondary project work track I stepped out of the traditional researcher-as-observer stance and stepped into taking an active role as an expert in user-centred design (UCD) [3.2.1], adversarial design [3.2.1] and practical software interface and process design and innovation.

It has been a tremendous privilege to spend over six years examining in great detail the nature of the problems facing our data-centric society, to translate those impacts into tangible needs as an understanding of the problems of the power imbalance over personal data, and to be able to map out the landscape and possibilities for improving the way we relate to data. Through this research, I have discovered rich evidence to quantify and qualify the losses of agency I had suspected prior to beginning the PhD, and done so in a far greater level of detail than existing research had thus far uncovered. The programme has also given me space to experiment with using both GDPR and web-scraping to access data and push boundaries, to really embrace my role as an HDR activist and adversarial designer

[3.2.1; Figure ARI7.1] while also exploring what can be achieved from 'within the system' through the work with Connected Health Cities and BBC R&D. It has allowed me to design and prototype new models and views of data and of information which have transformed the way I look at digital information and how we relate to it, in particular:

- the five types of held data [Table 5.2],
- the two purposes of HDR [Figure 7.1],
- the approaches to effecting change in the HDR landscape [Figures 9.1, 9.2, 9.18, and 9.20], and
- the principles for generative action in pursuing better HDR [10.1].

I hope these models, as well as the other contributions [1.2], can help others to develop their thinking in the same way, to become more HDR-literate and contribute to the crusade for HDR reform that the world so desperately needs.

The collaborative opportunities have been significant. Without this PhD, I would never have had the opportunities to discuss and develop models for personal data interaction and improved ecosystem negotiability with experts at the BBC, Hestia.ai, Sitra and in the wider *MyData* community. Alongside these formal collaborations, I have disseminated ideas through blogs, tweets, workshop papers and lectures, which has helped to refine and clarify ideas but also to stimulate valuable discussions with interested people to gain feedback that helped develop the models and my own learning further.

This opportunity has opened doors that have allowed me to pivot my career towards putting these learnings into action, working on important projects [See Section IV Introduction] to explore how data interaction reforms can be realised in practice, and how we can become not just innovators but social data activists. Thanks in large part to this research, I am now a Senior UX Designer with the BBC, in a team specialising on data user experience. What started for me as a passion project and a journey in research has become a journey into the design profession. As a designer, I aspire to bring the benefits of this research to the general public through the design of services and interfaces that aim to inform, educate and entertain people in their everyday lives. I now know how to begin to have an impact, how to work on building that better HDR future I and my participants have imagined. It is the journey of a lifetime, and also one that is in many ways just beginning. I

hope that my work and this thesis can contribute to a better, more human-centric digital world, and I can't wait to see where this leads.

10.5 Legacy of This Thesis

This thesis offers five key contributions, which summarised in 1.2:

- An Understanding of what People want from Personal Data [1.2.1];
- Methodologies for Participatory Work around Personal Data [1.2.2];
- Best Practices and Design Guidelines for Systems and Processes involving Personal Data [1.2.3];
- Principles for Generative Action towards Better Data Relations [1.2.4]; and
- A Detailed and Actionable Research Agenda and Strategy for Empowerment and Systemic Change [1.2.5].

The constituent parts of these contributions can be easily located using the HDR index, which is located after the Appendices.

Its detailed understanding of individual needs around data interaction and data-centric service relationships [Chapter 6] is backed by participatory action research in both public sector and private sector Case Studies [Chapter 4; Chapter 5], which provides a clear answer to the two primary research questions RQ1 [3.3.1] and RQ2 [3.3.2]: People want visible, understandable and useable[11] data, process transparency, individual oversight capabilities and involvement in decision making.

Based on a solid grounding in existing literature, policy and innovation around Data Access, Personal Information Management, Human Data Interaction and Human-centric Innovation [Chapter 2], these needs have been synthesised into a clearly-defined new agenda for future research and innovation, *Human Data Relations (HDR)* [7.2] which encompasses four clear objectives [7.5] for improving individual agency and societal power imbalances around data:

(i) data awareness & understanding,

(ii) data useability[11],

(iii) data ecosystem awareness & understanding, and

(iv) data ecosystem negotiability.

Sections III-V of the thesis took this body of work much further than a traditional HCI PhD, drawing on the author's experiences with the practical pursuit of better Human Data Relations in four different real-world academic and industrial project settings [see Section IV Introduction]. Through additional insights, designs and implementation strategies [Chapter 9], the thesis offers not just a theoretical frame for this area of research, but clear and actionable Principles [10.1] that could be immediately explored by researchers and innovators - this thesis becomes **an anthology of reference material, designs and strategies for HDR reform**. This practical contribution of the thesis is delivered in four distinct parts:

- first, a **map of the landscape for improving HDR** [Chapter 8], outlining the key **obstacles** that are likely be faced in pursuing HDR objectives, including illegible, immobile, scattered and unmalleable data; a complex ecosystem lacking metadata; exploitations of power by data holders, introspective practices, insufficient machine understanding of human information, and gaps in interoperability, investment and demand;
- second, **four detailed approaches** for making progress in the pursuit of better HDR, illustrated with reference to real-world projects situated in the HDR space: (i) discovery-driven activism (ii) life interface design, (iii) protection of, and progressive action within, the information landscape and (iv) motivational efforts to make better HDR viable, investable and well-understood across society [Chapter 9];
- third, through **a series of Principles** that can aid the pursuit of better HDR, including conceptualisations around **life information** and **ecosystem information**, deep understandings of the ways in which service providers exert power over the data economy and at the seams of their products; practical trajectories for change including entity identification, individual and collective data activism; and methods for acquiring additional metadata, provenance and context so that systems can better understand and represent human information [10.1]; and
- fourth, **an HDR index**, located after the Appendices, making the novel findings, insights, obstacles, approaches and strategies of this thesis easy to locate,

accompanied by a glossary explaining existing terms and nomenclature this thesis makes use of.

Through its Case Studies, this thesis has made additional contributions to the fields of Early Help and GDPR Data Access, as detailed in 1.2, as well as more generally to any research or design context that deals with how individuals or service providers relate to personal data. The work of this thesis has had demonstrable impact, with methodologies and models from this work having been adopted in projects in the European Union and in the UK. Over a dozen publications, workshops and presentations of the work in this thesis have been delivered [1.3], and the research contributed value to real-world industrial projects at BBC R&D in the UK, Hestia.ai in Switzerland and their client Sitra in Finland, through the sharing of the design Principles and the adoption of models of personal data and methods for data interviewing pioneered here.

In answer to the research gap identified in 2.2.5, this thesis has examined the role of personal data in a 'whole life' context. The insights and Principles of this thesis show that personal data should not be an owned commodity or viewed as something static or inert. It is reflective of something real, both in the value it can offer to people's lives and in the role it can play as their proxy in service relationships. Personal data is ultimately an aspect of the human, and as such to separate or extricate it from that human without their involvement or consent is a violation of individual trust, liberty and respect. Personal data is not a commodity. It is part of a fundamental human right to private life, and this privacy is being regularly violated by the treatment of personal data without sufficient consideration to the individual behind the data. The growth of systems which view data as a resource to be mined, a thing to be transacted with, overlook these moral and socially important human values. It is critical that individuals are given greater agency and control over their personal data, and this thesis has shown that current laws and regulations are, so far, inadequate for this purpose. This research has provided a greater understanding of what is lacking, and what might be done, enabling anyone with a desire to do so to move forward in this space.

Through the grounded and detailed references and examples in Section IV, this work moves beyond conducting research to understand human personal data wants, and **sets the scene for an progressive and activist agenda** to take action in service of those wants, with the objective to reconfigure society to one where those human-centric needs are better met. It

constitutes **a call to arms** for future research, innovation and activism in Human Data Relations, combined with a detailed guide to understand the data economy landscape, what needs to change, and an arsenal of design and implementation strategies for how HDR reformers might fulfil their role as a recursive public [7.6]. Armed with these insights, practitioners of HDR reform can **drive us towards a better future** to deliver increased agency for individuals, greater data use capabilities, and a more balanced landscape around the use of personal data by service providers across society—in short, a better world for us all.

Bibliography

'A guide to lawful basis' (2022) *Information Commissioner's Office*. available at: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/a-guide-to-lawful-basis/#what (accessed: 21 May 2023).

Abbattista, F. *et al.* (2007) 'Shaping personal information spaces from collaborative tagging systems', in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. (PART 3), pp. 728–735. doi: 10.1007/978-3-540-74829-8_89.

Abiteboul, S., André, B. and Kaplan, D. (2015) *Managing your digital life with a Personal information management system*. 5. ACM, pp. 32–35. doi: 10.1145/2670528.

'About digipower.academy' (2022). available at: https://digipower.academy/about.

'About the Quantified Self' (no date). available at: https://quantifiedself.com/about/what-is-quantified-self/ (accessed: 22 March 2021).

'About Us' (no date). Rita. available at: https://ritapersonaldata.com/about.html (accessed: 10 August 2022).

Abowd, G. D. *et al.* (1999) 'Towards a better understanding of context and context-awareness', in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, pp. 304–307. doi: 10.1007/3-540-48157-5_29.

Abowd, G. D. (2012) 'What next, ubicomp?: celebrating an intellectual disappearing act', in *Proceedings of the 2012 ACM conference on ubiquitous computing*. New York, New York, USA: ACM Press, pp. 31–40. doi: http://dx.doi.org/10.1145/2370216.2370222.

Abowd, G. D. and Mynatt, E. D. (2000) *Charting Past, Present, and Future Research in Ubiquitous Computing*. 1, pp. 29–58. available at: https://www.cc.gatech.edu/fce/pubs/abowd-mynatt-tochi-millenium.pdf.

'Access my info' (2022). available at: https://accessmyinfo.ca/.

Ackoff, R. L. (1989) ‘From data to wisdom’, *Journal of Applied Systems Analysis*, 16(1), pp. 3–9.

Adams, R. (2017) ‘Michel Foucault: Discourse’. available at: https://criticallegalthinking.com/2017/11/17/michel-foucault-discourse/ (accessed: 7 May 2021).

Alexander, J., Conrad, A. and Eno, B. (2022) *Citizens: Why the key to fixing everything is all of us*. Canbury Press.

Alizadeh, F. *et al.* (2019) ‘GDPR-reality check on the right to access data’, in *ACM international conference proceeding series*. New York, New York, USA: ACM Press, pp. 811–814. doi: 10.1145/3340764.3344913.

Allen, D. (2015) *Getting things done: The art of stress-free productivity*. Penguin.

‘AllofMe Company Profile’ (2007). available at: https://www.crunchbase.com/organization/allofme (accessed: 23 March 2021).

‘AllOfMe.com Teaser Clip’ (2008). YouTube. available at: https://www.youtube.com/watch?v=JWyqt4WL6xE (accessed: 21 March 2021).

Altsitsiadis, E. (2021) ‘The maker movement – the quiet, game-changing revolution near you’. available at: http://www.bos-cbscsr.dk/2021/01/05/the-maker-movement/ (accessed: 21 August 2022).

Ananya (2020) ‘Java: Pass by value or pass by reference’, *Medium*. available at: https://medium.com/swlh/java-passing-by-value-or-passing-by-reference-c75e312069ed.

Andrews, R. (2005) ‘GTD : A New Cult for the Info Age’, *Wired*. available at: https://www.wired.com/2005/07/gtd-a-new-cult-for-the-info-age/.

Apple (2009) ‘iPhone 3G Commercial: "There’s an app for that"’. YouTube. available at: https://www.youtube.com/watch?v=mFlITzqRBWY.

Apple (2022) ‘About app activity report’. available at: https://support.apple.com/en-us/HT212958.

Arfelt, E., Basin, D. and Debois, S. (2019) 'Monitoring the GDPR', in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, pp. 681–699. doi: 10.1007/978-3-030-29959-0_33.

Arthur, C. (2017) 'Theresa May wants to tackle online extremism. Here's how to do it', *The Guardian*. available at: https://www.theguardian.com/commentisfree/2017/may/26/theresa-may-online-extremisim-google-facebook-twitter-manchester.

'Article 13: Information to be provided where personal data are collected from the data subject' (2020). doi: 10.1093/oso/9780198826491.003.0044.

Article 29 Data Protection Working Party (2010) *Opinion 3/2010 on the principle of accountability*. Brussels: European Commission.

Aslam, H. *et al.* (2016) 'Harnessing Smartphones as a Personal Informatics Tool towards Self-Awareness and Behavior Improvement', *Proceedings - 2016 IEEE 14th International Conference on Dependable, Autonomic and Secure Computing, DASC 2016, 2016 IEEE 14th International Conference on Pervasive Intelligence and Computing, PICom 2016, 2016 IEEE 2nd International Conference on Big Data*. IEEE, pp. 467–474. doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2016.92.

Attoresi, M. and Moraes, T. (2020) 'Personal information management systems', *EDPS TechDispatch, European Union*. European Data Protection Supervisor, European Union, (3).

Ausloos, J. (2019) 'GDPR Transparency as a Research Method', *SSRN Electronic Journal*, (May), pp. 1–23. doi: 10.2139/ssrn.3465680.

Ausloos, J. and Dewitte, P. (2018) *Shattering one-way mirrors-data subject access rights in practice*. available at: www.irissproject.eu https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3106632.

Ausloos, J. and Veale, M. (2020) 'Researching with Data Rights', *Technology and Regulation*, pp. 136–157.

Various Authors (2022) 'Naive bayes classifiers', *GeeksForGeeks*. available at: https://www.geeksforgeeks.org/naive-bayes-classifiers/.

Bakardjieva, M. and Feenberg, A. (2001) 'Involving the Virtual Subjects', *Ethics and Information Technology*, 2, pp. 233–240. doi: 10.1023/A:1011454606534.

Baker, K. S. and Karasti, H. (2018) 'Data care and its politics: Designing for local collective data management as a neglected thing', in *Proceedings of the 15th participatory design conference: Full papers-volume 1*, pp. 1–12.

Bakshy, E., Messing, S. and Adamic, L. A. (2015) 'Exposure to ideologically diverse news and opinion on facebook', *Science*. American Association for the Advancement of Science, 348, pp. 1130–1132. doi: 10.1126/SCIENCE.AAA1160/SUPPL_FILE/PAPV2.PDF.

Bannon, L. J. (1995) 'From Human Factors to Human Actors: The Role of Psychology and Human-Computer Interaction Studies in System Design', *Readings in Human–Computer Interaction*, pp. 205–214. doi: 10.1016/b978-0-08-051574-8.50024-8.

Bansal, A. (2018) 'An introduction to SOLID, tim berners-lee's new, re-decentralized web', *FreeCodeCamp*. available at: https://www.freecodecamp.org/news/an-introduction-to-solid-tim-berners-lees-new-re-decentralized-web-25d6b78c523b/.

Barbosa Neves, B. and Casimiro, C. (2018) *Connecting Families?: Information & Communication Technologies, Generations, and the Life Course*. Policy Press.

Barclay, L. (2021) 'Facebook banned me for life because i help people use it less', *Slate*.

Barreau, D. K. (1995) 'Context as a factor in personal information management systems', *Journal of the American Society for Information Science*, 46(5), pp. 327–339. doi: 10.1002/(SICI)1097-4571(199506)46:5<327::AID-ASI4>3.0.CO;2-C.

Barreau, D. and Nardi, B. A. (1995) 'Finding and reminding', *ACM SIGCHI Bulletin*, 27(3), pp. 39–43. doi: 10.1145/221296.221307.

Bate, A. and Bellis, A. (2018) *The Troubled Families programme (England)*. July.

Battarbee, K. and Koskinen, I. (2005) 'Co-experience: user experience as interaction', *CoDesign*. Taylor & Francis, 1(1), pp. 5–18.

Baxter, G. and Sommerville, I. (2011) 'Socio-technical systems: From design methods to systems engineering', *Interacting with Computers*. OUP, 23(1), pp. 4–17. doi: 10.1016/j.intcom.2010.07.003.

BBC Cornmarket Team (2021) 'Building trusted data services and capabilities'. available at: https://paper.dropbox.com/doc/Building-trusted-data-services-and-capabilities--BmF6FlBd0cKGw0nC654th7V3Ag-Us49Ek0nex7yClKughPN4.

BBC News (2014) 'More than 400,000 freedom of information requests made since 2005', *BBC News*. available at: https://www.bbc.co.uk/news/uk-politics-30632221.

BBC R&D (2017) 'Human Data Interaction - BBC R&D'. available at: https://www.bbc.co.uk/rd/projects/human-data-interaction.

Beck, K. *et al.* (2001) 'The Agile Manifesto'. available at: http://agilemanifesto.org/.

Beghelli, A., Huerta-Cánepa, G. and Segal, R. (2019) 'Data materialisation: A new undergraduate course for a data driven society', in *Proceedings of the design society: International conference on engineering design*. Cambridge University Press (1), pp. 2061–2070.

Bell, G. and Gemmell, J. (2009) *Total recall: how the E-memory revolution will change everything*. Dutton (09), pp. 47-5062-47-5062. doi: 10.5860/choice.47-5062.

Bennett, C. J. and Raab, C. D. (2017) *The governance of privacy: Policy instruments in global perspective*. Routledge.

Bentham, J. and Bozovic, M. (1791) *The panopticon writings (1791/2011)*. Verso Books (Radical thinkers). available at: https://books.google.co.uk/books?id=VbpvDwAAQBAJ.

Bergman, O. *et al.* (2008) 'Improved search engines and navigation preference in personal information management', *ACM Transactions on Information Systems*, 26(4). doi: 10.1145/1402256.1402259.

Bergman, O. *et al.* (2012) 'How do we find personal files?: The effect of OS, presentation & depth on file navigation', *Conference on Human Factors in Computing Systems - Proceedings*, pp. 2977–2980. doi: 10.1145/2207676.2208707.

Bergman, O. (2013) 'The Effect of Folder Structure on Personal File Navigation', *Journal of the American Society for Information Science and Technology*, 64(July), pp. 1852–1863. doi: 10.1002/asi.

Bergman, O., Beyth-Marom, R. and Nachmias, R. (2003) 'The user-subjective approach to personal information management systems', *Journal of the American Society for Information Science and Technology*, 54(9), pp. 872–878. doi: 10.1002/asi.10283.

Berners-Lee, T. (2022) 'Solid: Sir tim berners-lee's vision of a vibrant web for all'. Inrupt. available at: https://inrupt.com/solid/.

Berners-Lee, T., Hendler, J. and Lassila, O. (2001) 'The Semantic Web', *Scientific American*, 284(5), pp. 34–43. available at: https://jstor.org/stable/10.2307/26059207.

Binns, R. (2022) 'Tracking on the Web, Mobile and the Internet-of-Things'. available at: http://arxiv.org/abs/2201.10831.

'Bits of freedom: Our focus' (2000). available at: https://www.bitsoffreedom.nl/english/.

Bjerknes, Gro. *et al.* (1987) *Computers and democracy : a Scandinavian challenge*. Aldershot [Hants, England]; Brookfield [Vt.], USA: Avebury, p. 434. available at: http://www.worldcat.org/title/computers-and-democracy-a-scandinavian-challenge/oclc/614994092?referer=di&ht=edition.

Björgvinsson, E., Ehn, P. and Hillgren, P.-A. (2010) 'Participatory design and" democratizing innovation"', in *Proceedings of the 11th biennial participatory design conference*, pp. 41–50.

Bødker, S. (2006) 'When second wave HCI meets third wave challenges', *ACM International Conference Proceeding Series*, 189(October), pp. 1–8. doi: 10.1145/1182475.1182476.

Bødker, S. (2015) 'Third-wave HCI, 10 years later—participation and sharing', *Interactions*, 22(5), pp. 24–31. doi: 10.1145/2804405.

Bogers, S. *et al.* (2016) 'Connected baby bottle', pp. 301–311. doi: 10.1145/2901790.2901855.

Boud, D., Keogh, R. and Walker, D. (1985) *Reflection: Turning experience into learning*. Routledge.

Bowcott, O. and Hern, A. (2018) 'Facebook and cambridge analytica face class action lawsuit', *The Guardian*. available at: https://www.theguardian.com/news/2018/apr/10/cambridge-analytica-and-facebook-face-class-action-lawsuit.

Bowker, G. C. (2005) *Memory practices in the sciences*. MIT Press, p. 261.

Bowker, G. C. *et al.* (2015) *Boundary objects and beyond : working with Leigh Star*. MIT Press, p. 548. available at: https://books.google.co.uk/books?hl=en&lr=&id=nmSkCwAAQBAJ&oi=fnd&pg=PR5&dq=Boundary+Objects+and+Beyond:+Working+with+Leigh+Star&ots=blmnW7yz4u&sig=F08uGeG_lT_klhhR64M18tQNI1s#v=onepage&q=Boundary Objects and Beyond%3A Working with Leigh Star&f=false.

Bowyer, A. (2009) 'The importance of persona: Why your computer needs to know which hat you're wearing'. available at: http://bitnorth.com/shortbits/.

Bowyer, A. (2010) 'Why files need to die'. available at: http://bitnorth.com/shortbits/.

Bowyer, A. (2011) 'Why files need to die'. available at: http://radar.oreilly.com/2011/07/why-files-need-to-die.html.

Bowyer, A. (2012a) 'How internet economics threaten useful computing'. available at: http://bitnorth.com/shortbits/.

Bowyer, A. (2012b) 'We deserve the time and space to be human', in Gupta, V. (ed.) *The Future We Deserve*. PediaPress GmbH. available at: https://eprints.ncl.ac.uk/273829.

Bowyer, A. (2013) 'The interfaces no-one is building: What would a human-centric computer look like'. available at: http://bitnorth.com/shortbits/.

Bowyer, A. (2016) 'An exploration of usefulness in smartphone applications'. available at: https://eprints.ncl.ac.uk/273836.

Bowyer, A. (2017) 'Designing for human autonomy: The next challenge that civic HCI must address'. available at: https://eprints.ncl.ac.uk/273832.

Bowyer, A. (2018a) ‘A grand vision for post-capitalist HCI: Digital life assistants"’, *CHI Workshops 2018*. available at: https://eprints.ncl.ac.uk/273826.

Bowyer, A. (2018b) ‘Free Data Interfaces: Taking Human- Data Interaction to the Next Level’, *CHI Workshops 2018*. available at: https://eprints.ncl.ac.uk/273825.

Bowyer, A. *et al.* (2018) ‘Understanding the Family Perspective on the Storage, Sharing and Handling of Family Civic Data’, in *Conference on human factors in computing systems - proceedings*. New York, New York, USA: ACM Press, pp. 1–13. doi: 10.1145/3173574.3173710.

Bowyer, A. *et al.* (2019) ‘Human-data interaction in the context of care: Co-designing family civic data interfaces and practices’, in *Conference on human factors in computing systems - proceedings*. doi: 10.1145/3290607.3312998.

Bowyer, A. (2020a) ‘Design research for cornmarket PDS, recommender & associated permissions: Report by alex bowyer (BBC research intern/open lab PhD)’. available at: https://bit.ly/bbc-pds-research-bowyer.

Bowyer, A. (2020b) ‘Designing personal data interfaces - a multi-disciplinary challenge’. available at: https://bit.ly/bbc-internship-alex-bowyer (accessed: 18 August 2022).

Bowyer, A. (2020c) ‘Personal data use: A human-centric perspective’. available at: https://eprints.ncl.ac.uk/273834.

Bowyer, A. (2021a) ‘Human-Data Interaction has two purposes: Personal Data Control and Life Information Exploration’. available at: https://eprints.ncl.ac.uk/273832#.

Bowyer, A. (2021b) ‘My thesis in 3 minutes: Understanding and designing human data relations’. available at: https://www.youtube.com/watch?v=YFHXc_TfM5c.

Bowyer, A., Pidoux, J., *et al.* (2022) *Digipower technical reports: Auditing the data economy through personal data access*. doi: 10.5281/zenodo.6554177.

Bowyer, A., Holt, J., *et al.* (2022) ‘Human-GDPR interaction : Practical experiences of accessing personal data’, *CHI ’22*.

Bowyer, A. (2022a) ‘Response to guidelines 01/2022 on data subject rights - right of access version 1.0’. available at: https://edpb.europa.eu/sites/default/files/webform/public_consultation_reply/Response%20to%20Guidelinesv1.0-AlexBowyer-March2022.pdf.

Bowyer, A. (2022b) ‘The power of personal data part 1: How companies exploit your data to shape your world’. available at: https://bit.ly/power-personal-data-part-1.

Bowyer, A. (2023) ‘The power of personal data part 2: Taking back control of our data and challenging powerful corporations’. available at: https://bit.ly/power-personal-data-part-2.

Bowyer, A. and Bourgeois, J. (2023) ‘Conversation with alex bowyer: Understanding and improving human data relations’. Data-centric Design Lab, Delft University of Technology. available at: https://podcasts.apple.com/gb/podcast/conversation-with-alex-bowyer-understanding-and/id1581554015?i=1000595165881.

Bowyer, A. and Croll, A. (2009-2011) ‘Human 2.0’. available at: https://web.archive.org/web/20111231165329/http://www.human20.com/.

Bowyer, A. and Wheater, S. (2017) ‘SILVER family health data interface: Demo screencasts / showcase’. available at: https://eprints.ncl.ac.uk/273839.

Bradbury, D. (2019) ‘You can probably be identified from your anonymized data’, *Naked Security by Sophos*. available at: https://nakedsecurity.sophos.com/2019/07/25/youre-special-after-alland-thats-bad-for-your-privacy/.

Brakel, M., Uuk, R. and Boine, C. (2022) ‘The artificial intelligence act’. Future of Life Institute. available at: https://artificialintelligenceact.eu/.

Brandt, E. and Messeter, J. (2004) ‘Facilitating collaboration through design games’, in *Proceedings of the eighth conference on participatory design artful integration: Interweaving media, materials and practices - PDC 04*. New York, New York, USA: ACM Press, p. 121. doi: 10.1145/1011870.1011885.

Braun, A. (2018) ‘Which email providers are scanning your emails?’, *maketecheasier*. available at: https://www.maketecheasier.com/which-email-providers-scanning-emails/ (accessed: 21 August 2022).

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*. Taylor & Francis, 3(2), pp. 77–101. doi: 10.1191/1478088706qp063oa.

Braw, E. (2022) 'How corporate boycotts could backfire', *Foreign Policy*. available at: https://foreignpolicy.com/2022/03/28/russia-sanctions-ukraine-corporate-boycotts-could-backfire/.

Brest, P. (2010) 'The Power of Theories of Change', *Stanford Social Innovation Review*, 8(2), pp. 47–51.

Bridle, J. (2016) 'Algorithmic Citizenship, Digital Statelessness', *GeoHumanities*. James Bridle, 2(2), pp. 377–381. doi: 10.1080/2373566x.2016.1237858.

British Broadcasting Corporation (1997) 'Our purpose'. available at: https://www.bbc.co.uk/rd/about/our-purpose (accessed: 18 August 2022).

Brooks, D. (2013) 'The Philosophy of Data'. available at: https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html.

Brown, D. (2015) 'Here's what 'fail fast' really means'. available at: https://venturebeat.com/2015/03/15/heres-what-fail-fast-really-means/.

Brynjolfsson, E. and Oh, J. H. (2012) 'The attention economy: Measuring the value of free digital services on the internet', *International Conference on Information Systems, ICIS 2012*, 4, pp. 3243–3261.

Bufalieri, L. *et al.* (2020) 'GDPR: When the right to access personal data becomes a threat'. doi: 10.1109/icws49710.2020.00017.

Bunge, M. (1999) *Social Science Under Debate: A Philosophical Perspective*. University of Toronto Press. available at: https://books.google.co.uk/books?id=-MLjZzJLbpkC.

Burgess, M. (2021) 'Why Amazon's £636m GDPR fine really matters', *Wired*. available at: https://www.wired.co.uk/article/amazon-gdpr-fine.

Burkeman, O. (2011) 'SXSW 2011: The internet is over'. available at: https://www.theguardian.com/technology/2011/mar/15/sxsw-2011-internet-online (accessed: 23 March 2021).

Bush, V. (1945) ‘As we may think’, *The Atlantic Monthly*, 3(2), pp. 35–46. doi: 10.1145/227181.227186.

Cambridge Dictionary (no date) ‘Empowerment’. available at: https://dictionary.cambridge.org/dictionary/english/empowerment.

Campbell, P. L. (2011) ‘Peirce, pragmatism, and the right way of thinking’, *Sandia National Laboratories, Albuquerque*. Citeseer.

Carter, J. (2015) ‘Who are the digital disruptors redefining entire industries?’ available at: https://www.techradar.com/uk/news/world-of-tech/who-are-the-digital-disruptors-redefining-entire-industries-1298171 (accessed: 23 March 2021).

Caruthers, M. (2018) ‘World Password Day: How to Improve Your Passwords’. available at: https://blog.dashlane.com/world-password-day/ (accessed: 5 May 2021).

Cavoukian, A. (2010) ‘Privacy by design: the definitive workshop. A foreword by Ann Cavoukian, Ph.D’, *Identity in the Information Society*, 3(2), pp. 247–251. doi: 10.1007/s12394-010-0062-y.

Cavoukian, A. (2012) ‘Privacy by Design and the Emerging Personal Data Ecosystem’, (October), pp. 1–39.

Chalmers, M., MacColl, I. and Bell, M. (2003) ‘Seamful design: Showing the seams in wearable computing’, *IEE Eurowearable*.

Chan, R. (2019) ‘The cambridge analytica whistleblower explains how the firm used facebook data to sway elections’, *Business Insider*. available at: https://www.businessinsider.com/cambridge-analytica-whistleblower-christopher-wylie-facebook-data-2019-10?r=US&IR=T.

Chang, A. (2018) ‘The Facebook and Cambridge Analytica scandal, explained with a simple diagram - Vox’. available at: https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram.

Cheetham, M. *et al.* (2018) 'Embedded research: A promising way to create evidence-informed impact in public health?', *Journal of Public Health (United Kingdom)*. Oxford University Press, 40(suppl_1), pp. i64–i70. doi: 10.1093/pubmed/fdx125.

Chevalier, J. M. and Buckles, D. J. (2008) *SAS2: A guide to collaborative inquiry and social engagement*. SAGE Publishing India.

Choe, E. K. *et al.* (2014) 'Understanding quantified-selfers' practices in collecting and exploring personal data', in *Proceedings of the 32nd annual ACM conference on human factors in computing systems - CHI '14*. New York, New York, USA: ACM Press, pp. 1143–1152. doi: 10.1145/2556288.2557372.

Chung, C. F. *et al.* (2016) 'Boundary negotiating artifacts in personal informatics: Patient-provider collaboration with patient-generated data', *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 27, pp. 770–786. doi: 10.1145/2818048.2819926.

CitizenMe (2021) 'Become a Citizen and unlock the value of your data'. available at: https://www.citizenme.com/for-citizens/ (accessed: 23 August 2021).

Claburn, T. (2021) 'Android's Messages, Dialer apps quietly sent text, call info to Google'. available at: https://www.theregister.com/2022/03/21/google_messages_gdpr/.

Clarke, N. *et al.* (2019) 'GDPR: an impediment to research?', *Irish Journal of Medical Science (1971-)*. Springer, 188(4), pp. 1129–1135.

Cockton, G. (2004) 'Value-centred HCI', *NordiCHI*.

Cogran, P. and Kinsley, S. (2012) 'Paying Attention: towards a critique of the attention economy', *Cultural Machine*, 13.

Colbert, M. (2022) 'Labour members left in the dark over data breach', *Byline Times*. available at: https://bylinetimes.com/2022/06/06/labour-members-left-in-dark-over-data-breach/.

Colcelli, V. (2019) 'Joint controller agreement under gdpr', *EU and comparative law issues and challenges series (ECLIC)*, 3, pp. 1030–1047.

Collins Dictionary (no date) 'bleeding edge'. available at: https://www.collinsdictionary.com/dictionary/english/bleeding-edge (accessed: 29 July 2022).

Collins English Dictionary (no date a) 'Useability'. available at: https://www.thefreedictionary.com/useability.

Collins English Dictionary (no date b) 'Useable'. available at: https://www.thefreedictionary.com/useable.

Comandè, G. and Schneider, G. (2021) 'Can the GDPR make data flow for research easier? Yes it can, by differentiating! A careful reading of the GDPR shows how EU data protection law leaves open some significant flexibilities for data protection-sound research activities', *Computer Law & Security Review*. Elsevier, 41, p. 105539.

Connected Health Cities (2017) 'SILVER Project: Smart Interventions for Local Residents'. available at: https://web.archive.org/web/20210308040602/https://www.connectedhealthcities.org/research-projects/troubled-families/ (accessed: 8 March 2021).

Connected Health Cities (2021) *Connected health cities impact report: 2016 - 2020*. available at: https://web.archive.org/web/20211225192408/https://www.chc-impact-report.co.uk/ (accessed: 25 December 2021).

Copeland, E. (2015) *Small Pieces Loosely Joined: How smarter use of technology and data can deliver real reform of local government*. available at: www.policyexchange.org.uk https://policyexchange.org.uk/publication/small-pieces-loosely-joined-how-smarter-use-of-technology-and-data-can-deliver-real-reform-of-local-government/.

Cormack, A. (2016) 'Is the Subject Access Right Now Too Great a Threat to Privacy?', *European Data Protection Law Review*, 2(1), pp. 15–27. doi: 10.21552/edpl/2016/1/5.

Cormack, A. (2021) 'Thinking with GDPR: A guide to better system design', *Information Services & Use*, 41(1-2), pp. 61–69. doi: 10.3233/isu-210107.

Cornford, J., Baines, S. and Wilson, R. (2013) 'Representing the family: how does the state 'think family'?', *Policy & Politics*, 41(1), pp. 1–19. doi: 10.1332/030557312X645838.

Corra, M. and Willer, D. (2002) 'The gatekeeper', *Sociological Theory*. SAGE Publications Sage CA: Los Angeles, CA, 20(2), pp. 180–207.

Costanza-Chock, S. (2020) 'Design practices: "Nothing about us without us"', *Design Justice*. MIT Press. available at: https://designjustice.mitpress.mit.edu/pub/cfohnud7/release/4 (accessed: 24 May 2023).

Coughlan, T., Leder Mackley, K., *et al.* (2013) 'Current issues and future directions in methods for studying technology in the home', *PsychNology Journal*, 11(2), pp. 159–184.

Coughlan, T., Brown, M., *et al.* (2013) 'Methods for studying technology in the home', in *CHI'13 extended abstracts on human factors in computing systems*, pp. 3207–3210.

Coulton, P. and Lindley, J. G. (2019) 'More-than human centred design: Considering other things', *The Design Journal*. Taylor & Francis, 22(4), pp. 463–481.

Council of the European Union (2015) 'Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)'. Brussels. available at: http://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf.

Cowburn, P. (2023) '26 civil society groups call on government to scrap data protection and digital information (DPDI) bill', *Open Rights Group Press Releases*. available at: https://www.openrightsgroup.org/press-releases/26-civil-society-groups-call-on-government-to-scrap-data-protection-and-digital-information-dpdi-bill/.

Crabtree, A. and Mortier, R. (2015) 'Human data interaction: Historical lessons from social studies and CSCW', in *ECSCW 2015: Proceedings of the 14th european conference on computer supported cooperative work, 19-23 september 2015, oslo, norway*. Springer, pp. 3–21.

Crabtree, A. and Mortier, R. (2016) 'Personal Data, Privacy and the Internet of Things: The Shifting Locus of Agency and Control', *SSRN Electronic Journal*, pp. 1–20. doi: 10.2139/ssrn.2874312.

Crabtree, A. and Tolmie, P. (2018) 'The practical politics of sharing personal data', in *Personal and Ubiquitous Computing*. Springer-Verlag (2), pp. 293–315. doi: 10.1007/s00779-017-1071-8.

Crabtree, J. (2007) 'Civic hacking: A new agenda for e-democracy', *openDemocracy*. available at: https://www.opendemocracy.net/en/civic_hacking_a_new_agenda_for_e_democracy/.

Crivellaro, C. *et al.* (2019) 'Not-equal: Democratizing research in digital innovation for social justice', *Interactions*, 26(2), pp. 70–73. doi: 10.1145/3301655.

Croll, A. (2009) 'The Three Economies of Online Currency'. available at: https://solveforinteresting.com/the-three-currencies-of-the-online-economy/.

Croll, A. and Bowyer, A. (2009) 'Why email clients need to change'. available at: https://web.archive.org/web/20120128050842/http://gigaom.com/2009/04/24/why-email-clients-need-to-change/.

Crossley, S. (2022) 'Misrecognising 'complex'families: A social harm perspective', in *Negotiating families and personal lives in the 21st century*. Routledge, pp. 30–45.

Crossley, S. and Lambert, M. (2017) 'Introduction: "Looking for trouble?" Critically examining the UK government's troubled families programme', *Social Policy and Society*. Cambridge University Press, 16(1), pp. 81–85. doi: 10.1017/S1474746416000476.

Ctrl-Shift (2014) 'Personal Information Management Services: An analysis of an emerging market'. Nesta, p. 38. available at: https://www.nesta.org.uk/report/personal-information-management-services-an-analysis-of-an-emerging-market/.

D. Urquhart, L. and J. Craigon, P. (2021) 'The moral-IT deck: A tool for ethics by design', *Journal of Responsible Innovation*. Taylor & Francis, 8(1), pp. 94–126.

Daniel Turner (2014) 'Quirkos cloud'. available at: https://www.quirkos.com/learn-qualitative/features.html.

'Data' (no date). Grammarist. available at: https://grammarist.com/usage/data/.

'Data Brokers: Everything You Need to Know' (2022). available at: https://www.avast.com/c-data-brokers (accessed: 11 February 2022).

‘datacy - About Us’ (no date). available at: https://www.datacy.com/personal/about-us (accessed: 22 March 2019).

Davies, S. (1990) ‘Privacy international: About us’. available at: https://privacyinternational.org/about.

Decker, S. and Frank, M. (2004) ‘The Networked Semantic Desktop’, *WWW Workshop on Application Design, Development and Implementation Issues in the Semantic Web*. doi: 10.1108/eb057368.

Deckers, E. (2018) ‘Data-enabled design’, *UXDX*. available at: https://uxdx.com/blog/data-enabled-design/.

Dehaye, P.-O. (2017) ‘Facebook forced to disclose more information about its ad targeting’, *Medium*. available at: https://medium.com/personaldata-io/facebook-forced-to-disclose-more-information-about-its-ad-targeting-7e6c0127722.

Dehaye, P.-O. (2018) ‘Post-hearing questions by Senator Blumenthal to Mark Zuckerberg’. available at: https://wiki.personaldata.io/wiki/Item:Q1800.

Dehaye, P.-O. (2019) ‘Hestia.ai: About us’. available at: https://hestia.ai/en/about/.

Dehaye, P.-O. (2021) ‘HestiaLabs’. Geneva, Switzerland. available at: https://hestialabs.org/en/ (accessed: 23 August 2021).

Dehaye, P.-O. (2022) ‘Hestia.ai: achievements’. available at: https://hestia.ai/en/#realization.

Delacroix, S. and Lawrence, N. D. (2019) ‘Bottom-up data trusts: Disturbing the ‘one size fits all’approach to data governance’, *International data privacy law*. Oxford University Press, 9(4), pp. 236–252.

Delacroix, S. and Montgomery, J. (2020) ‘From research data ethics principles to practice: Data trusts as a governance tool’, *Available at SSRN 3736090*.

‘Delicious’ (2003). available at: https://en.wikipedia.org/wiki/Delicious_(website).

Department for Education (2018) *Working Together to Safeguard Children*. March, p. 393. doi: 10.1080/13561820020003919.

Department of Health, Home Office and Department for Education and Skills (2006) *Working together to safeguard children: A guide to inter-agency working to safeguard and promote the welfare of children*. Department for Education; Skills, p. 256.

Design Council UK (2004) 'What is the framework for innovation? Design Council's evolved Double Diamond'. available at: https://www.designcouncil.org.uk/news-opinion/what-framework-innovation-design-councils-evolved-double-diamond (accessed: 20 May 2021).

DeSimone, C. (2010) 'Pitting karlsruhe against luxembourg? German data protection and the contested implementation of the EU data retention directive', *German Law Journal*. Cambridge University Press, 11(3), pp. 291–317.

Dewey, J. (1938) 'Experience and education'.

Dewey, J. and Archambault, R. D. (1964) 'John Dewey on education: Selected writings'.

Dey, A. K. (2000) *Providing Architectural Support for Building Context-Aware Applications*. PhD thesis.

Dey, A. K. (2001) 'Understanding and using context', *Personal and ubiquitous computing*, pp. 4–7. available at: http://dl.acm.org/citation.cfm?id=593572.

digi.me (2019) 'Digi.me private sharing: See how you can do more with your personal data'. YouTube video; YouTube. available at: https://www.youtube.com/watch?v=pGcnK_KraXs.

'Digi.me' (no date). available at: https://digi.me/ (accessed: 23 August 2021).

Dijck, J. van (2014) 'Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology', *Surveillance and Society*. Surveillance Studies Network, 12(2), pp. 197–208. doi: 10.24908/ss.v12i2.4776.

DiSalvo, C. (2010) 'Design, Democracy and Agonistic Pluralism', *Proceedings of the Design Research Society Conference 2010*, pp. 366–371.

DiSalvo, C. (2012) *Adversarial Design*. MIT Press (Design thinking, design theory). doi: 10.7551/mitpress/8732.003.0007.

Dix, A. (2007) 'Designing for appropriation', in *Proceedings of HCI 2007 the 21st british HCI group annual conference university of lancaster, UK 21*, pp. 1–4.

Doctorow, C. (2019) 'Adversarial interoperability', *Electronic Frontier Foundation*. available at: https://www.eff.org/deeplinks/2019/10/adversarial-interoperability.

Dourish, P. *et al.* (2000) 'Extending document management systems with user-specific active properties', *ACM Transactions on Information Systems*, 18(2), pp. 140–170. doi: 10.1145/348751.348758.

Dourish, P. (2001) *Where the action is: the foundations of embodied interaction*. MIT press.

Dourish, P. (2003) 'The appropriation of interactive technologies: Some Lessons From Placeless Documents', *Computer Supported Cooperative Work*, 12(4), pp. 465–490.

Dourish, P. (2004) 'What we talk about when we talk about context', *Personal and Ubiquitous Computing*, 8(1), pp. 19–30. doi: 10.1007/s00779-003-0253-8.

Dudovskiy, J. (2012) 'Interpretivism (interpretivist) research philosophy', *Business Research Methodology*. available at: https://research-methodology.net/research-philosophy/interpretivism/.

Edwards, A. and Elwyn, G. (2006) 'Inside the black box of shared decision making: distinguishing between the process of involvement and who makes the decision', *Health Expectations*. Wiley/Blackwell (10.1111), 9(4), pp. 307–320. doi: 10.1111/j.1369-7625.2006.00401.x.

Edwards, J. (2004) 'Manifesto of the programmer liberation front', *Alarming Development*. available at: https://web.archive.org/web/20051219171732/http://alarmingdevelopment.org/?p=5 (accessed: 7 May 2023).

Edwards, L. *et al.* (2019) 'Data subjects as data controllers: A fashion (able) concept', *Internet Policy Review*, 13.

Edwards, L. and Veale, M. (2017) 'Slave to the algorithm? Why a'right to an explanation'is probably not the remedy you are looking for', *Duke L. & Tech. Rev.* HeinOnline, 16, p. 18.

Edwards, L. and Veale, M. (2018) 'Enslaving the algorithm: From a "right to an explanation" to a "right to better decisions"?', *IEEE Security & Privacy*. IEEE, 16(3), pp. 46–54.

Eliasson, J., Cerratto Pargman, T. and Ramberg, R. (2009) 'Embodied interaction or context-aware computing? An integrated approach to design', in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. Springer, Berlin, Heidelberg (PART 1), pp. 606–615. doi: 10.1007/978-3-642-02574-7_68.

Elsden, C. *et al.* (2016) 'Metadating: Exploring the romance and future of personal data', in *Proceedings of the 2016 chi conference on human factors in computing systems*, pp. 685–698.

Engelbart, D. C. (1962) 'Augmenting human intellect: A conceptual framework'. Menlo Park, CA, USA: Stanford Research Institute.

Erdos, D. (2015) 'Data protection confronts freedom of expression on the" new media" internet: The stance of european regulatory authorities', *European law review*. Sweet & Maxwell, (4), pp. 531–562.

Erdos, D. (2016) 'European union data protection law and media expression: Fundamentally off balance', *International & Comparative Law Quarterly*. Cambridge University Press, 65(1), pp. 139–183.

Erdos, D. (2019) 'European data protection regulation, journalism, and traditional publishers: Balancing on a tightrope?'

Erdos, D. (2022) 'Identification in personal data: Authenticating the meaning and reach of another broad concept in EU data protection law', *Computer Law & Security Review*. Elsevier, 46, p. 105721.

Es, M. van, Guijt, I. and Vogel, I. (2015) 'Hivos ToC Guidelines: Theory of Change Thinking in Practice'. The Hague, The Netherlands: Hivos.

'Ethi' (no date). available at: https://www.ethi.me/.

Etzel, B. (1995) 'New strategy and techniques to cope with information overload', in *IEE colloquium on information overload*. IEE (223), pp. 2–2. doi: 10.1049/ic:19951427.

European Commission (2014) *Research and Innovation in the field of ICT for Health, Wellbeing and Ageing Well: an overview*, p. 39.

European Commission (2022a) 'Data governance act explained | shaping europe's digital future'. available at: https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained.

European Commission (2022b) 'The digital markets act: Ensuring fair and open digital markets'. available at: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en.

European Commission (2022c) 'The digital services act: Ensuring a safe and accountable online environment'. available at: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en.

European Data Protection Board (2022a) 'Guidelines 01/2022 on data subject rights - right of access version 1.0 - public consultation'. available at: https://edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-012022-data-subject-rights-right_en.

European Data Protection Board (2022b) 'The EU's data act: Data protection must prevail to empower data subjects'. available at: https://edpb.europa.eu/news/news/2022/eus-data-act-data-protection-must-prevail-empower-data-subjects_en.

European Union Agency for Fundamental Rights (2020) 'Your Rights Matter: Data Protection and Privacy 2020', p. 20. doi: 10.2811/031862.

Eurostat (2020) 'First population estimates: EU population in 2020.' available at: https://ec.europa.eu/eurostat/documents/2995521/11081093/3-10072020-AP-EN.pdf/d2f799bf-4412-05cc-a357-7b49b93615f1 (accessed: 23 August 2021).

Evans, W. (2021) 'Amazon's dark secret: It has failed to protect your data'.

‘Exist.io’ (no date). available at: https://exist.io/ (accessed: 23 August 2021).

‘Explainable AI: Making machines understandable for humans’ (no date). available at: https://explainableai.com/ (accessed: 16 June 2022).

‘Facebook - Data Policy’ (no date). available at: https://www.facebook.com/about/privacy (accessed: 9 August 2021).

‘Facebook–Cambridge Analytica Data Scandal’ (2014). available at: https://en.wikipedia.org/wiki/Facebook–Cambridge_Analytica_data_scandal.

Faife, C. (2021) ‘Facebook rolls out news feed change that blocks watchdogs from gathering data’, *The Markup*. available at: https://themarkup.org/citizen-browser/2021/09/21/facebook-rolls-out-news-feed-change-that-blocks-watchdogs-from-gathering-data.

Feng, Y. and Agosto, D. E. (2019) ‘Revisiting personal information management through information practices with activity tracking technology’, *Journal of the Association for Information Science and Technology*, 70(12), pp. 1352–1367. doi: 10.1002/asi.24253.

Field, F. (2010) *The Foundation Years: preventing poor children becoming poor adults*. available at: www.frankfield.co.uk http://www.inspiredbybabies.org.uk/Page2NationalrelevantDocsresources/Frank Field Preventing poor children becoming poor adults 2011.pdf.

‘Finland: Broadband Access Made Legal Right in Landmark Law’ (2010). available at: https://www.huffpost.com/entry/finland-broadband-access_n_320481 (accessed: 23 March 2021).

Firth, E. (2019) ‘Personal data has value in so many different ways’. digi.me. available at: https://blog.digi.me/2019/09/04/personal-data-has-so-much-more-value-than-pure-cash/.

Fisch, J. (2015) ‘Individual self-determination’, in. Cambridge University Press.

Forrester, I. (2021) 'Talking about human values and design', *BBC Research & Development*. available at: https://www.bbc.co.uk/rd/blog/2021-07-talking-about-human-values-and-design.

Forrester, I., Thompson, B. and Jones, R. (2016) 'Databox and the BBC box', *BBC R&D Blog*. available at: https://www.bbc.co.uk/rd/projects/databox (accessed: 26 February 2019).

Foucault, M. (1975) 'Discipline and punish: The birth of the prison', *New York*. Pantheon Books.

Foucault-Dumas, C. (2021) '"This case forces uber to do a lot more than ever before", says data protection expert', *Medium*. available at: https://medium.com/personaldata-io/uber-vs-drivers-trial-interview-data-protection-expert-rene-mahieu-55359f8cdd9d.

Foucault-Dumas, C. (2022) 'The eyeballs: We are the targeted audience'. available at: https://eyeballs.hestialabs.org/en/.

Foulonneau, Muriel. and Riley, Jenn. (2008) *Metadata for digital resources : implementation, systems design and interoperability*. Chandos Pub, p. 203.

Fowler, M. and Highsmith, J. (2001) 'The agile manifesto', *Software Development*. [San Francisco, CA: Miller Freeman, Inc., 1993-, 9(8), pp. 28–35.

Freeman, E. and Gelernter, D. (1996) 'Lifestreams: A Storage Model for Personal Data', *SIGMOD Record (ACM Special Interest Group on Management of Data)*. Association for Computing Machinery (ACM), 25(1), pp. 80–86. doi: 10.1145/381854.381893.

Friedman, B. and Hendry, D. G. (2019) *Value Sensitive Design: Shaping Technology with Moral Imagination*. MIT Press (The MIT press). available at: https://books.google.co.uk/books?id=8ZiWDwAAQBAJ.

Friedman, R. L. (2006) 'Deweyan Pragmatism', *William James Studies*, 1. available at: https://williamjamesstudies.org/deweyan-pragmatism/.

Frost, A. (2019) 'Forget Folders: The Best Ways to Organize Your Files with Tags and Labels'. available at: https://zapier.com/blog/how-to-use-tags-and-labels/.

Fu, S. *et al.* (2020) 'Social media overload, exhaustion, and use discontinuance: Examining the effects of information overload, system feature overload, and social overload', *Information Processing and Management*, 57(6). doi: 10.1016/j.ipm.2020.102307.

Gauntlett, D. (2015) 'What is digital DIY?', *Digital Do It Yourself*. available at: https://www.didiy.eu/blogs/what-digital-diy/.

Gayomali, C. (2015) 'Why twitter is killing TweetDeck'. The Week. available at: https://theweek.com/articles/467040/why-twitter-killing-tweetdeck.

Gelernter, D. (1994) 'The cyber-road not taken: Lost on the info-highway? Here's some stuff that could really change your life.', *The Washington Post*, 3.

Gellman, B. (2013) 'Edward Snowden, after months of NSA revelations, says his mission's accomplished', *The Washington Post*, 23. available at: http://www.washingtonpost.com/world/national-security/edward-snowden-after-months-of-nsa-revelations-says-his-missions-accomplished/2013/12/23/49fc36de-6c1c-11e3-a523-fe73f0ff6b8d_story.html%5Cnhttp://www.washingtonpost.com/world/national-security/edward-.

Gemmell, J., Bell, G. and Lueder, R. (2006) 'MyLifeBits: A personal database for everything', *Communications of the ACM*, 49(1), pp. 88–95. doi: 10.1145/1107458.1107460.

Gener8 (2021) 'Gener8'. available at: https://gener8ads.com/ (accessed: 23 August 2021).

Gheera, M. (2011) *The ContactPoint database*. The House of Commons Library. available at: https://researchbriefings.files.parliament.uk/documents/SN05171/SN05171.pdf.

Giaccardi, E. and Redström, J. (2020) 'Technology and more-than-human design', *Design Issues*. MIT Press, 36(4), pp. 33–44.

Gillespie, T. and Seaver, N. (2016) 'Critical Algorithm Studies - A Reading List'. available at: https://socialmediacollective.org/reading-lists/critical-algorithm-studies/.

Gitelman, L. (2013) *Raw data is an oxymoron*. edited by Lisa Gitelman. MIT Press, p. 182. available at: https://mitpress.mit.edu/books/raw-data-oxymoron.

Glavic, B. *et al.* (2021) 'Trends in Explanations: Understanding and Debugging Data-driven Systems', *Foundations and Trendsin Databases*. Now Publishers, Inc., 11(3), pp. 226–318. doi: 10.1561/XXXXXXXXX.Boris.

Goffe, L. *et al.* (2021) 'Appetite for disruption: Designing human-centred augmentations to an online food ordering platform', in *34th british HCI conference*, pp. 155–167.

Goffe, L. *et al.* (2022) 'Web augmentation for well-being: The human-centred design of a takeaway food ordering digital platform', *Interacting with Computers*.

Golder, S. A. and Huberman, B. A. (2006) 'Usage patterns of collaborative tagging systems', *Journal of Information Science*, 32, pp. 198–208. doi: 10.1177/0165551506062337.

Golembewski, M. and Selby, M. (2010) 'Ideation decks', in *Proceedings of the 8th ACM conference on designing interactive systems - DIS '10*. New York, New York, USA: ACM Press, p. 89. doi: 10.1145/1858171.1858189.

Gonscherowski, S. and Bieker, F. (2018) 'Who You Gonna Call When There's Something Wrong in Your Processing? Risk Assessment and Data Breach Notifications in Practice', in *IFIP international summer school on privacy and identity management*. Springer, pp. 35–50.

Goodwins, R. (2021) 'Sir tim berners-lee and the BBC stage a very british coup to rescue our data from facebook and friends', *The Register*. available at: https://www.theregister.com/2021/10/04/column_data_privacy/ (accessed: 25 August 2022).

'Google Desktop Search' (2004). available at: https://en.wikipedia.org/wiki/Google_Desktop.

Guba, E. G. (1990) 'The alternative paradigm dialog', *The paradigm dialog*. Sage Publications, Inc, pp. 17–30. available at: http://www.jstor.org/stable/3340973.

Gulite, K. (2014) 'What is civil libertarianism?', *ThoughtsOnLiberty: Women Writing for a Free World*.

Gurstein, M. (2003) 'Effective use: A community informatics strategy beyond the digital divide', *First Monday*, 8(12). doi: 10.5210/fm.v0i0.1798.

Gurstein, M. B. (2011) 'Open data: Empowering the empowered or effective data use for everyone?', *First Monday*. First Monday, 16(2). doi: 10.5210/fm.v16i2.3316.

Haddadi, H. *et al.* (2015) 'Personal data: Thinking inside the box', *arXiv preprint arXiv:1501.04737*. doi: https://doi.org/10.48550/arXiv.1501.04737.

Haddadi, H. *et al.* (2016) 'About the databox project', *databox project*. available at: https://web.archive.org/web/20180825180114/https://www.databoxproject.uk/about (accessed: 25 August 2018).

Hall, W., Tinati, R. and Jennings, W. (2018) 'From brexit to trump: Social media's role in democracy', *Computer*. IEEE, 51(1), pp. 18–27.

Hamon, R. *et al.* (2021) 'Impossible Explanations? Beyond explainable AI in the GDPR from a COVID-19 use case scenario', in *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pp. 549–559.

Harbird, R. (2006) 'Novel Applications for Information Technology in Risk Assessment for Children's Social Care in the UK', *Rn*. available at: http://www.cs.ucl.ac.uk/research/researchnotes/documents/RN_06_11.pdf.

Härkönen, T. *et al.* (2022) *Tracking digipower: How data can be used for influencing decision-makers and steering the world*. Sitra. available at: https://www.sitra.fi/en/publications/tracking-digipower/.

Härkönen, T. and Vänskä, R. (2021). Sitra. available at: https://www.sitra.fi/en/projects/digipower-investigation/#what-is-it-about.

Harris, T. (2013a) 'A Call to Minimize Distraction Respect Users' Attention'. available at: http://www.minimizedistraction.com/.

Harris, T. (2013b) 'Who We Are: Center for Humane Technology (CHT)'. available at: https://www.humanetech.com/who-we-are.

Harris, T. (2016) 'How Technology Hijacks People's Minds — from a Magician and Google's Design Ethicist'. available at: https://www.tristanharris.com/2016/05/how-technology-

hijacks-peoples-minds-from-a-magician-and-googles-design-ethicist/ (accessed: 22 March 2019).

Hart-Davidson, W., Zachry, M. and Spinuzzi, C. (2012) 'Activity streams: Building context to coordinate writing activity in collaborative teams', in *SIGDOC'12 - proceedings of the 30th ACM international conference on design of communication*. New York, New York, USA: ACM Press, pp. 279–287. doi: 10.1145/2379057.2379109.

Hartman, T. *et al.* (2020) 'Public perceptions of good data management: Findings from a UK-based survey:', *https://doi.org/10.1177/2053951720935616*. SAGE PublicationsSage UK: London, England, 7. doi: 10.1177/2053951720935616.

Hatmaker, T. (2018) 'Twitter is killing its twitter for mac desktop client'. TechCrunch.

Havlak, H. and Abelson, B. (2016) 'The definitive list of what everyone likes on facebook', *The Verge*. available at: https://www.theverge.com/2016/2/1/10872792/facebook-interests-ranked-preferred-audience-size.

Hayes, G. R. (2011) 'The relationship of action research to human-computer interaction', *ACM Transactions on Computer-Human Interaction*, 18(3), pp. 1–20. doi: 10.1145/1993060.1993065.

'HDI Lab, Heerlen' (2020). available at: https://hdilab.com/.

'HDI Network Plus, University of Glasgow' (2018). available at: https://hdi-network.org/.

Helmond, A. (2015) 'The platformization of the web: Making web data platform ready', *Social media+ society*. Sage Publications Sage UK: London, England, 1(2), p. 2056305115603080.

Hemp, P. (2009) 'Death by Information Overload'. available at: https://hbr.org/2009/09/death-by-information-overload (accessed: 23 March 2021).

Henderson, I. and B-s Working Group (2020) 'Customer — Supplier Engagement Framework Explained', pp. 1–7. available at: https://me2ba.org/wp-content/uploads/2020/09/customer-supplier-engagement-framework-updated-9-28.pdf.

Hendler, J. and Berners-Lee, T. (2010) 'From the Semantic Web to social machines: A research challenge for AI on the World Wide Web'. doi: 10.1016/j.artint.2009.11.010.

Herselman, M. *et al.* (2016) 'A Digital Health Innovation Ecosystem for South Africa', in *2016 IST-africa conference, IST-africa 2016*. doi: 10.1109/ISTAFRICA.2016.7530615.

Hixon, J. G. and Swann, W. B. (1993) 'When Does Introspection Bear Fruit? Self-Reflection, Self-Insight, and Interpersonal Choices', *Journal of Personality and Social Psychology*, 64(1), pp. 35–43. doi: 10.1037/0022-3514.64.1.35.

'Hobson's Choice' (no date). available at: https://www.britannica.com/dictionary/Hobson%27s-choice.

Hoffman, W. (2010) 'Rethinking Personal Data'. available at: https://web.archive.org/web/20110220013300/http://www.weforum.org/issues/rethinking-personal-data.

Hoffman, W. (2011) *Personal data : The emergence of a new asset class*. World Economic Forum, pp. 1–40. available at: http://www.weforum.org/reports/personal-data-emergence-new-asset-class.

Hoffman, W. (2013) *Unlocking the Value of Personal Data: From Collection to Usage Prepared in collaboration with The Boston Consulting Group Industry Agenda*. February. World Economic Forum.

Hoffman, W. (2014a) *Rethinking Personal Data : A New Lens for Strengthening Trust*. May. World Economic Forum, p. 35. available at: http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf.

Hoffman, W. (2014b) *Rethinking personal data: Trust and context in user-centred data ecosystems*. May. World Economic Forum, p. 35. available at: http://www3.weforum.org/docs/WEF_RethinkingPersonalData_TrustandContext_Report_2014.pdf.

Hogan, T. (2012) *Toward a phenomenology of human-data relations*. available at: http://www.manovich.net/DOCS/data_art.doc,.

Holmes, L. *et al.* (2012) *Exploration of the costs and impact of the common assessment framework*. Department for Education.

Honeyman, M., Dunn, P. and Mckenna, H. (2016) *A digital NHS?*

Hoofnagle, C. J., Sloot, B. van der and Borgesius, F. Z. (2019) 'The European Union general data protection regulation: What it is and what it means', *Information and Communications Technology Law*. Taylor & Francis, 28(1), pp. 65–98. doi: 10.1080/13600834.2019.1573501.

Hori, S. (2021) 'Self-sovereign identity: The future of personal data ownership?', *World Economic Forum*. available at: https://www.weforum.org/agenda/2021/08/self-sovereign-identity-future-personal-data-ownership/.

Horwitz, J. *et al.* (2021) 'The facebook files', *Wall Street Journal*. available at: https://www.wsj.com/articles/the-facebook-files-11631713039.

Hosch, W. L. (2017) 'Web 2.0'. available at: https://www.britannica.com/topic/Web-20 (accessed: 26 April 2021).

Hotho, A., Nürnberger, A. and Paaß, G. (2005) 'A brief survey of text mining.', in *Ldv forum*. Citeseer (1), pp. 19–62.

Huberman, M. and Miles, M. B. (2002) *The qualitative researcher's companion*. Sage.

'Human Data Interaction Project at the Data to AI Lab, MIT' (2015). available at: https://hdi-dai.lids.mit.edu/.

Human, S. and Cech, F. (2021) 'A human-centric perspective on digital consenting: The case of GAFAM', *Smart Innovation, Systems and Technologies*, 189, pp. 139–159. doi: 10.1007/978-981-15-5784-2_12.

Hutton, D. M. (2012) 'Turing's Cathedral: The Origins of the Digital Universe'. Emerald Group Publishing Limited.

Hwang, E. (2021) 'Sketching Dialogue : Incorporating Sketching in Emphatic Semi-structured Interviews for HCI'.

Ihde, D. (1990) *Technology and the lifeworld: From garden to earth*. Indiana University Press.

'Information' (no date). available at: https://en.wikipedia.org/wiki/Information.

Information Commissioner's Office (2014) 'Data controllers and data processors: what the difference is and what the governance implications are', p. 20. available at: https://ico.org.uk/for-organisations/guide-to-data-protection/introduction-to-data-protection/some-basic-concepts/.

Information Commissioner's Office (2018) 'Your data matters - Your rights'. available at: https://ico.org.uk/your-data-matters/.

Information Commissioner's Office (2021a) 'Your right of access'. available at: https://ico.org.uk/your-data-matters/your-right-to-get-copies-of-your-data/ (accessed: 23 August 2021).

Information Commissioner's Office (2021b) 'Your right to data portability'.

'Infovark Company Profile' (2007). available at: https://www.crunchbase.com/organization/infovark.

Jasperson, J. (Sean). *et al.* (2002) 'Review: Power and Information Technology Research: A Metatriangulation Review'. Society for Information Management; The Management Information Systems Research Center. doi: 10.2307/4132315.

Jeffers, S. and Webb, L. K. (2017) 'About who targets me'.

Jelly, M. (2021) 'The Mission'. ethi.me. available at: https://web.archive.org/web/20220312232859/https://www.ethi.me/the-mission (accessed: 12 March 2021).

Jenkins, H. (2006) *Convergence Culture: Where Old and New Media Collide*. New York, USA: New York University Press. doi: 10.7551/mitpress/9780262036016.003.0012.

Jensen, C. *et al.* (2010) 'The life and times of files and information: A study of desktop provenance'.

Jilek, C. *et al.* (2018) 'Context spaces as the cornerstone of a near-transparent and self-reorganizing semantic desktop', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11155 LNCS, pp. 89–94. doi: 10.1007/978-3-319-98192-5_17.

Johnson, S. L., Kim, Y. M. and Church, K. (2010) 'Towards client-centered counseling: Development and testing of the WHO Decision-Making Tool', *Patient Education and Counseling*. Elsevier Ireland Ltd, 81(3), pp. 355–361. doi: 10.1016/j.pec.2010.10.011.

Jones, T. (2011) 'Designing for second screens : The Autumnwatch Companion'. available at: https://www.bbc.co.uk/blogs/researchanddevelopment/2011/04/the-autumnwatch-companion---de.shtml.

Jones, W. *et al.* (2006) '"It's about the information stupid!": Why we need a separate field of human-information interaction', *Conference on Human Factors in Computing Systems - Proceedings*, pp. 65–68. doi: 10.1145/1125451.1125469.

Jones, W. (2011a) 'The Future of Personal Information Management Part I: Our Information, Always and Forever'.

Jones, W. (2011b) 'The Future of Personal Information Management Part I: Our Information, Always and Forever', p. 72.

Kalvet, T. (2005) 'Digital divide and the ICT paradigm generally and in estonia', in *Encyclopedia of developing regional communities with information and communication technology*. IGI Global, pp. 182–187. doi: 10.4018/978-1-59140-575-7.ch032.

Kanter, J. (2021) 'BBC and sir tim berners-lee app mines netflix data to find shows viewers like', *The Times*. available at: https://www.thetimes.co.uk/article/bbc-and-sir-tim-berners-lee-app-mines-netflix-data-to-find-shows-viewers-like-lxp002gg8 (accessed: 25 August 2022).

Karger, D. R. *et al.* (2005) 'Haystack: A customizable general-purpose information management tool for end users of semistructured data', in *2nd biennial conference on innovative data systems research, CIDR 2005*, pp. 13–27. available at: https://s3.amazonaws.com/academia.edu.documents/46870765/haystack.pdf.

Karger, D. R. and Jones, W. (2006) 'Data unification in personal information management', *Communications of the ACM*, 49(1), p. 77. doi: 10.1145/1107458.1107496.

Kasirzadeh, A. and Clifford, D. (2021) *Fairness and Data Protection Impact Assessments*. Association for Computing Machinery (1), pp. 146–153. doi: 10.1145/3461702.3462528.

Kaye, J. *et al.* (2015) 'Dynamic consent: a patient interface for twenty-first century research networks', *European Journal of Human Genetics*. Nature Publishing Group, 23(2), pp. 141–146. doi: 10.1038/ejhg.2014.71.

Keegan, J. and Ng, A. (2022) 'Lawsuit highlights how little control brokers have over location data', *The Markup*. available at: https://themarkup.org/privacy/2022/03/21/lawsuit-highlights-how-little-control-brokers-have-over-location-data (accessed: 25 May 2023).

Kelly, K. (2023) 'The ring video doorbell and the entry of amazon into the smart home: Implications for consumer-initiated surveillance', *Journal of Consumer Policy*. Springer, 46(1), pp. 95–104.

Kelly, K. and Wolf, G. (2007) 'What is the quantified self'. available at: https://web.archive.org/web/20100507215130/http://www.kk.org/quantifiedself/2007/10/what-is-the-quantifiable-self.php.

Kelly, R. (2020) 'The Biggest ICO Fines Ever Issued'. available at: https://digit.fyi/data-protection-2020-the-biggest-fines-ever-issued-by-the-ico/.

Kelty, C. M. (2008) *Geeks and Recursive Publics*. Duke University Press, pp. 27–63.

Kensing, F. and Blomberg, J. (1998) 'Participatory design: Issues and concerns', *Computer supported cooperative work (CSCW)*. Springer, 7(3), pp. 167–185.

Kirven, A. (2018) 'Whose gig is it anyway: Technological change, workplace control and supervision, and workers' rights in the gig economy', *U. Colo. L. Rev.* HeinOnline, 89, p. 249.

Klatzky, S. R. (1970) 'Automation, size, and the locus of decision making: the cascade effect', *The Journal of Business*. JSTOR, 43(2), pp. 141–151. available at:

https://www.jstor.org/stable/pdf/2352107.pdf?refreqid=excelsior%3A24bde6bf7de0eccf42c6ea11f8446d38.

Klein, B. *et al.* (2004) 'Enabling flow - A paradigm for document-centered personal information spaces', in *Proceedings of the eighth IASTED international conference on artificial intelligence and soft computing*, pp. 187–192. available at: https://www.semanticscholar.org/paper/Enabling-flow%3A-%7BA%7D-paradigm-for-document-centered-Klein-Agne/22be4a7b25e75de235e5d96bad6ab4ab4583daac.

Knight, M. (2016) 'What is data literacy?', *Dataversity*. available at: https://www.dataversity.net/what-is-data-literacy/.

Koene, A. *et al.* (2018) 'UnBias: Emancipating users against algorithmic biases for a trusted digital economy'.

Kollnig, K. (2021) 'TrackerControl'. available at: https://www.trackercontrol.org/.

Kollnig, K., Datta, S. and Kleek, M. V. (2021) 'I want my app that way: Reclaiming sovereignty over personal devices; i want my app that way: Reclaiming sovereignty over personal devices', *CHI Extended Abstracts*. ACM. doi: 10.1145/3411763.3451632.

Kosta, E., Leenes, R. and Kamara, I. (2022) *Research handbook on EU data protection law*. Edward Elgar Publishing.

Kostkova, P. (2015) 'Grand Challenges in Digital Health', *Frontiers in Public Health*. Frontiers Media SA, 3. doi: 10.3389/fpubh.2015.00134.

Kriisk, K. and Minas, R. (2017) 'Social rights and spatial access to local social services: The role of structural conditions in access to local social services in Estonia', *Social Work and Society*, 15(1). available at: https://www.socwork.net/sws/article/view/503/1007.

Krishnan, A. (2010) 'Pervasive Personal Information Spaces'. University of Waikato. available at: https://researchcommons.waikato.ac.nz/handle/10289/4590.

Krishnan, A. and Jones, S. (2005) 'TimeSpace: Activity-based temporal visualisation of personal information spaces', *Personal and Ubiquitous Computing*, 9(1), pp. 46–65. doi: 10.1007/s00779-004-0291-x.

Kröger, J. L., Miceli, M. and Müller, F. (2021) 'How Data Can Be Used Against People: A Classification of Personal Data Misuses', *SSRN Electronic Journal*, (December). doi: 10.2139/ssrn.3887097.

Kuchler, H. (2018) 'Max schrems: The man who took on facebook — and won', *The Financial Times*. available at: https://www.ft.com/content/86d1ce50-3799-11e8-8eee-e06bde01c544.

Lansdale, M. W. (1988) 'The psychology of personal information management', *Applied Ergonomics*, 19(March 1988), pp. 55–66. doi: 10.1016/0003-6870(88)90199-8.

Lansdale, M. and Edmonds, E. (1992) 'Using memory for events in the design of personal filing systems', *International Journal of Man-Machine Studies*, 36(1), pp. 97–126. doi: 10.1016/0020-7373(92)90054-O.

Larsson, S. (2018) 'Algorithmic governance and the need for consumer empowerment in data-driven markets', *Internet Policy Review*, 7(2). doi: 10.14763/2018.2.791.

Le Dantec, C. A. (2016) *Designing publics*. MIT Press.

Lecluijze, I. *et al.* (2015) 'Co-production of ICT and children at risk: The introduction of the Child Index in Dutch child welfare', *Children and Youth Services Review*. Elsevier Ltd, 56, pp. 161–168. doi: 10.1016/j.childyouth.2015.07.003.

Lemley, M. A. (2021) 'The splinternet', *Duke Law Journal*, pp. 1397–1428. available at: https://perma.cc/92LZ-B8DN].

Leprince-Ringuet, D. (2021). available at: https://www.zdnet.com/article/gdpr-fines-increased-by-40-last-year-and-theyre-about-to-get-a-lot-bigger/.

Lessig, L. (2000) 'Code is law: On liberty in cyberspace', *Harvard Magazine*. available at: https://www.harvardmagazine.com/2000/01/code-is-law-html.

Levine, R. (2011) 'How the internet has all but destroyed the market for films, music and newspapers'. available at: https://www.theguardian.com/media/2011/aug/14/robert-levine-digital-free-ride (accessed: 23 March 2021).

Levitas, J. (2013) 'Defining civic hacking', *Medium*. available at: https://medium.com/civic-innovation/defining-civic-hacking-16844fc161cd.

Lewin, K. (1946) 'Action Research and Minority Problems', *Journal of Social Issues*, 2(4), pp. 34–46. doi: 10.1111/j.1540-4560.1946.tb02295.x.

Lewin, K. (1951) 'Problems of research in social psychology', *Field theory in social science: Selected theoretical papers*, pp. 155–169.

Li, I. (2009) 'Designing Personal Informatics Applications and Tools that Facilitate Monitoring of Behaviors', *Uist*. available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.232.8536.

Li, I., Forlizzi, J. and Dey, A. (2010) 'Know thyself: Monitoring and reflecting on facets of one's life', *Conference on Human Factors in Computing Systems - Proceedings*, pp. 4489–4492. doi: 10.1145/1753846.1754181.

Lindley, S. E. *et al.* (2018) 'Exploring new metaphors for a networked world through the file biography', *Conference on Human Factors in Computing Systems - Proceedings*, 2018-April, pp. 1–12. doi: 10.1145/3173574.3173692.

'List of target companies for GDPR requests' (no date). available at: https://wiki.personaldata.io/wiki/Item:Q2369 (accessed: 22 September 2021).

Local Government Association (2017) *Adult social care funding: State of the nation*. Local Government Association.

Lohr, S. (2021) 'He created the web. Now he's out to remake the digital world', *The New York Times*. available at: https://www.nytimes.com/2021/01/10/technology/tim-berners-lee-privacy-internet.html.

Lomas, N. (2020) 'UK's ICO faces legal action after closing adtech complaint with nothing to show for it'. available at: https://techcrunch.com/2020/11/05/uks-ico-faces-legal-action-after-closing-adtech-complaint-with-nothing-to-show-for-it/.

Lomas, N. (2021) 'Dutch court rejects uber drivers' "robo-firing" charge but tells ola to explain algo-deductions', *TechCrunch*. available at:

https://techcrunch.com/2021/03/12/dutch-court-rejects-uber-drivers-robo-firing-charge-but-tells-ola-to-explain-algo-deductions/.

Lomas, N. (2022) 'TikTok 'pauses' privacy policy switch in europe after regulatory scrutiny'. TechCrunch. available at: https://techcrunch.com/2022/07/12/tiktok-pauses-privacy-policy-switch/.

Loon, A. van (2018) 'Publishers and google as joint controllers. How to avoid joint and several liability?', *LinkedIn Pulse*. available at: https://www.linkedin.com/pulse/publishers-google-joint-controllers-how-avoid-several-ad/.

Lowe, T. and Wilson, R. (2015) 'Playing the game of outcomes-based performance management', *Is Gamesmanship Inevitable*.

Lufkin, B. (2017) '50 grand challenges for the 21st century', *BBC Future*. available at: https://www.bbc.com/future/article/20170331-50-grand-challenges-for-the-21st-century.

Luger, E. and Rodden, T. (2013) 'An informed view on consent for ubicomp', in *UbiComp 2013 - proceedings of the 2013 ACM international joint conference on pervasive and ubiquitous computing*. New York, New York, USA: ACM Press, pp. 529–538. doi: 10.1145/2493432.2493446.

Lukes, S. M. (2020) 'Individualism', in *Encyclopedia britannica*. available at: https://www.britannica.com/topic/individualism (accessed: 31 July 2022).

Lupton, D. (2016) 'Personal data practices in the age of lively data', *Digital sociologies*. Policy Press Bristol, 2016, pp. 335–350.

Lupton, D. (2020) *Data selves: More-than-human perspectives*. Polity Cambridge.

Lynskey, O. (2015) *The foundations of EU data protection law*. Oxford University Press.

Lyon, D. (2001) *Surveillance society*. McGraw-Hill Education (UK).

Macaskill, E. *et al.* (2013) 'NSA files decoded: Edward snowden's surveillance revelations explained', *The Guardian*. available at:

https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded.

Mahieu, R. L. P. and Ausloos, J. (2020) 'Harnessing the collective potential of GDPR access rights : towards an ecology of transparency', *Internet Policy Review*. available at: https://policyreview.info/articles/news/harnessing-collective-potential-gdpr-access-rights-towards-ecology-transparency/1487.

Mahieu, R. and Ausloos, J. (2020) 'Recognising and Enabling the Collective Dimension of the GDPR and the Right of Access'.

Maldre, M. (2012) 'Amazon makes kindle less social -'. available at: https://www.mattmaldre.com/2012/10/01/amazon-makes-kindle-less-social/.

Malomo, F. and Sena, V. (2017) 'Data Intelligence for Local Government? Assessing the Benefits and Barriers to Use of Big Data in the Public Sector', *Policy and Internet*, 9(1), pp. 7–27. doi: 10.1002/poi3.141.

Malone, T. W. (1983) 'How do people organize their desks?: Implications for the design of office information systems', *ACM Transactions on Information Systems*, 1(1), pp. 99–112. doi: 10.1145/357423.357430.

Mannay, D. and Morgan, M. (2015) 'Doing ethnography or applying a qualitative technique? Reflections from the 'waiting field'', *Qualitative research*. Sage Publications Sage UK: London, England, 15(2), pp. 166–182.

Mansour, E. *et al.* (2016) 'A demonstration of the solid platform for social web applications', in *Proceedings of the 25th international conference companion on world wide web*, pp. 223–226.

Marchionini, G. (2008) 'Human-information interaction research and development', *Library and Information Science Research*, 30(3), pp. 165–174. doi: 10.1016/j.lisr.2008.07.001.

Marshall, C. (2019) 'What is named entity recognition (NER) and how can i use it?', *Medium*. available at: https://medium.com/mysuperai/what-is-named-entity-recognition-ner-and-how-can-i-use-it-2b68cf6f545d.

Marshall, C. C. and Jones, W. (2006) 'Keeping encountered information', *Communications of the ACM*, 49(1), pp. 66–67. doi: 10.1145/1107458.1107493.

Marshall, M. (2020) 'Markdown thesis'. available at: https://gitlab.com/mrshll1001/markdown-thesis.

Martin, M. (2007) *Research note: Representing identity and relationships in information systems*, *International Journal of Business Science & Applied Management (IJBSAM) Suggested Citation: Martin*, pp. 47–51. available at: http://hdl.handle.net/10419/190583https://creativecommons.org/licenses/by/2.0/uk/.

Martin, M. (2022) 'The trustworthy, governable platform: A concept of safe social spaces in the public network'. in prep.

Maskell, T. *et al.* (2018) 'Spokespeople: Exploring routes to action through citizen-generated data', in *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–12.

McAdams, D. P. (1996) 'Personality, modernity, and the storied self: A contemporary framework for studying persons', *Psychological inquiry*. Taylor & Francis, 7(4), pp. 295–321.

McCarthy, J. and Wright, P. (2004) 'Technology as experience', *Interactions*, 11(5), pp. 42–43. doi: 10.1145/1015530.1015549.

McCullagh, K. (2009) 'Protecting 'privacy' through control of 'personal' data processing: A flawed approach', *International Review of Law, Computers and Technology*, 23(1-2), pp. 13–24. doi: 10.1080/13600860902742562.

McNamee, R. (2019) *Zucked: Waking up to the Facebook Catasrophe*, p. 336.

Melendez, S. and Pasternack, A. (2019) 'The data brokers quietly buying and selling your personal information'. available at: https://www.fastcompany.com/90310803/here-are-the-data-brokers-quietly-buying-and-selling-your-personal-information.

Merriam-Webster Dictionary (no date a) 'Usability'. available at: https://www.merriam-webster.com/dictionary/usability.

Merriam-Webster Dictionary (no date b) ‘Usable’. available at: https://www.merriam-webster.com/dictionary/usable.

Meschtscherjakov, A., Wilfinger, D. and Tscheligi, M. (2014) ‘Mobile attachment- Causes and consequences for emotional bonding with mobile phones’, *Conference on Human Factors in Computing Systems - Proceedings*, pp. 2317–2326. doi: 10.1145/2556288.2557295.

Miagkov, A., Gillula, J. and Cyphers, B. (2019) ‘Google’s plans for chrome extensions won’t really help security’, *Electronic Frontier Foundation*. available at: https://www.eff.org/deeplinks/2019/07/googles-plans-chrome-extensions-wont-really-help-security.

Microsoft (2021) ‘Project Bali’. available at: https://www.microsoft.com/en-us/research/project/bali/ (accessed: 23 August 2021).

Miettinen, R. (2013) *Innovation, human capabilities, and democracy: Towards an enabling welfare state*. Oxford University Press.

Mill, J. van and Quintais, J. P. (2022) ‘A matter of (joint) control? Virtual assistants and the general data protection regulation’, *Computer Law & Security Review*. Elsevier, 45, p. 105689.

Millar, S. (2002) ‘UK singled out for criticism over protection of privacy’. available at: https://www.theguardian.com/technology/2002/sep/05/security.humanrights.

Miller, A. (2021) ‘The right to repair movement, explained – BMC software | blogs’, *The Business IT Blog*. available at: https://www.bmc.com/blogs/right-to-repair/.

Milmo, D. (2021) *The Guardian*. available at: https://www.theguardian.com/uk-news/2021/oct/14/amazon-asks-ring-owners-to-respect-privacy-after-court-rules-usage-broke-law (accessed: 25 May 2023).

Moore-Colyer, R. (2022) ‘Apple launches iPhone self service repair kits — but there’s a big catch | tom’s guide’, *Tom’s Guide*. available at: https://www.tomsguide.com/news/apple-launches-iphone-self-service-repair-kits-but-theres-a-big-catch.

Mora, S., Gianni, F. and Divitini, M. (2017) 'Tiles: A card-based ideation toolkit for the internet of things', in *Proceedings of the 2017 conference on designing interactive systems*, pp. 587–598.

Moraveji, N. *et al.* (2007) 'Comicboarding: Using comics as proxies for participatory design with children', in *Conference on Human Factors in Computing Systems - Proceedings*. ACM, pp. 1371–1374. doi: 10.1145/1240624.1240832.

Morgan, J. (2020) 'Making your Solid Apps interoperable with ShapeRepo.com'. available at: https://medium.com/@JacksonMorgan/making-your-solid-apps-interoperable-with-shaperepo-com-8da512936073.

Morozov, E. (2013) *To save everything, click here: The folly of technological solutionism*. Public Affairs.

Mortier, R. *et al.* (2013) 'Challenges & opportunities in human-data interaction', *University of Cambridge, Computer Laboratory*. Citeseer. doi: 10.5210/fm.v17i5.4013.

Mortier, R. *et al.* (2014) 'Human-data interaction: The human face of the data-driven society', *Available at SSRN 2508051*. doi: 10.2139/ssrn.2508051.

Mühlhoff, R. (2023) 'Predictive privacy: Collective data protection in the context of artificial intelligence and big data', *Big Data & Society*. SAGE Publications Sage UK: London, England, 10(1), p. 20539517231166886.

Mundie, C. (2014) 'Privacy pragmatism; focus on data use, not data collection', *Foreign Affairs*. HeinOnline, 93, p. 28.

Murton, D. (2011) 'A Brief History of the Evolution of Social Technology'. available at: https://www.scottmonty.com/2011/04/brief-history-of-evolution-of-social.html.

MyData (2017) 'Declaration - MyData.org'. available at: https://web.archive.org/web/20210325143142/https://www.mydata.org/declaration/ (accessed: 25 March 2021).

'MyData Comparison of Principles document' (2017). available at: http://bit.ly/pd-principles.

MyData.org (2018) ‘MyData - Who we are’. available at: https://web.archive.org/web/20210624020733/https://www.mydata.org/about/ (accessed: 24 June 2021).

Mydex CIC (2010) ‘The Case for Personal Information Empowerment : The rise of the personal data store’, *World*, pp. 1–44.

mySociety (2004) ‘About TheyWorkForYou’. available at: https://www.theyworkforyou.com/about/.

‘myTimeline’ (2018). available at: https://www.timelineinc.com/ (accessed: 23 March 2021).

Nadeem, D. and Sauermann, L. (2007) ‘From Philosophy and Mental-Models to Semantic Desktop Research: Theoretical Overview’.

Neef, D. (2015) *Digital exhaust: what everyone should know about big data, digitization and digitally driven innovation*. Pearson Education.

Neff, G. (2013) ‘Why Big Data Won’t Cure Us’, *Big Data*, 1(3), pp. 117–123. doi: 10.1089/big.2013.0029.

Negroponte, N. and Bolt, R. A. (1978) *Spatial data management system*. MASSACHUSETTS INST OF TECH CAMBRIDGE ARCHITECTURE MACHINE GROUP.

Nelson, T. (2006) ‘Lost in hyperspace’, *New Scientist*, 191(2561). doi: 10.1002/elsc.200620112.

Nelson, T. H. (1965) ‘Complex information processing’, pp. 84–100. doi: 10.1145/800197.806036.

Newton, C. (2018) ‘Twitter officially kills off key features in third-party apps’. The Verge. available at: https://www.theverge.com/2018/8/16/17699626/twitter-third-party-apps-streaming-api-deprecation.

Ng, I. C. (2018) ‘Can you own your personal data? The hat (hub-of-all-things) data ownership model’. Warwick Manufacturing Group.

Nielsen, J. (2012) 'Usability 101: Introduction to Usability'. available at: https://www.nngroup.com/articles/usability-101-introduction-to-usability/.

Norman, D. A. and Draper, S. W. (1986) 'User Centered System Design; New Perspectives on Human-Computer Interaction'. L. Erlbaum Associates Inc.

Northern Health Science Alliance (2020) *Connected Health Cities Impact Report*. Newcastle upon Tyne, UK: Northern Health Science Alliance, pp. 129–130.

O'Donnell, B. (2020) 'Zoom, the office and the future: What will work look like after coronavirus?' available at: https://eu.usatoday.com/story/tech/columnist/2020/09/07/zoom-work-from-home-future-office-after-coronavirus/5680284002/.

O'Donoghue, T. and Rabin, M. (2001) 'Choice and procrastination', *The Quarterly Journal of Economics*. MIT Press, 116(1), pp. 121–160.

O'Sullivan, C. (2021) 'What is responsible AI?', *Medium*. available at: https://towardsdatascience.com/what-is-responsible-ai-548743369729.

O'Sullivan, D., Duffy, C. and Fung, B. (2022) 'Ex-twitter exec blows the whistle, alleging reckless and negligent cybersecurity policies', *CNN Business*. available at: https://edition.cnn.com/2022/08/23/tech/twitter-whistleblower-peiter-zatko-security/index.html.

Odom, W. *et al.* (2018) 'Time, Temporality, and Slowness', pp. 383–386. doi: 10.1145/3197391.3197392.

OFSTED (2015) *Early help: whose responsibility?*, p. 32. available at: www.ofsted.gov.uk https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/410378/Early_help_whose_responsibility.pdf.

Olivo, O. (2020) 'The road ahead for inrupt and solid'. available at: https://inrupt.com/Solid-roadmap-preview.

Open Lab (2014) 'Centre for doctoral training in digital civics: About the CDT'. Newcastle University. available at:

https://web.archive.org/web/20180727024332/https://digitalcivics.io/apply/ (accessed: 6 July 2022).

'Open rights group: Who we are' (no date). available at: https://www.openrightsgroup.org/who-we-are/ (accessed: 16 June 2022).

Organisation for Economic Co-operation and Development (1980) *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. available at: https://www.oecd.org/digital/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm.

Orphanides, K. G. (2021) 'The BBC's radical new data plan takes aim at netflix', *Wired UK*. available at: https://www.wired.co.uk/article/bbc-data-personalisation.

Owen, L. H. (2012) 'Betaworks' findings shifts to web clipping, as amazon bans kindle clips', *GigaOm*. available at: https://web.archive.org/web/20120922222936/http://gigaom.com/2012/09/18/betaworks-findings-pivots-as-amazon-bans-kindle-clips/.

P., R. (2021) 'Web scraping extensions: The easy gateway to web data', *Medium*. available at: https://raluca-p.medium.com/web-scraping-extensions-the-easy-gateway-to-web-data-40e8592e13bf.

Pansardi, P. (2012) 'Power to and power over: Two distinct concepts of power?', *Journal of Political Power*, 5(1), pp. 73–89. doi: 10.1080/2158379X.2012.658278.

Papert, S. (1980) 'Mindstorms: children, computers, and powerful ideas'. Basic Books, Inc.

Parr, S. *et al.* (2016) 'State intervention in family life'.

Pegoraro, R. (2022) 'Facebook will soon stop tracking your location and delete your location history', *Fast Company*. available at: https://www.fastcompany.com/90750241/facebook-will-soon-stop-tracking-your-location-and-delete-your-location-history.

Perera, C. *et al.* (2017) 'Valorising the IoT databox: Creating value for everyone', *Transactions on Emerging Telecommunications Technologies*. Wiley Online Library, 28(1), p. e3125.

Perez, S. (2018) 'Facebook is shutting down Friend List Feeds'. available at: https://techcrunch.com/2018/08/09/facebook-is-shutting-down-friend-list-feeds-today/.

Peysakhovic, A. and Wu, L. (2020) 'CS 4803 / 7643: Deep learning guest lecture: Embeddings and world2vec, a guest lecture by ledell wu'. available at: https://web.archive.org/web/20211018015836/https://www.cc.gatech.edu/classes/AY2020/cs7643_spring/slides/L13_Embedding_world2vec_final_version.pdf.

Phillips, P. J. *et al.* (2020) 'Four principles of explainable artificial intelligence'. National Institute of Standards; Technology, p. 18. doi: https://doi.org/10.6028/NIST.IR.8312-draft.

Pidoux, J. *et al.* (2022) *Digipower technical reports: Understanding influence and power in the data economy*. doi: 10.5281/zenodo.6554155.

Pidoux, J. (2022) 'PersonalData.IO: A space and an integrative toolbox for individuals and collectivities addressing surveillance capitalism and wanting to take control of their personal data.' available at: https://wiki.personaldata.io/wiki/Main_Page.

Pink, S. *et al.* (2013) 'Applying the lens of sensory ethnography to sustainable HCI', *ACM Transactions on Computer-Human Interaction (TOCHI)*. ACM New York, NY, USA, 20(4), pp. 1–18.

Plantin, J. C. *et al.* (2018) 'Infrastructure studies meet platform studies in the age of google and facebook', *New Media and Society*. SAGE Publications Ltd, 20, pp. 293–310. doi: 10.1177/1461444816661553.

Pollock, R. (2011) 'Building the (Open) Data Ecosystem – Open Knowledge Foundation Blog'. available at: https://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/ (accessed: 23 July 2019).

Pór, G. (1997) 'Designing Knowledge Ecosystems for Communities of Practice', in *Advancing organizational capability via knowledge management*.

Prawat, R. and Floden, R. (1994) 'Philosophical perspectives on constructivist views of learning', *Educational Psychologist - EDUC PSYCHOL*, 29, pp. 37–48. doi: 10.1207/s15326985ep2901_4.

Precisely Editor (2022) 'Data literacy: What is it and why it matters', *Precisely*. available at: https://www.precisely.com/blog/data-integrity/data-literacy-what-it-is-and-why-it-matters.

Price Ball, M. (no date) 'About Us'. available at: https://www.openhumans.org/about/ (accessed: 31 March 2021).

'Privacy' (no date). available at: https://privacy.linkedin.com/ (accessed: 9 August 2021).

'Privacy - Apple (UK)' (no date). available at: https://www.apple.com/uk/privacy/ (accessed: 9 August 2021).

'Privacy & Terms – Google' (no date). available at: https://policies.google.com/ (accessed: 9 August 2021).

'Privacy salon & my data done right' (2022). available at: http://www.privacysalon.org/my-data-done-right-right.

Puussaar, A., Clear, A. K. and Wright, P. (2017) 'Enhancing personal informatics through social sensemaking', in *Proceedings of the 2017 CHI conference on human factors in computing systems*, pp. 6936–6942.

Quinn, P. (2021) 'Research under the GDPR–a level playing field for public and private sector research?', *Life Sciences, Society and Policy*. Springer, 17(1), pp. 1–33.

Raskin, J. (2000) *The humane interface: new directions for designing interactive systems*. Addison-Wesley Professional.

Reber, M. and Duffy, A. (2005) 'Value centred design: Understanding the nature of value', *International Conference on Engineering Design*.

'Recursive Public (Discussion Page)' (no date). available at: https://wiki.p2pfoundation.net/Recursive_Public (accessed: 16 June 2022).

Regoje, O. (2021) ‘Reddit’s disrespectful design’. available at: https://ognjen.io/reddits-disrespectful-design/.

Ricklefs, H. *et al.* (2021) ‘Stronger together: Cross service media recommendations’, *International Broadcasting Convention*. available at: https://www.ibc.org/download?ac=18659 (accessed: 25 August 2022).

Ries, E. (2011) *Wiki: The Lean Startup: How Today’s Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. Crown. available at: http://en.wikipedia.org/wiki/Lean_Startup.

Rivera-Pelayo, V. *et al.* (2012) ‘A framework for applying Quantified Self approaches to support reflective learning’, *Proceedings of the IADIS International Conference Mobile Learning 2012, ML 2012*, pp. 123–131.

Roche, M. (2011) ‘Full internet ban for sex offenders ruled unlawful’. available at: https://ukhumanrightsblog.com/2011/08/12/full-internet-ban-for-sex-offenders-ruled-unlawful/ (accessed: 23 March 2021).

Rogers, Y. (2006) ‘Moving on from Weiser’s Vision of Calm Computing: Engaging UbiComp Experiences’, *LNCS*, 4206, pp. 404–421. available at: http://www.inf.ufg.br/$\sim$vagner/courses/mobilecomputing/docs/papers/03-Rogers_Ubicomp06.pdf.

Ross, G. (2005) ‘An introduction to Tim Berners-Lee’s Semantic Web’. available at: https://www.techrepublic.com/article/an-introduction-to-tim-berners-lees-semantic-web/.

Rughiniș, R. *et al.* (2021) ‘From social netizens to data citizens: Variations of GDPR awareness in 28 european countries’, *Computer Law & Security Review*. Elsevier, 42, p. 105585.

Saha, D. and Mukherjee, A. (2003) ‘Pervasive computing: A paradigm for the 21st century’. IEEE. doi: 10.1109/MC.2003.1185214.

Sailaja, N. *et al.* (2021) 'Human-data interaction through design: An explorative step from theory to practice using design as a vehicle', in *Extended abstracts of the 2021 CHI conference on human factors in computing systems*, pp. 1–5.

Sambra, A. V. *et al.* (2016) 'Solid: A platform for decentralized social applications based on linked data', *MIT CSAIL & Qatar Computing Research Institute, Tech. Rep.*

Sauermann, L., Bernardi, A. and Dengel, A. (2005) 'Overview and outlook on the semantic desktop', in *CEUR workshop proceedings*.

Savage, A. and Hyde, R. (2014) 'Using freedom of information requests to facilitate research', *International Journal of Social Research Methodology*. Routledge, 17(3), pp. 303–317. doi: 10.1080/13645579.2012.742280.

Schneider, H. *et al.* (2018) 'Empowerment in HCI - A survey and framework', in *Conference on human factors in computing systems - proceedings*. Association for Computing Machinery. doi: 10.1145/3173574.3173818.

Schrems, M. (2017) 'Noyb.eu - our detailed concept', *noyb.eu*. available at: https://noyb.eu/en/our-detailed-concept.

Schrock, A. R. (2016) 'Civic hacking as data activism and advocacy: A history from publicity to open government data', *New Media and Society*. SAGE Publications Ltd, 18, pp. 581–599. doi: 10.1177/1461444816629469.

Schumacher, K., Sintek, M. and Sauermann, L. (2008) 'Combining fact and document retrieval with spreading activation for semantic desktop search', in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, pp. 569–583. doi: 10.1007/978-3-540-68234-9_42.

Searls, D. (2008) 'The Intention Economy: What Happens When Customers Get Real Power'. available at: https://web.archive.org/web/20101226073246/http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/2009_03_24_lunchtalk.ppt.

Searls, D. (2012) *The intention economy: when customers take charge*. Harvard Business Press (04), pp. 50-2168-50-2168. doi: 10.5860/choice.50-2168.

Seligman, C. and Darley, J. M. (1976) 'Feedback as a means of decreasing residential energy consumption', *Journal of Applied Psychology*, 62(4), pp. 363–368. doi: 10.1037/0021-9010.62.4.363.

Shah, C. and Bender, E. M. (2022) 'Situating search', in *CHIIR 2022 - Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*. Association for Computing Machinery, Inc, pp. 221–232. doi: 10.1145/3498366.3505816.

Shannon, C. E. (1948) 'A mathematical theory of communication', *The Bell system technical journal*. Nokia Bell Labs, 27(3), pp. 379–423.

'ShapeRepo: Make your apps interoperable' (2022). available at: https://shaperepo.com/.

Sharp, E. (2021) 'Personal data stores: Building and trialling trusted data services - BBC r&d', *BBC R&D Blog*. available at: https://www.bbc.co.uk/rd/blog/2021-09-personal-data-store-research.

Sharp, E. and Bowyer, A. (2022) 'Building trusted data services and capabilities'. available at: https://paper.dropbox.com/doc/Building-trusted-data-services-and-capabilities-Us49Ek0nex7yClKughPN4 (accessed: 18 August 2022).

Shemtob, L. (2021) 'GPDPR versus GDPR', *British Journal of General Practice*. British Journal of General Practice, 71(710), pp. 419–419.

Shilton, K. (2011) 'Participatory Personal Data: An Emerging Research Challenge for the Information Sciences', *Advances in Information Science*.

Shipman, F. M. (III). and Marshall, C. C. (1999) 'Formality Considered Harmful : Experiences , Emerging Themes , and Directions on the Use of Formal Representations in Interactive Systems', pp. 333–352.

Shneiderman, B. (1996) *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualisations*.

Siegal, J. (2022) 'Twitter is killing TweetDeck for mac on july 1st and everyone's angry'. BGR. available at: https://bgr.com/tech/twitter-is-killing-tweetdeck-for-mac-on-july-1st-and-everyones-angry/.

Siegel, D. (2009) *Pull: The power of the semantic web to transform your business*. Penguin.

Siegel, D. (2010) 'Personal Data Locker Vision Video'. available at: https://vimeo.com/14061238.

Siegler, M. G. (2011) 'Facebook Unveils Timeline: The Story Of Your Life On A Single Page'. available at: https://techcrunch.com/2011/09/22/facebook-timeline/ (accessed: 21 March 2021).

Silver, J. (2014) 'Shamed on twitter, corporations do an about-face', *ArsTechnica*. available at: Shamed on Twitter, corporations do an about-face.

Simon, H. A. (1971) 'Designing Organizations for an Information-Rich World', *Computers, Communication, and the Public Interest.*, pp. 37–72.

Simon, H. A. and Newell, A. (1958) 'Heuristic Problem Solving: The next advance in operations research'. doi: 10.1057/978-1-349-94848-2_792-1.

Sisco, J. (2022) 'New internal documents show tech giants pushing out competitors', *Politico*. available at: https://www.politico.com/news/2022/07/19/documents-antitrust-case-google-amazon-00046522.

Sitra (1967). available at: https://www.sitra.fi/en/topics/strategy-2/#what-is-sitra (accessed: 18 August 2022).

Sitra (2018) 'Sitra's fair data economy theme: What is it about?' available at: https://www.sitra.fi/en/themes/fair-data-economy/#what-is-it-about (accessed: 18 August 2022).

Slater, C. L. (2003) 'Generativity versus stagnation: An elaboration of erikson's adult stage of human development', *Journal of Adult development*. Springer Nature BV, 10(1), pp. 53–65.

Smith, R. C., Bossen, C. and Kanstrup, A. M. (2017) 'Participatory design in an era of participation', *CoDesign*. Taylor & Francis, 13(2), pp. 65–69. doi: 10.1080/15710882.2017.1310466.

Soja, E. (2015) 'Supporting Healthcare of the Elderly through ICT: Socio-demographic Conditions and Digital Inclusion', in *Knowledge economy society - challenges and development trends of modern economy, finance and information technology.*, pp. 279–290.

Solanki, M. R. (2021) 'Solid: A web system to restore the control of users' personal data', in *ICT systems and sustainability: Proceedings of ICT4SD 2020, volume 1*. Springer, pp. 257–267.

Solon, O. (2012) 'How much data did facebook have on one man? 1,200 pages of data in 57 categories', *Wired*. available at: https://www.wired.co.uk/article/privacy-versus-facebook.

Solove, D. J. (2008) 'Understanding privacy'. Harvard University Press, May.

Spagnuelo, D., Ferreira, A. and Lenzini, G. (2019) 'Accomplishing Transparency within the General Data Protection Regulation.', in *ICISSP*, pp. 114–125.

Spector, P. E. (1982) 'Behavior in organizations as a function of employee's locus of control', *Psychological Bulletin*, 91(3), pp. 482–497. doi: 10.1037/0033-2909.91.3.482.

Spencer, D. and Warfel, T. (2004) 'Card sorting: A definitive guide', *Boxes and arrows*, 2(2004), pp. 1–23.

Spiekermann, S. and Korunovska, J. (2017) 'Towards a value theory for personal data', *Journal of Information Technology*, 32(1), pp. 62–84. doi: 10.1057/jit.2016.4.

Spinuzzi, C. (2005) 'The methodology of participatory design', *Technical Communication*. Society for Technical Communication, 52, pp. 163–174.

Star, S. L. (1989) 'The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving', in *Distributed artificial intelligence*. Elsevier, pp. 37–54. doi: 10.1016/b978-1-55860-092-8.50006-x.

Star, S. L. (2010) 'This is not a boundary object: Reflections on the origin of a concept', *Science Technology and Human Values*, 35(5), pp. 601–617. doi: 10.1177/0162243910377624.

Steinberg, S. G. (1997) 'Lifestreams', *Wired*. available at: https://www.wired.com/1997/02/lifestreams/.

Steinfeld, N. (2016) '"I agree to the terms and conditions":(how) do users read privacy policies online? An eye-tracking experiment', *Computers in human behavior*. Elsevier, 55, pp. 992–1000.

Steyaert, J. and Gould, N. (2009) 'Social work and the changing face of the digital divide', *British Journal of Social Work*, 39(4), pp. 740–753. doi: 10.1093/bjsw/bcp022.

Storni, C. (2014) 'The problem of De-sign as conjuring: Empowerment-in-use and the politics of seams', *Proceedings of the 13th Participatory Design Conference on Research Papers - PDC '14*. New York, New York, USA: ACM Press, pp. 161–170. doi: 10.1145/2661435.2661436.

Strohmayer, A. *et al.* (2021) *Trust and Abusability Toolkit: Centering Safety in Human-Data Interactions*.

Svantesson, D. J. B. (2012) 'Data protection in cloud computing–the swedish perspective', *Computer Law & Security Review*. Elsevier, 28(4), pp. 476–480.

Svensson, S. *et al.* (2018) 'The emerging 'right to repair'legislation in the EU and the US', *Proceedings from Going Green–Care Innovation*, pp. 27–29.

Symons, T. *et al.* (2017) 'Me, my data and I: The future of the personal data economy', *DECODE (DEcentralised Citizen Owned Data Ecosystems) Report*, (732546), p. 88. available at: https://media.nesta.org.uk/documents/decode-02.pdf%0Ahttps://decodeproject.eu/publications/me-my-data-and-ithe-future-personal-data-economy%0Ahttps://media.nesta.org.uk/documents/decode-02.pdf%0Ahttps://decodeproject.eu/publications/me-my-data-and-ithe-fu.

'Tap my data' (2022). available at: https://tapmydata.com/.

Taplin, D. H. and Clark, H. (2012) 'Theory of change basics: A primer on theory of change', *ActKnowledge*, p. 9. available at: http://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/ToCBasics.pdf.

Tauberer, J. (2014) 'Civic hacking', *Open Government Data: The Book*. available at: https://opengovdata.io/2014/civic-hacking/.

Taylor, A. S. *et al.* (2015) 'Data-in-place: Thinking through the relations between data and community', in *Conference on human factors in computing systems - proceedings*, pp. 2863–2872. doi: 10.1145/2702123.2702558.

Taylor, L. (2017) 'What is data justice? The case for connecting digital rights and freedoms globally', *Big Data and Society*, 4(2). doi: 10.1177/2053951717736335.

Taylor, L., Floridi, L. and Van der Sloot, B. (2016) *Group privacy: New challenges of data technologies*. Springer.

Teevan, J. *et al.* (2004) 'The perfect search engine is not enough: A study of orienteering behavior in directed search', in *Conference on human factors in computing systems - proceedings*, pp. 415–422. available at: http://people.csail.mit.edu/teevan/work/publications/papers/chi04.pdf.

Teevan, J. B. (2001) 'Displaying dynamic information', in *Conference on human factors in computing systems - proceedings*, pp. 417–418. doi: 10.1145/634067.634311.

Televiciute, J. (2020) 'What is faceted search and how does it improve ecommerce UX?' Searchanise. available at: https://searchanise.io/blog/what-is-faceted-search/.

Terdiman, D. (2008) 'Using tags to improve the Flickr experience'. available at: https://www.cnet.com/news/using-tags-to-improve-the-flickr-experience/.

Tett, G. (2022) '"Right to repair" campaign forces rethink by big tech', *Financial Times*. available at: https://www.ft.com/content/aabe2aee-cd2b-42e4-9a0b-51f838da89db.

'The citizens - about us' (2020). available at: https://the-citizens.com/about-us/.

The European Parliament and the Council of the European Union (2016b) 'Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data', pp. 16–32. available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=ES.

The European Parliament and the Council of the European Union (2016a) 'Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data'. available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=ES.

'The GDPR: Does it Benefit Consumers in Any Practical Way?' (2020). Atebits.com. available at: https://www.atebits.com/the-gdpr-does-it-benefit-consumers-in-any-practical-way/.

'The personal computer revolution' (no date) in *Britannica*. available at: https://www.britannica.com/technology/computer/The-personal-computer-revolution.

Thompson, R. (2011) 'Radicalization and the use of social media', *Journal of strategic security*. JSTOR, 4, pp. 167–190.

Tim Davie (BBC Director-General), Richard Sharp (BBC Chairman) and Clare Sumner (Director of Policy) (2022) 'House of lords communications and digital select committee', *parliamentlive.tv*. available at: https://parliamentlive.tv/event/index/7d249bcf-78e9-447b-907c-81df72b87542?in=15:01:35 (accessed: 25 August 2022).

Timely (2020) 'The attention economy: what it is, what it's doing to you'. available at: https://memory.ai/timely-blog/the-attention-economy.

Tisné, M. (2020) 'The data delusion: Protecting individual data isn't enough when the harm is collective', *Luminate, July*.

Toonders, J. (2014) 'Data Is the New Oil of the Digital Economy'. available at: https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/.

Tregeagle, S. and Darcy, M. (2008) 'Child welfare and information and communication technology: Today's challenge', *British Journal of Social Work*, 38(8), pp. 1481–1498. doi: 10.1093/bjsw/bcm048.

Tufecki, Z. (2019) 'We are tenants on our own devices', *Wired*. available at: https://www.wired.com/story/right-to-repair-tenants-on-our-own-devices/.

Tufekci, Z. (2017) 'We're building a dystopia just to make people click on ads', *TED*. TED. available at: https://www.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads.

Tunikova, O. (2018) 'Are We Consuming Too Much Information?' available at: https://medium.com/@tunikova_k/are-we-consuming-too-much-information-b68f62500089 (accessed: 23 March 2021).

Turitzin, M. and Patel, J. (2010) 'Workflowy'. available at: https://www.workflowy.com/features/.

Udaptor (2021) 'Udaptor Assistant'. available at: https://udaptor.io/assistant.html (accessed: 23 August 2021).

UK Health Security Agency (2020) 'Mixed methods study', *gov.uk*. available at: https://www.gov.uk/guidance/mixed-methods-study.

Urquhart, L., Lodge, T. and Crabtree, A. (2019) 'Demonstrably doing accountability in the internet of things', *International Journal of Law and Information Technology*. Oxford University Press, 27(1), pp. 1–27.

Urquhart, L., Sailaja, N. and McAuley, D. (2018) 'Realising the right to data portability for the domestic internet of things', *Personal and Ubiquitous Computing*. Springer, 22, pp. 317–332.

US Department of Health Education and Welfare (1973) 'Records Computers and the Rights of Citizens'.

Varghes, S. (2019) 'Pokémon Go was a warning about the rise of surveillance capitalism'. available at: https://www.wired.co.uk/article/the-age-of-surveillance-capitalism-facebook-shoshana-zuboff.

Various Authors (2018) 'Our Digital Lives', in *TED talks*. TED. available at: https://www.ted.com/playlists/26/our_digital_lives.

Various Authors (2022) 'ARIA accessibility tags', *Mozilla Developer Network*. available at: https://developer.mozilla.org/en-US/docs/Web/Accessibility/ARIA.

Véliz, C. (2020) *Privacy Is Power: Why and How You Should Take Back Control of Your Data*. Transworld Publishers Limited, p. 224. available at: https://b-ok.lat/book/11000161/b53144.

Verborgh, R. (2017) 'Paradigm shifts for the decentralized web'. available at: https://ruben.verborgh.org/blog/2017/12/20/paradigm-shifts-for-the-decentralized-web/.

Vlachokyriakos, V. *et al.* (2016) 'Digital civics: Citizen empowerment with and through technology', *Conference on Human Factors in Computing Systems - Proceedings*, 07-12-May-, pp. 1096–1099. doi: 10.1145/2851581.2886436.

Wagner, A. (2012) 'Is internet access a human right?' available at: https://www.theguardian.com/law/2012/jan/11/is-internet-access-a-human-right (accessed: 23 March 2021).

Wakkary, R., Oogjes, D. and Behzad, A. (2022) 'Two years or more of co-speculation: Polylogues of philosophers, designers, and a tilting bowl', *ACM Transactions on Computer-Human Interaction*. ACM New York, NY, 29(5), pp. 1–44.

Walby, K. and Larsen, M. (2012) 'Access to information and freedom of information requests: Neglected means of data production in the social sciences', *Qualitative Inquiry*, 18, pp. 31–42. doi: 10.1177/1077800411427844.

Waldman, A. E. (2020) 'Data Protection by Design ? A Critique of Article 25 of the GDPR', 1239(2019), pp. 147–168.

Wallace, D. P. (2007) *Knowledge management: Historical and cross-disciplinary themes*. Libraries unlimited.

Watzman, N. (2021) 'The political ads facebook won't show you', *Medium*. available at: https://medium.com/cybersecurity-for-democracy/the-political-ads-facebook-wont-show-you-e0d6181bca25.

Weiser, M. (1991) 'The computer for the 21st century', *Scientific American*, 265(3), pp. 94–105. doi: 10.1145/329124.329126.

Weiser, M. (1994) 'Creating the invisible interface: (Invited talk)', in *Proceedings of the 7th annual ACM symposium on user interface software and technology*, p. 1.

Weiser, M. and Brown, J. S. (1996) 'The coming age of calm technology', *Beyond Calculation: The Next Fifty Years of Computing*. available at: http://link.springer.com/content/pdf/10.1007/978-1-4612-0685-9_6.pdf.

Weiser, M. and Brown, J. S. (1997) 'The coming age of calm technology', in *Beyond calculation*. Springer, pp. 75–85.

Wellisch, H. H. (1996) *Abstracting, indexing, classification, thesaurus construction: A glossary*. American Society of Indexers.

'What is text mining and content analytics?' (2022). OpenText. available at: https://www.opentext.com/products-and-solutions/products/ai-and-analytics/opentext-magellan/magellan-text-mining.

White, S. *et al.* (2010) 'When policy o'erleaps itself: The "tragic tale" of the integrated children's system', *Critical Social Policy*, 30(3), pp. 405–429. doi: 10.1177/0261018310367675.

Whittaker, S. and Hirschberg, J. (2001) 'The Character, Value, and Management of Personal Paper Archives', *ACM Transactions on Computer-Human Interaction*, 8(2), pp. 150–170. doi: 10.1145/376929.376932.

'Whose data is it anyway?' (2019). 04: UBDI. available at: https://www.ubdi.com/blog/whose-data-is-it-anyway (accessed: 31 March 2021).

Wiki.personaldata.io (no date) 'Subject Access Request Template'. available at: https://wiki.personaldata.io/wiki/Template:Access (accessed: 21 September 2021).

Williams, H. *et al.* (2015) 'Dynamic consent: a possible solution to improve patient confidence and trust in how electronic patient records are used in medical research.', *JMIR medical informatics*. JMIR Publications Inc., 3(1), p. e3. doi: 10.2196/medinform.3525.

Wilson, L., Wilson, R. and Martin, M. (2020) *Health and Care Practitioner Insights: Understanding Information Sharing in Constellations of Care - Report on Amy's Page*

workshop series. Great North Care Record. available at: https://www.greatnorthcarerecord.org.uk.

Wilson, R. *et al.* (2011) 'Re-Mixing Digital Economies in the Voluntary Community Sector? Governing Identity Information and Information Sharing in the Mixed Economy of Care for Children and Young People*', *Social Policy and Society*. Cambridge University Press, 10(3), pp. 379–391. doi: 10.1017/s1474746411000108.

Windeyer, R. C. (2021) *Black box exposures: Enriching public engagement with human-data relations through intermedial performance strategies*.

'WinFS' (no date). available at: https://en.wikipedia.org/wiki/WinFS.

Wong, J. and Henderson, T. (2018) 'How Portable is Portable ? Exercising the GDPR ' s Right to Data Portability', *Acm*, pp. 911–920.

Woods, B. (2022) 'BBC wages war on online echo chambers with "unbiased" tech', *The Telegraph*. available at: https://www.telegraph.co.uk/business/2022/06/09/bbc-wages-war-online-echo-chambers-unbiased-tech/ (accessed: 25 August 2022).

Woolgar, S. (2014) 'Configuring the User: The Case of Usability Trials', *The Sociological Review*, 38(1_suppl), pp. 58–99. doi: 10.1111/j.1467-954x.1990.tb03349.x.

'Worker info exchange' (2022). available at: https://www.workerinfoexchange.org/.

Wright, P. and McCarthy, J. (2008) 'Empathy and experience in HCI', *Conference on Human Factors in Computing Systems - Proceedings*, pp. 637–646. doi: 10.1145/1357054.1357156.

Xie, A., Ho, J. C. F. and Wang, S. J. (2021) 'Data City: Leveraging Data Embodiment Towards Building the Sense of Data Ownership', pp. 365–378. doi: 10.1007/978-3-030-73426-8_22.

Zanfir, G. (2015) 'Tracing the right to be forgotten in the short history of data protection law: The "new clothes" of an old right', *Reforming European Data Protection Law*. Springer, pp. 227–249.

Zichichi, M., Ferretti, S. and D'Angelo, G. (2020) 'On the Efficiency of Decentralized File Storage for Personal Information Management Systems'. available at: https://arxiv.org/abs/2007.03505.

Zins, C. (2015) ‘What is the meaning of "data", "information", and "knowledge"?’, *Institute of Knowledge Sharing*, 3(1).

Ziogas, G. (2020) ‘The Inventor of the World Wide Web Says the Internet Is Broken’. available at: https://medium.com/digital-diplomacy/the-inventor-of-the-world-wide-web-says-the-internet-is-broken-fbce1c8bf6cf.

Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile. available at: https://books.google.co.uk/books?id=W7ZEDgAAQBAJ.

Zuckerman, E. (2021) *Mistrust: Why Losing Faith In Institutions Provides The Tools To Transform Them*. New York, NY, USA: W. W. Norton & Company, pp. 1–3. doi: 10.1017/ipo.2021.30.

Zúñiga, H. de, Garcia-Perdomo, V. and McGregor, S. C. (2015) ‘What is second screening? Exploring motivations of second screen use and its effect on online political participation’, *Journal of communication*. Oxford University Press, 65(5), pp. 793–815.

Additional Reference Information

ARI2. Additional Reference Information for Chapter 2

ARI2.1 Data Protection Terminology and a Legal Definition of Personal Data

From the GDPR (Hoofnagle, Sloot and Borgesius, 2019) and its antecedents, a number of concepts have been established which are relevant to this thesis, specifically (Information Commissioner's Office, 2014; The European Parliament and the Council of the European Union, 2016b):

- *Personal data* is legally defined as any information relating to an identifiable natural person - one who can be identified directly or indirectly by reference to an identifier such as a name, identification number or location or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person.
- The *data subject* is the identified individual, living or deceased, who the personal data relates to.
- A *data controller* is the legal entity (company, public authority, agency, individual or other body) which collects or stores personal data about an individual and determines the means and purposes for which it is processed. Liability for data protection compliance rests with the data controller.
- A *data processor* is a legal entity (company, public authority, agency, individual or other body) which deals with personal data as instructed by a controller for specific purposes and services offered to the controller that involve personal data processing.
- *Personal data processing* refers to any manual or automated handling of digital or analogue data including collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.
- A *Subject access request* is the right to a copy of your personal data.

- *Data portability* is the right to receive a copy of all stored data about you, not just that which you provided, in an accessible and machine-readable format such as a CSV file, so that you can transport it to another service or make use of it.

The terms *Subject Access Request* and *Data portability* are used in Case Study Two, and referenced also in Chapter 7.

For simplicity, this thesis uses everyday layperson-friendly terms rather than the legal terms defined in this section. Data subjects are referred to simply as *individuals* and both data controllers and data processors as *data holders*, because for this thesis, focusing as it does on the individual perspective, there is no need to draw a distinction between data controllers and data processors.

ARI3. Additional Reference Information for Chapter 3

ARI3.1 The Private Data Viewing Monitor

By removing the filter layer on an old monitor and modifying cinema IMAX glasses, a monitor was created that only allowed viewing by the holder of the viewing glasses, which would be ideal for interviewing someone about their data while respecting privacy. Face to face interviewing had to be abandoned due to COVID-19, so this technique was sadly never used in practice.



Figure ARI3.1: Private Data Viewing Monitor with Viewing Glasses

ARI4. Additional Reference Information for Chapter 4

ARI4.1 Family Civic Data Categories

The table below illustrates the types of *family civic data* identified in the pilot study [4.2.2; Bowyer *et al.* (2018); Appendix A], and referenced in Case Study One [4.3.1].

Category	**Type of data**	**Examples/Details**
Family	Personal details	Date of birth, address, telephone number.
	Relationships	Marital status, exs, step-parents, living arrangements.
	Children	Parentage, adoption, fostering, childcare.
Education	School Records	Attendance (truancy), special needs.
	Academic Results	SATs, reports, exam failures, training courses.
Welfare	Social Support	Social worker visits & notes, details of family crises, interventions, allegations.
	Welfare Benefits	Jobseeker's Allowance, child support, Disability Living Allowance, tax credits
Money/Work	Family Finances	Salary, savings, credit cards, spending, debt
	Employment	Job history, periods of unemployment, performance at work, NI, PAYE, pensions.
Civil	Housing data	Council house provision, eligibility criteria.
	Legal documents	Birth / marriage / death certificates, citizenship /immigration status, work permits.
Crime	Criminal records	Arrests, cautions, offenders' registers, prison time, speeding tickets, spent convictions.
	Court orders	Restraining orders, lawsuits, custody, ASBOs.
	Domestic Violence	Allegations made, medical records, social / legal interventions, victim support.

Category	Type of data	Examples/Details
Medical	GP records	GP's notes, prescriptions, tests, referrals.
	Hospital records	Operations, hospital stays, emergency care.
	Medical conditions	Diagnoses, diseases, allergies, blood type.
	Mental health	PTSD, breakdowns, depression, sectioning.
	Addictions	Substance abuse, gambling, rehab, crime.
Leisure[21]	Library Usage	Books/CDs borrowed, computer access.
	Sports & Health	Gym usage, class attendance.
	Shopping Habits	Loyalty cards, store & online purchases.
	Transport Data	Buses used, ANPR tracking, walking patterns.

Table ARI4.1 - Example Categories of Family Civic Data.

ARI4.2 Sentence Ranking - List of Sentences and Analysis Approach

In this section, additional details are provided on the *Sentence Ranking* exercise referenced in 4.3.5.

The sentences offered to participants across the 4 workshops were as follows:

- S1 A family's data should all be joined up and looked at together.
- S2 Any information from more than 5 years ago should be hidden from staff.
- S3 Asking families for consent to share data just once at the start is enough.
- S4 Councils should treat families like people, not records in a database.
- S5 Families don't want to be responsible for looking after their data.
- S6 Families find setting privacy preferences to be annoying and tedious.

[21] Some leisure categories (namely Shopping and Transport) were included that are not strictly civic data, as these would be useful for exploring issues around ethics. These also provided a reference point for participants to better consider the 'big data' benefits of data linking.

- S7 Families should always be able to talk to someone from the authorities about their data.
- S8 Families should have rights to see their data and how it is used.
- S9 Families will be willing to spend time checking their data is correct.
- S10 Families won't mind lots of data being collected about them if they can see it.
- S11 Families' data should be private unless they say it can be shared.
- S12 Information stored about families must be fair and accurate.
- S13 It is important for support workers to know mental health details.
- S14 Just looking at data doesn't tell you everything about a family.
- S15 Labels like 'domestic abuse' are damaging to families & hard to shake off.
- S16 Numerical scores are a good way to compare the progress families have made.
- S17 Officials should be able to see historical records about families.
- S18 Public sector officials can make good judgements just by looking at families' data.
- S19 Support workers make better decisions if they have more data about a family.
- S20 Support workers should be able to see family medical records.
- S21 The police should be able to see all of a family's data.

Where participants unanimously or mainly disagreed with a sentence, it is referenced in the inverse using a prime notation, e.g. S18', which would imply a reference to the opposite of the statement - in this case 'Public sector officials can **not** make good judgements just by looking at families' data.'

In each of the workshops, families ranked the sentences according to:

(a) whether they agreed, disagreed or were neutral on that statement, and

(b) whether or not they felt that statement was important.

This produced numerical ranking data which was analysed as follows:

- Sentence rankings were encoded on two scales. Sentences which contained a negative statement were inverted so that disagreement with them could be considered as agreement with a positive statement.

a. *Agreement*: neutral (0) -> agree (+1.0)
b. *Importance*: not important (0.0) -> important (+1.0)

- Rankings from different groups within workshops were aggregated, using mean averaging, with a weighting to ensure each workshop contributes equally regardless of attendance.

- This gave four values for each sentence, for each participant group (families only, staff only, and combined). *Variance* can be understood as 'unanimity of opinion': i.e. variance 0.0 indicates total agreement and 1.0 would indicate disagreement.

    a. *Mean agreement*
    b. *Variance of agreement*
    c. *Mean importance*
    d. *Variance of importance*.

- Prioritising variance in agreement over variance of importance, the four dimensions were reduced to three to allow a visualisation to be produced.

The resulting visualisation is shown in Figure 4.1.

ARI4.3 Storyboarding Action Cards

Drawing from the world of film production, storyboarding is a well-established technique in participatory design (Spinuzzi, 2005; Moraveji *et al.*, 2007). Usually it involves the participants drawing out a series of sketches in the form of a comic strip 'telling the story' of an interaction, encounter or activity. However, it had already been determined, both in terms of the research approach of this thesis [3.2.2], and in terms of responding to participants [4.3.5] that it would be more important to understand the interpersonal interactions between family and support worker and the actual actions performed upon or with data, rather than the mechanisms by which the data interaction would occur. Focusing on the visual aspects of information visualisation could be distracting. Therefore, I developed a novel technique for use in the phase 2 workshop: **Storyboarding Action Cards**. Each storyboard card denotes a possible action that can be carried out by a family member (yellow border), support worker (blue border) or an action performed together (green border). Each card includes a simple action summary such as 'Give Information' and

an iconographic representation of the action, along with a short description of which actor is doing what. It includes blank lines which the participant can 'fill in' to describe the specifics of this occurrence of the action.

Figure ARI4.1: Extract of Sample Scenario Storyboarding Exercise walkthrough

Based on the accumulated knowledge of Early Help processes amongst myself and SILVER project colleagues, enhanced for this purpose through consultation with a former social worker, I developed a total of 43 different cards to represent the suite of possible actions that would be interesting to track. These are grouped into eight different types of card:

- **Conversation Cards** – representing actions relating to exchanges of information in a conversation as well as discussions, decisions, and questions;
- **Consent Cards** – representing actions relating to acquiring, revoking or changing family consent to data sharing or storage;
- **Data Access Cards** – representing the searching, browsing, reading, requesting and storing of information;
- **Motivation Cards** – for representing the internal wishes of either family member or support worker;
- **Feeling Cards** – for representing the emotional state of either family member or support member (This included a blank emoji face which could be filled in as well as describing the emotion in words);
- **'Elsewhere' Cards** – for those actions performed by either actor outside of their support engagement, such as sharing information with or obtaining information from a third party;
- **Problem Cards** – to represent actions where either party experiences a problem, for example either party having an issue with information handling or content, or a disagreement between worker and family member; and
- **'Custom' Cards** – a catch-all for any remaining actions that do not fall into one of the above categories.

The intent behind the storyboarding action cards is that they serve as both a boundary object and *things to think with* (as with the Family Data Cards described in (Bowyer *et al.*, 2018)) to provoke discussion among participants. They have an additional function over the Family Data Cards, however: they can be arranged in a sequence, much like a storyboard or comic strip, and filled in, to tell the story of exactly who would do what and how in the process of a support conversation involving shared data interaction. In this way they lend themselves to model processes rather than object design. Figure ARI4.1 shows an example of three cards having been filled in and arranged in sequence to tell a simple story

of how a scenario of a worker seeking out an address following new information from the family member.

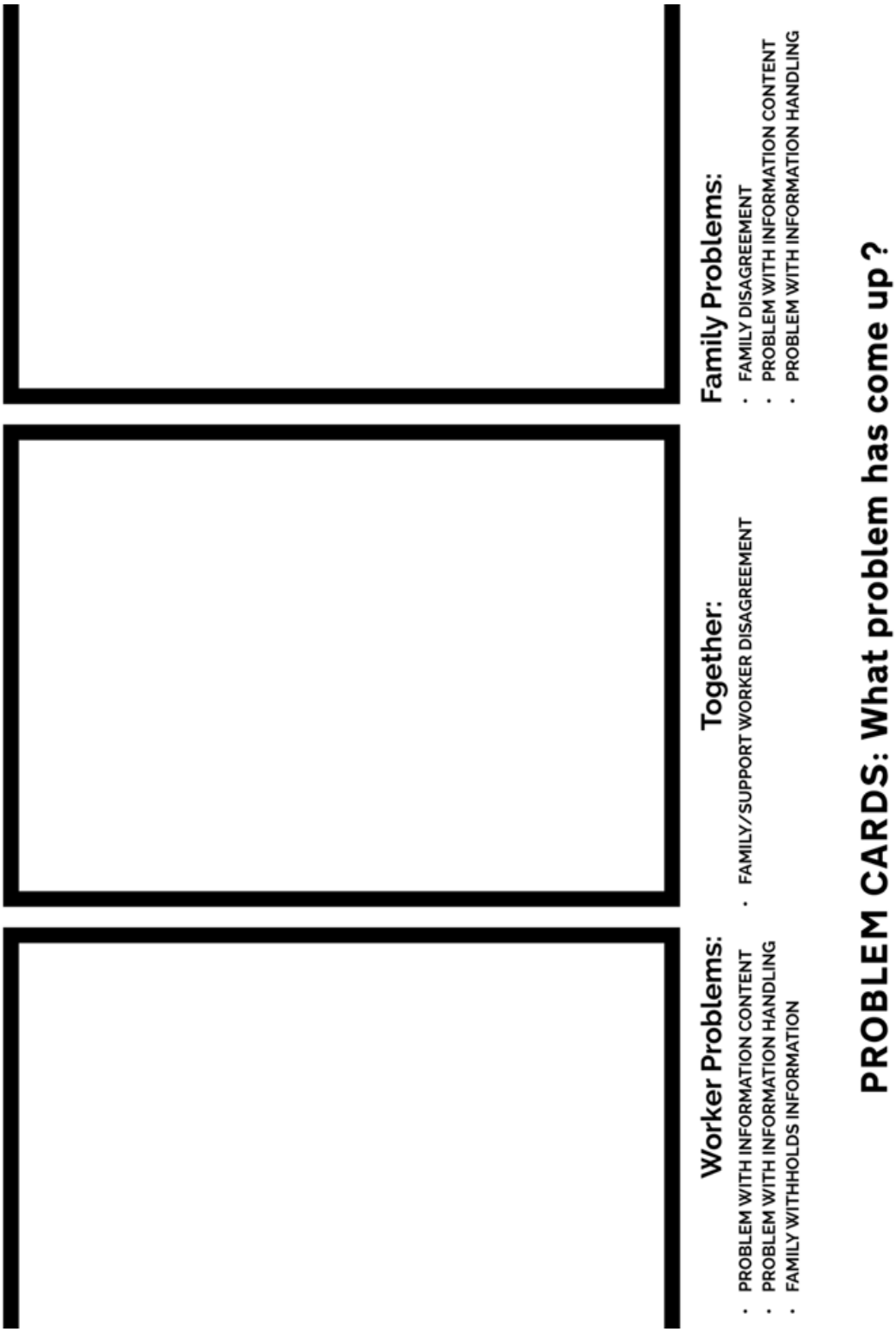


Figure ARI4.2: Example Backing Mat for Storyboard Decks

In addition to the storyboard cards, I also designed 'backing mats' for each of the eight card types. These were printed on large coloured card corresponding to each card type's backing colour, and provided areas for the 'decks' of available cards to be picked from. Each backing mat provided a separate home for family member actions, staff actions, and joint actions. Additionally, each backing mat included a summary of the available action cards of this type, and a prompt question. An example of a backing mat, in this case for Problem Cards, is shown in Figure ARI4.2.

Introduction and Practice

In order to familiarise the participants with the storyboarding action cards and the available actions, participants were first presented with an introduction to the storyboarding concept, as used in film-making and participatory design, then the card design and intended usage was explained.A very simple scenario of a family going through a breakup was used to talk through an illustrated example of how to map out the subsequent worker/parent conversation using the action cards. Then participants were invited to use the same scenario and practice mapping out the scenario themselves; however, this time they were to map out a 'problematic' version of the scenario, where things do not go so smoothly.

Scenario-Based Storyboarding Discussions

After the participants were acquainted with the cards and had practiced the storyboarding method, the main activity began, to which was allocated the majority of the time in the session. This involved each group mapping out two stories for a more substantial scenario; one version where things go smoothly and another 'negative' version where things do not go smoothly. It was highlighted to participants that the aim was to identify what would or should happen at each stage, and why.

The scenarios used for this activity by the two groups were (a) a new scenario where a couple is looking at their historical medical records (which contain various matters of concern such as missed appointments and historical mental health issues) and (b) a 'labels and judgement' scenario that had been used in the phase 1 workshops. Additional scenarios were prepared but not used. The layouts of the completed storyboards were photographed for reference, and to provide context during analysis of the discussion transcripts.

For a completed storyboard layout example, see Figure 3.11.

ARI4.4 Notation for Quotations in Chapter 4

Quotations included in section 4.4 are references using the following notation:

- FQnn = Family Quote - a quote from the families-only workshop (A)
- SQnn = Staff Quote - a quote from a staff-only workshop (B)
- CQnn = Combined Quote - a quote from the combined workshop (C).
- Sn = Sentence n - a sentence from the *Sentence Ranking* exercise, see ARI4.2.

The number after FQ/CQ/SQ provides a unique identifier for each quote. Individual speakers are identified only by their role. Within each quote, or in brackets afterwards, the speakers are identified as *Worker*, *Parent*, *Child*, or *Researcher*.

Most quotes and conversation extracts are directly embedded into section 4.4. All other quotes referenced in the text (excluded for reasons of space and flow) are included in ARI4.5.

ARI4.5 Additional Participant Quotations

The majority of quotations and conversation extracts in Case Study Two are embedded inline throughout section 4.4. The following quotes were referenced in the text but excluded for reasons of space and flow. The following list also includes some quotes or extracts which were abridged in the Chapter body but are included in full here.

Quotes from Families-Only Workshop [A]

FQ1 [Researcher(A), Parent(B) & Daughter(C)]
A: "So [you think that she should be able to be] selective about the things she wants her worker to know and leave out things that she doesn't?"
B: "Yes, like only her mental health and what tablets she's on and things."
C, talking to B: "It sounds like you." […]
B: "If she trusted her worker, I think she'd tell her herself though."
A: "Do you think that makes a big difference?"
B: "I had a worker and my daughter didn't like her and it made it really difficult when she came out. But she likes the new one."

C: “I don’t.”

B: “Why?”

C: “She’s annoying.”

A: “So do you think the relationship makes a difference to how much you tell?”

C: “Yes. Because if you don’t like them, why should you tell them?”

FQ3 [Researcher (A) & Parent (B)]

A: “What do you think could be done? What would help [this family] feel a bit happier?”

B: “Give them a one-to-one support worker who they can build up a trust and understanding where you feel like they’re not going to share your information. I don’t know, maybe come up with a computer thing so you [the family] know what they’re [the workers] putting in or maybe sign paperwork [to give your approval].””

FQ6 [Parents]

A: “It’s so hard because we’ve all done things in our past […]”

B: “I think for him to see [old medical records] the doctor should have requested it, it shouldn’t just be there for him to see. I don’t know, if he was going for some mental health problems or something and then [he can] look back… […] It should be like you have to request to look at that data. I know when I’ve been to the doctors and they actually go into a different part of the system to find my old records, which I think is a bit bad. It shouldn’t just be there.”

FQ9 [Child]

“I [designed] a graph to show how you are feeling day by day.”

FQ11A [Parent]

[discussing the sentence ‘Numerical scores are a good way to judge a family’s progress’]

“No, I disagree, because just anybody can tick any numbers. You could have a good day, you could have a bad day.”

FQ11B [Researcher (A) & Parents (B, C, D)]

A: “Do people have a right to know [past incidents with police]?”

B: “Not really. The past is the past, isn’t it?”

C: “No, because…”

D: “You shouldn’t be judged on your past, but I think it should be there [accessible in the

data] because I think at the end of the day, you can fall back into old ways. The thing is, if you're putting a child at risk or a person at risk, I think you [the worker] need to know everything, don't you?"

FQ12 [Parent]
"[The parent could] countersign. [The worker would] say, 'I feel that we've talked about this today so I'm going to write that down. I'm going to show you. Can you sign and me sign if you're happy and I'm going to share this.' That's a bit different [better]".

FQ15 [Parent]
"You would think that it would help with your benefits, [that] you wouldn't mind sharing your data, would you, because they [support workers] are trying to help you. It's not like they're saying, 'Well she gets too much money,' They're not trying to cut [families'] benefits, they're trying to help [families]."

FQ16 [Parent]
"[Families need to] feel they're being involved. […] [We need to be able to] sit together and say, 'Right, that's the information I'll allow you to share. I don't want that bit shared. But this bit, because it will help me and the family […]'. Say in this [scenario] family, she might have been married before and had domestic violence so she doesn't want that bit shared, that's in the past. So it's [only] certain up-to-date information about the family [that would be shared] because this [the family suggested by the data] isn't her family." [Parent, SQ76]

FQ17 (Worker (A) & Parents (B, C))
A: "'Families don't want to be responsible for looking after their data'"?
B: "It's one of those things where …"
C: "You've enough on without all that."
A: "You just don't think about it."
C: "And if you were to think about it, would you actually do anything?"

Quotes from Staff-Only Workshop [B]

SQ3 [Workers]
A: "I think we would have to see all the data." […]
B: "If you're going out to visit a family, you don't know what you're going to."

A: “It’s about protecting ourselves as well.”
B: “Yes, we have to check for markers, potential violence, things like that.”

SQ4 [Worker]
[imagining an interface that would allow workers to see missed appointments]
“Often they can lie to you, can’t they, and say,”Well yes, I’ve been to the doctor. Yes, I’ve been to the dentist. Yes, I’ve done that and yes, I’ve done that. But then [with this] we’ve kind of got the proof.

SQ5 [Worker]
“[a benefit of having family’s data is that] families don’t have to tell the tale over and over again […] they don’t want to have to keep verbally telling everybody.”

SQ6 [Worker]
‘I had one [client] yesterday where she was nearly all fives [out of 5] because they’d made that much progress. I had to put that on. So, she saw that as a real positive […] She was like, “I don’t need your support on this, I don’t need your support on this.”’

SQ14 [Worker]
“Parents might not want certain information [shared] so it might not be on [the visible data records] anyway…”

SQ9 [Workers]
A: “Sometimes they might have been out and had a drink, had an argument but the police have been called and it’s recorded as domestic abuse.”
B: “That’s what I’m saying about it [the “domestic violence” label] being overused.”
A: “In isolation, it probably wouldn’t be classed as domestic abuse. It was just an argument.”

SQ10 [Worker]
“I think we make a lot of assumptions on information that we get about families without actually talking to them to find out why.”

SQ11 [Workers]
A: “I think you should never make a judgement on data …, that data could be wrong.”
B: “It takes individuality, working with that person as well, doesn’t it?”

SQ12 [Worker]

"It all depends on what data they've got. You take that family I worked with, if there was nothing on there about the mental health, she just looked like a really, really poor parent when in fact she's not. I think a lot of the professionals over the years have just thought that. So, I disagree [with that sentence]."

SQ13 [Researcher (A) & Workers (B, C)]

A: "Was that fair and appropriate and is that accurate in terms of [what data has been viewed]?"

B: "I think it would be fair… I think for me it's fair if it's current because…"

C: "It can only be fair if it's complete, [if] you've got all the information there."

SQ15 [Worker]

"They [families] don't like people knowing what's going on in their lives."

SQ17 [Worker]

"You often get [that] by the time they've got back from the doctors, it's ten times worse than the conversation actually was and three other things were thrown in and then they started spiralling out of control thinking about 'What has been said behind my back?' sort of thing."

SQ18 [Worker]

"It hasn't been explained property to this family that their information will be shared with other professionals. So, they've been left feeling really let down and probably quite angry about it. So, although that information does need to be shared, they [the support workers involved] ought to make the family properly aware that information will be shared."

SQ20 [Worker]

"A lot of the families we work with have got the fear that we're still social workers or attached to social workers. So, they're saying, 'I'm not going to share with you or work with you.' […] [They might] say,"You're not social services are you? We're not going to have the kids taken away?"

SQ23 [Workers]

A: I think [the medical data we can access] has to be issue-specific. I think to be able to see somebody's full medical history is not always relevant to why we're working with them.

B: I had a gran who had residency and the GP sent everything from when she was 15 [including the details of her lost pregnancies]. That wasn't relevant to what they were doing at the time with the grandchildren and residency. It's got to be relevant. […] A: Relevant to what you're doing with the family. B: Yes, relevant with the priorities and the issues what's affecting them.

SQ24 [Workers]
A: "Yes." [to the sentence 'Families' data should be private unless they say it can be shared.']
B: "Unless it's safeguarding, obviously."
A: […] "It's private, but I guess if there was a real significant need for us to know or somebody else to know that information for safeguarding…"
B: "The law will overrule."

SQ25 [Workers]
A: "Imagine somebody doing that [checking all the different data sources] though, that would be a lot of work, wouldn't it?" […]
B: "But actually, that's a really good idea to have it all in one place."

SQ26 [Workers]
A: "[In this imagined ideal system] you would press on 'Mum' and then get all the data."
B: "You'd get all the data, anything you want."
A: "Crime, financial, just the things that we get. Then everything for Dad."

SQ30 [Worker]
"I think for some parents it will be good for them to visually see it as well. […] So you're able to give them almost a visual context rather than just talking at them. Different people take information in different ways, don't they?"

SQ31 [Worker]
"I guess the things with [tables of data] is that might just be like a number or a percentage… whereas [using a pie chart or graph] is actually giving some context."

SQ32 [Workers]
A: "A lot of the time they say, 'I'm not going to get into any more trouble,' [but with the ability to show them data] you can say,"But if you did, this could happen."

B: “If you get into more bother, you’re going to go straight back down to there [acts pointing at data]. Look where you are now. If you carry on you’re going to end up up there but if you go back, if you continue to smoke that weed and smash that phone box, you’re going to go straight back down to there.”

SQ34 [Workers]
A: “[Our idea is] an app for checking that data, with graphs and charts.”
B: “That would be amazing if we just sat down with them and handed them [a tablet] and said,”We’ve just updated [our records]. Can I just check the accuracy?”

SQ35 [Researcher(A) & Worker (B)]
A: What do you think determines whether [families] do or don’t have an interest in [checking their data]?
B: I think the experiences that they’ve had […] If it’s historical to say a safeguarding, [they’ll just think] ‘we know what the process is, we know how things are kept, we’re not going to be able to do anything about it.’ [Worker & Researcher, SQ35]

SQ38 [Worker]
“Families don’t know [what] data was being collected anyway […] If they knew what data was being collected about them and why it was being collected about them, I think they would mind – but I think that regardless of the fact whether they can see it or not, a lot of families don’t know how to access it because it all comes in the small print.”

SQ39 [Worker]
“Not many families ask to see the case notes, whether it’s a social worker or whether it’s a family partner, other members of the authority or any other services. So […] even if they’ve seen the data, [I’m not sure] whether they’d be confident with everything that’s been on it.”

SQ40 [Worker]
“Some families will go, ‘Well you know that information because it’s all there somewhere.’ We’re like, ‘Yes, but we don’t want to trawl back to eight years ago.’ There’s reams and reams and reams of it [data].”

SQ41 [Worker]
“The information that we hold […] you would verbalise this as well when you go to visit the

family. But what we [imagine] is expanding that a little bit more so: explaining why we hold the information that we hold, the process of why we store data, the information that we've got."

SQ42 [Worker]
"A lot of […] families talk to us about data we've collected and not one family I've ever met has got an issue with that. We go to them and say, 'We're aware that you've got these issues going on,' and it might be antisocial behaviour or school attendance, health or a domestic violence incident and they've never said,"How on earth have you got that information?"

SQ44 [Worker]
"For me, there's so much data that's stored. For me, for a parent to understand that through a text or email but just in point form. […] The less written, the better for the parent. [What we need is] a small synopsis […] like a summary view."

SQ45 [Workers]
A: "You know when people do have difficulties in terms of reading, on the computer you [could] press the sound button and it can read it for you. […] like text to audio."
B: "[It needs to be in an] easily understandable format, taking into account the family's needs."

SQ46 [Workers]
A: "[using a data interface to convey data to families] is quite verbal, isn't it?"
B: "It is. The way you use your words, the way you use your language […] [the] husband's needs are completely different to what [the] wife's are. Her levels are really low and your levels are really high. I think that's about the way you use your words…"
A: "It's how you explain it."

SQ47 [Workers]
A: "In terms of children, [you would need to have] more pictures and it would [need to] be clearer. [… Let's write down] 'Using age appropriate information'."
B: "Yes […] so it [would be] tailored content for the individual, if the age is there it might be sensitive information."

SQ48 [Workers]
A: "[There should be] separate data for each member."
B: "So really, if you want to talk to the daughter, she's not going to see the mum or dad's data. If you're talking to the dad, he's not going to see…"
A: "Unless they get permission. So you [could] have a tick box system at the start about who can see what…"

SQ51 [Worker] "[The families would have] a little app which they can log in to and read all their information - what's recorded about themselves, they can read the consent policy, who we share the information with, who we have shared the information with. If they're not happy — this would be a read-only app for them — if they're not happy they can fire off an email to us and let us know what they disagree with or if they want their information taken down or their consent."

SQ52 [Worker] "You'd just have a different page for each one of the priorities what we work with and all the information stored under there. So our key feature would be you'd be able to have individual family members log in. That would be to prevent the child seeing what mum and dad's issues were and stuff like that if it wasn't relevant. You'd be able to select what information is visible to other family members."

SQ55 [Workers]
A: "[It'd be good to have a way to] capture young person's voice and conversations." […]
B: "Self-help buttons [would be good] so say if somebody is feeling depressed […] There is a lot of self-harm going on at the moment."

SQ56 [Worker] "[our app design] would allow [families] to record audios and then the workers can then access those transcriptions. […] There's no chat, it's just about getting their worries, if they can't sit and talk to you in a face to face, one on one conversation…"

SQ57 [Worker (A) & Researcher (B)]
A: "There's times when I've been totally stuck in terms of getting information from professionals, GP, CAMHS9, so I'll say to the family,"I need this information, can you ring and get it?"
B: "So the family point you in the right direction, so they fill in the gaps for you?" A: "Yes."

SQ58 [Workers]

A: "There's loads of things where [families] make massive improvements, it's just not recorded. [They might have] changed their diet or lifestyle. There are loads and loads of things…"

B: "But it's not recorded as data."

SQ62 [Workers]

A: 'I would be inclined to agree because they can't get away from it.'

B: 'I think it depends on how you would pass it back, really.'

A: 'Well, it would be useful in meetings to know that she'd suffered from domestic abuse.'

C: 'Yes, I can see the benefits and the downsides, yes.' B: 'Yes, so, they can shake it off but it also gets in the way.'

SQ63 [Workers]

A: "[reading sentence] 'Asking families for consent to share data just once at the start is enough.' This is what we do now but how many times, when things go wrong families say to you, 'I didn't consent to that, I didn't. That's not what you asked me at the beginning.'" […] I don't know if there should be a regular…"

B: "…like an update, because things change in their life." […]

A: "[Should] we then [have] reviews, every six weeks [or so …], say to them, 'Well let's just remind each other what share consent is for and about.'? […] Obviously it's got to be regularly done because […] circumstances change."

SQ64 [Workers]

A: "[You would] click on the feed [an imagined feed of updates concerning the family] and it would bring up if they've been in trouble."

B: "Absolutely. This [would] definitely [be] your perspective of families."

SQ65 [Workers]

A: "We would get a report through to say…"

B: "They've recorded something."

A: "Yes. Then I suppose we would follow it up […] face to face."

SQ67 [Researcher (A) & Worker (B)]

A: "So is the key point of this one, that the families have input, as well and agree on what is

put on there?"
B: "Yes, so, agree on it and then they can add their signature."

SQ72 [Worker]
"You will have parents who will say that they don't want to share because they know the consequences. One of our families, the little one, she's six, and there was a DV [= Domestic Violence] incident and her mum was like, 'Don't say anything at school.'"

SQ75 [Worker]
"[This imagined data interface] would be accessible to both worker and family member so that we can be in sync but [would be] encouraging the family to take full accountability for their own responsibilities."

SQ76 [Worker]
"Let's say dad was sexually abused when he was a child, I think that's important that we know that because dad could have mental health problems now which would be a result and we didn't know that and he didn't want to speak about it."

SQ77 [Researcher (A) & Workers (B,C)]
A: "Was that fair and appropriate and is that accurate in terms of [what data has been viewed]?"
B: "I think it would be fair… I think for me it's fair if it's current because…"
C: "It can only be fair if it's complete, you've got all the information there."

SQ78 [Worker]
"So maybe you've got groups of young people who are, I don't know, there's something going on maybe in [local park], you've got some antisocial behaviour and they might be putting on their things that they like to do it with their friends. Then we pull from that, actually you've got a group of these young people who are involved in this. Then from that you can have focus groups. So, I think [if] we all as family partners know that we've got groups of young people where they are hanging out together so instead of just being one worker, I might think,"Well actually, there's so many people in my team have got these kids so we can have a focus group."

Quotes from Combined Parents and Staff Workshop [C]

CQ1 [Worker (A) & Parents(B, C)]

A: "I think most families wouldn't think about [checking their data] until […] something happens and they go, 'Hang on a minute, that's not right.'"

B: "Yes, 'Where've you got that from?'"

C: "Yes, yes."

A: "But I think, other than that, we tend to just trust that everything that has been put down is right, don't we?"

C: "Yes."

CQ2 [Worker]

"That happens a lot, doesn't it? It does happen where information is shared and then somebody gets upset because they didn't think that level of information would be made available, even though permission had been given at the start of the plan."

CQ8 [Parents (A, D), Worker (B)& Researcher(C)]

A: if you find [a criminal record for burglary], you're looking and thinking, "God! She's gone out and committed a bloody burglary."

B: Well, it could affect your employment chances if that comes back on your DBS. But I explored it and talk about it and she said, "Well, I don't agree with that. That's not what happened." I mean, she did break in but she wasn't stealing anyone else's stuff, it was her own stuff. […] If there is breaking and entering and burglary, and no explanation of that, and no way for that person to give you an explanation …

C: It's just somebody's version of what happened?

B: Well, it is, isn't it?

D: Well, the Courts need to change what's recorded because if you broke into a house and stole a telly, that would come to the top. Whereas, something like that, which is more or less trespassing. In the eyes of any decent solicitor, it's trespassing, to get your own stuff but, technically, you've stolen your own stuff. That should be put on a scale of severity, of 1 to 5, in the circumstances. If you're homeless and you break into an empty house, is that burglary? Is that worth three years in prison? You know what I mean? [Parents, Worker & Researcher, CQ8]

CQ11 [Parent (A) & Researcher (B)]

A: "I would want to see what information is held about me but then there are people out there who aren't very confident in being able to ask or if they can't read, if they've got learning [difficulties]"

B: "What should happen for those people then?"

A: "They should be supported by whoever is around them to access it in some form or another."

B: "They need to have someone talk them through it, or something?"

A: "Yes."

CQ12 [Parent]

"I think a lot of people would like to be able to [access their data]. I think the prospect of, if you want to see your medical records […] having to make an appointment and go up and sit down and read paper records [is not something people would choose, whereas] if they were able to access it, in their own time, at their own pace [that would work better]. I'd love to see what's been written about me in my medical records, I think some of it could be quite interesting."

CQ15 [Parent]

"I think [whether support workers should be able to access mental health details] depends on how long ago it was. […] I went through a really, really rough patch […] nearly 20 years ago and I had a brief patch of about three weeks where I was really not controlling my depression and I self-harmed and made an absolute fool of myself, and I'm fine with that now but I wouldn't want people, everybody, to know about that because I wouldn't want people to jump to the conclusion — because they still do — that there's something wrong and I'm going to do it again and things like that. Because people change, and situations change."

CQ17 [Worker]

"I think most families wouldn't think about [looking at or checking their data] until […] something happens and they go, 'Hang on a minute, that's not right.'"

ARI5. Additional Reference Information for Chapter 5

ARI5.1 GDPR Data Analysis Approach

In this section, the methodology used for the analysis of data from Case Study Two is explained. The content of this appendix is identical to Appendix 3 in the Supplemental Materials of the CHI 2022 paper from this study (Bowyer, Holt, *et al.*, 2022). Case Study Two was written first as a paper and then expanded to produce Chapter 5. While the paper was co-written, Chapter 5 was written entirely by Alex Bowyer.

All coding was carried out by Alex Bowyer and Jack Holt, who followed the following process over a nine-month period, comprising at least 200 person-hours:

- **EXTRACTION AND ANALYSIS OF SEMI-QUANTITATIVE DATA**: Identifying closed question (or brief) responses that might be processable quantitatively.
- **TEXT FILE PROCESSING**: Splitting, organising, anonymising and some cleaning of auto-transcribed and time-coded text files.
- **CATEGORISATION INTO CSVs**: Categorised extraction of timecoded text sections from text files into cells of 6-topic spreadsheet, then generation of CSV files for importing into Quirkos Cloud (Daniel Turner, 2014)
- **INDUCTIVE CODING**: Importing of CSVs into Quirkos Cloud and labelling by Participant, Company, and Topic. Inductive coding of source texts, ensuring good coverage per topic and per participant.
- **REDUCTIVE CYCLES**: Reductive cycles of merging, renaming and reorganising the codes hierarchy, resulting in 10 top-level codes with hierarchies of coded texts underneath them.
- **THEME IDENTIFICATION & QUOTE EXTRACTION**: Construction of 3 paper-focussed themes using Workflowy (Turitzin and Patel, 2010) and quote gathering using the organised codes hierarchy.

Some additional detail on the stages:

1. Semi-Quantitative Data Extraction & Analysis

Prior to beginning coding the data, responses to some key closed questions from the transcripts were combined with field notes, response emails from companies forwarded by participants, sketches and tables from Interview 1/2, data from the interview 2/3 spreadsheet cells, and other data collected, and used to populate a spreadsheet that featured summaries of those responses. For example, where participants had been asked to outline their hopes for the outcomes of their GDPR data requests, these responses were recorded on the spreadsheet to be used as a resource for summarising participant hopes in a manner that could be easily quantified and referred back to. In some cases this data was analysed within the spreadsheet to produce insights, graphs and percentages. Such data was later used to support and illustrate findings from the coding process. This spreadsheet also included important information relating to each participant's GDPR process experience, such as the timeliness and completeness of their data returns, which could serve as a reference point when analysing the transcripts.

The semi-quantitative data areas captured or derived from captured data were:

- Company Response Timelines
- Power Scores
- Trust Scores
- Hopes, Goals and Imagined Uses
- Term Definitions
- For each participant + target company + data type (+ subtype in some cases):
- Provided or Not?
- Perceived Value
- Completeness
- Understandability
- Accuracy
- Useability
- Usefulness
- Meaningfulness
- Feelings about data (general, and company-specific)
- General questions (general, and company-specific)

- Best and Worst Companies (taking into account provided, completeness, understandability, accuracy, usability, usefulness)
- Sankey analysis of participant journeys

2. Text File Processing (Splitting & Recombination)

The researchers then moved on to prepare for the fully qualitative analysis. All interview audio was auto-transcribed using Zoom and Google Recorder, and then the generated text files were cleaned. Cleaning consisted of listening to sections of audio where transcription seemed inaccurate and correcting the transcripts. Due to the volume of data this cleaning was not done for all texts, only where ambiguity or typos meant it was needed for accurate coding and for quotes. Some anonymisation of source texts was also carried out at this stage and later, with a particular focus on quotes included in the chapter. The researchers used this data preparation stage as an initial means of (re)familiarising with the dataset. With reference to the structured interview schedules, the initial 33 text transcripts were split up by participant, company and topic using the labelling scheme outlined in 'Text File Labelling Strategy' below.

At the end of this process, roughly 100 'pieces' had been identified for each participant (slightly more for P11 whose interview 1 covered a broader scope and considerably less for P9 who only did interview 1).

3. Categorisation into CSVs

The pieces from stage 1 were then recombined, across all participants, into 233 source files. These 233 source files were then further grouped into 6 topics areas. (The aim of the analysis was to identify common opinions and ideas around different topics, not to explore individual participant journeys end-to-end). The six topic areas were:

- **POWER** – discussions and scoring around the power of data holding companies
- **TRUST** – discussions and scoring around participants' subjective trust in data holding companies
- **LIFE** – life sketching and annotation discussions, and 'digital life' questioning
- **HOPES & USES** – discussions around motivations, expectations, goals and hopes, and imagined uses of data

- **COMPANY-SPECIFIC** – (repeated once per target company per participant) – all discussions around the data return from a particular company
- **GENERAL** – all non-company specific discussions not captured elsewhere

This produced too many files for import into Quirkos Cloud, so once organised by topic, these six groups of files were further combined into 11 General files and 46 Company-Specific, files (with **Life** and **General** going into the General files and everything else going into **Company-Specific**). This gave 57 organised CSV files ready for use in the first coding phase.

4. Inductive Coding

The majority of the analysis took place with the use of Quirkos Cloud (Daniel Turner, 2014), a computer-assisted qualitative data analysis software (CAQDAS) package that allows for collaborative analysis by more than one researcher. The 57 files from stage 3 were imported into Quirkos Cloud, with each having a unique number. The sources in Quirkos were labelled by Participant, Company and Topic for easy search and retrieval. The researchers then collaboratively coded sections of the interview transcripts to develop and ensure a consistent approach, based on established techniques (Huberman and Miles, 2002; Braun and Clarke, 2006). Codes were identified inductively and not according to a fixed or predetermined set. Once a baseline codeset and strategy had been established, they each coded sections of interviews in parallel, regularly regrouping to discuss generated codes and any new questions or challenges arising. At first, these codes were created in an unstructured/flat state with only occasional clustering on the Quirkos interface. Due to the volume of data, not every piece of every transcript was coded, however care was taken to ensure a representative sample of views from across the participant pool was included. These were clustered into loose code-topic areas, an example is shown in the following screenshot taken approximately 6 weeks into coding:

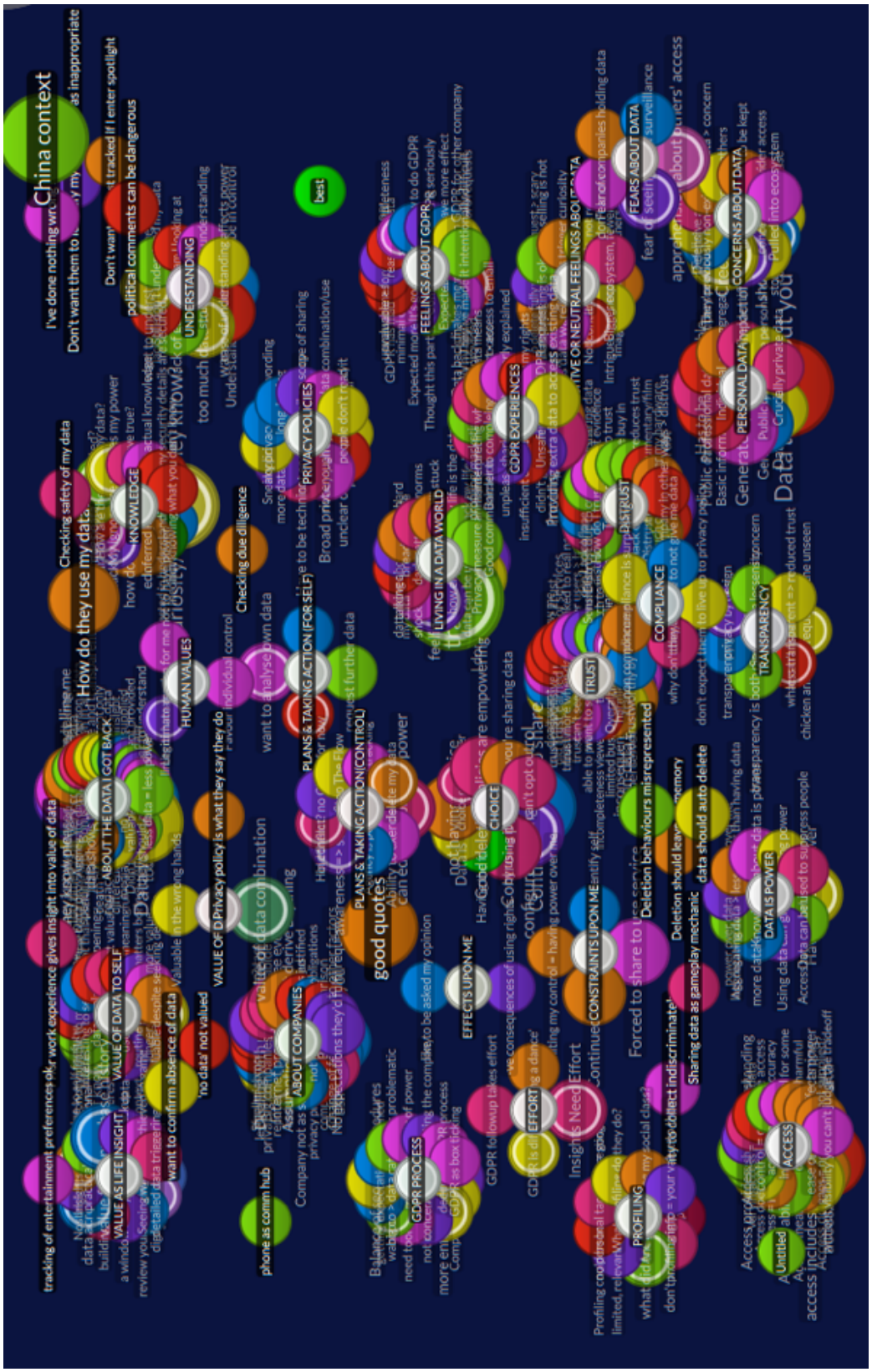


Figure ARI5.1: Screenshot from Quirkos During Coding Process

5. Reductive Cycles

As more codes were identified and structures and commonalities between them were formed, existing codes were merged or absorbed into one another and grouped together in small clusters. The researchers regularly met to discuss each other's codes according to their context and occasionally amended wording or merged concepts that were labelled differently but semantically equivalent. All codes were checked and agreed between these two researchers. Over time, the codes were iteratively structured and restructured, creating top-level thematic clusters around different research questions that held multiple layers of related codes. These clusters were then summarised with a short sentence or paragraph of text, allowing summaries to be produced at different levels of hierarchy. These summaries were kept in the Description fields of codes in Quirkos and also in external structured text-based documents. These can be seen in the following screenshot, taken 5 months into coding:

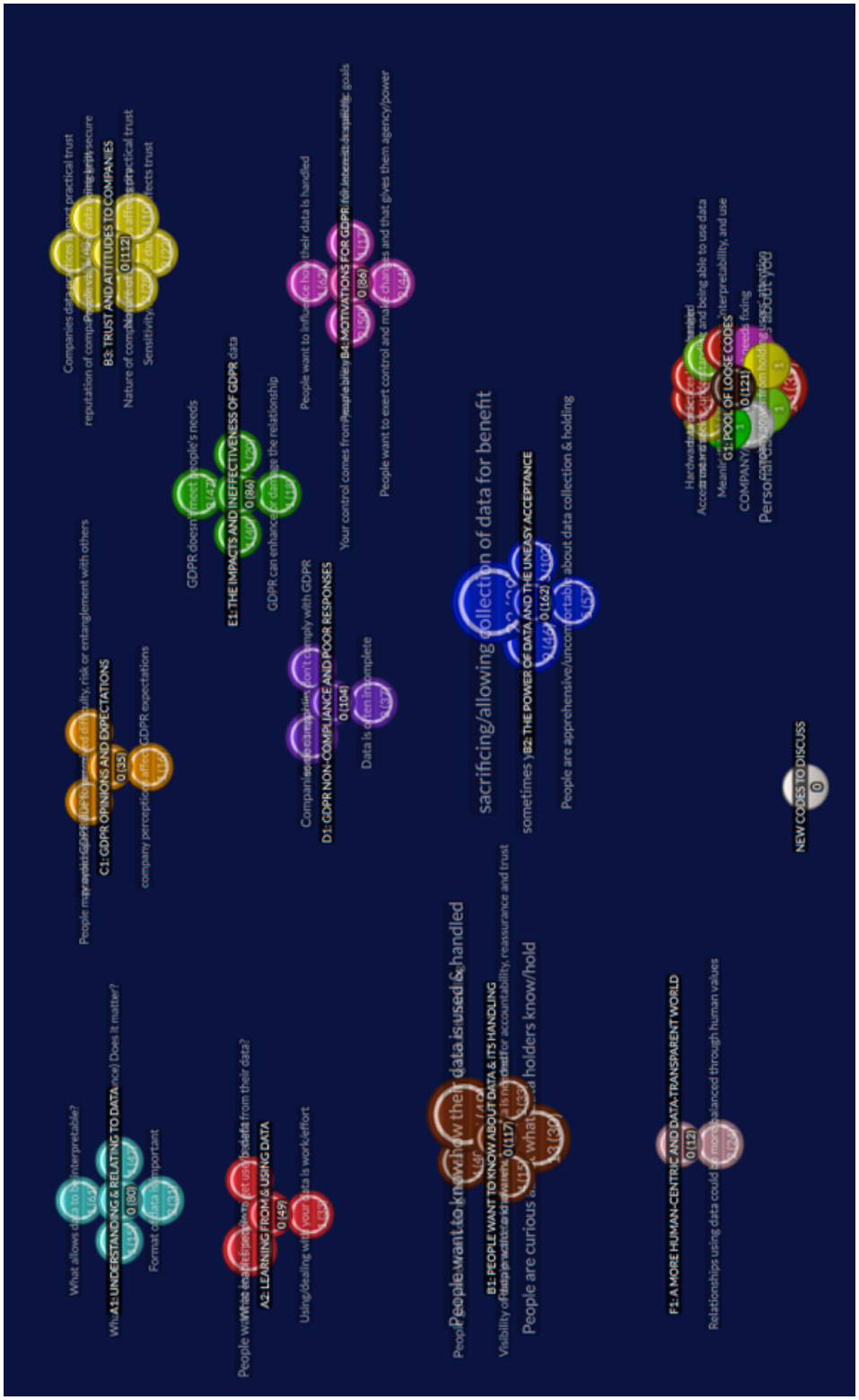


Figure ARI5.2: Screenshot from Quirkos at End of Coding Process

The above-pictured structure of the coded corpus at the end of the Quirkos Cloud phase was as follows:

- A (129 codes): What do people/need want from their data and how do they feel about data?
  - A1 (80 codes): **Understanding and Relating to Data**: People want to understand and need to be able to relate to data.
  - A2 (49 codes): **Learning from and Using Data**: People want to learn more from and do more with their data.
- B (279 codes): What do people feel about the data-centric world?
  - B1 (117 codes): **People want to know about data and its handling**: People want to know what data exists and how it is handled, so they can understand what companies do to hold them to account, and inform their choices/trust.
  - B2 (162 codes): **The Power of Data and the Uneasy Acceptance**: People accept a certain amount of data collection and use but are apprehensive and sometimes feel they have no choice, because data holding is a form of power.
  - B3 (112 codes): **Trust and Attitudes to Companies**: Trust placed in companies is influenced by both the nature and reputation of the company, as well as what data they hold and how that data is kept and handled.
  - B4 (86 codes): **Motivations for GDPR**: People want to make use of their data and influence how it is handled and see GDPR as having the potential to help them achieve this.
- C (35 codes): What do people think about GDPR?
- C1 (35 codes) **GDPR Opinions and Expectations**: People's expectations for GDPR are affected by their perception of the company and its perceived difficulty, risk and entanglement; people expect non-compliance.
- D: What is the experience of GDPR as a means to gain awareness of and access to useable and understandable data?
- D1 (104 codes) **GDPR Non-Compliance and Poor Responses**: The data returned from GDPR is often incomplete, hard to deal with, lacking explanation, or poorly formatted. Many companies are not complying.

- E: What is the experience of GDPR as a means to gain influence and achieve goals with data/What is the practical impact of GDPR?
- E1 (86 codes) **The Impacts and Ineffectiveness of GDPR**: People's interest in GDPR comes from curiosity to exert their rights or from specific questions about data handling or data use goals. GDPR rarely delivers upon on any of their goals but it does change people's outlook and affects the relationship with the data holder.
- F: How should the world change or be different?
- F1 (12 codes) **A more human-centric and data-transparent world**: People want companies to provide greater transparency and data control/agency and act in a more human manner so they can trust them.
- G: Loose/ungrouped codes (121 codes)

Total codes = 645.

6. Theme Identification & Quote Extraction

Having produced the structure above as a reduced representation of *'what the codes say'* that the participants think, the researchers used outlining tool Workflowy (Turitzin and Patel, 2010) to develop the arguments and primary narrative of the chapter into a structured three-theme-based summary of the most important items from these findings. The code hierarchy was used as source material to populate the three key themes with illustrative quotes and observed findings. An example from later in this process (around 8-9 months since Stage 1 began) is shown in the screenshot below:

Figure ARI5.3: Screenshot from Workflowy During Theme Construction

The themes are broken down in detail in 5.5 and can be summarised as:

- **Insufficient Transparency**: Organisations appear evasive over data when responding to GDPR, leaving people "in the dark" even after making GDPR requests.
- **Confusing Data**: When presented with their data, people struggle to understand it and relate it to their lives and are not able to make use of it.
- **Fragile Relationships**: Companies' data practices, and in particular their privacy policies and GDPR response handling, can be impactful to customer relationships, carrying a risk of damaging trust but also the potential to improve relations.

In all, the process from commencing data analysis to writing up thematic findings in the chapter took over 200 person-hours over a 9-month period from January to September 2020.

Text File Labelling Strategy used in Stage 2

In stage 2, text files were initially broken down into small pieces and labelled according to the following strategy:

Interview 1 (Sensitisation / Poster Display Chat)

Break into 5 parts:

- `Comp` - list of companies
- `Type` - types of data
- `DoWt` - potential uses of data ['what would you do with the data?']
- `GDPR` - GDPR
- `Motv` - motivation for taking part

Interview 1 (Main Sketch Interview)

Break down as follows:

- `SktR` - review of previous sketch interview from prior study [p11 only]
- `DPer` - definition of personal data
- `DAcc` - definition of access to data
- `DCon` - definition of control of data

- `DPow` - definition of power
- `Sket` - sketching
- `Anno` - annotation
- `SelC` - company selection
- `XXXX` - per company [use first four letters of company]
- `Powr` - power
- `Hope` – hopes
- `Uses` – uses
- `Wrap` - [Wrap up]/What happens next

Format: `NN-pXX-iX-[Comp/Type/Uses/GDPR/Motv]-[company first three letters].txt`

e.g. `01-p01-i1-Comp.txt` or `02-p01-i1-Powr-Face.txt`

Interview 2

Break down as follows:

- `XXXX` - per company [use first four letters of company name]
  - `Priv` - viewing privacy policy
  - `Powr` - power
  - `HopU` - hopes & uses
  - `Trst` - trust [p10 & p11]
  - `Pow2` - end power
  - `Trs2` - end trust
  - `Hop2` - end hopes and uses

Format: `NN-pXX-iX-[….]-[company first three letters].txt`

e.g. `01-p01-i2-Priv-Goog.txt`

Interview 3

Break down as follows:

- [intro & consent] - no need to transcribe/code

- XXXX - per company [use first four letters of company name]
  - Powr - power rating
  - Trst - trust rating
  - RPow - retro power
  - RTrs - retro trust
  - Hope - hope (for company) and uses (how well have hopes been met / how practical are the envisaged data uses
  - Data - Overall data overview
  - Prov - Data provided by you
  - Indr - Data indirectly / automatically collected
  - Derv - Data derived about you
  - Othr - Data from other sources
  - Meta – Metadata
  - GenQ - general questions about this company
  - Pow2 - end power
  - Trs2 - end trust
  - Next - what next for this company specifically
  - Genr - General topics
  - Hope - Hope (general)
  - Wrap - Wrap up questions / the future

Format: NN-pXX-iX-[….]-[company first three letters].txt

e.g. 01-p01-i3-Cred-Indr.txt or 02-p01-i3-Genr-Wrap.txt

ARI5.2 Best and Worst Companies for GDPR Handling

The quality and coverage datapoints described in 5.4.3 also allowed insights about which service providers were strongest or weakest in each category, and overall, to be drawn. This was done by tallying the 'Yes' responses for each category and overall, then dividing by the number of times that provider was selected, to avoid inflating scores for popular companies. The outcome of this analysis is shown in Table ARI5.1. The companies that fared

worst overall were those that did not return any data at all in response to a GDPR request (Sainsbury's, Freeprints, Tyne Tunnels, LinkedIn, Huawei, Bumble, LNER). It should be noted that Sainsbury's and Huawei *did* respond, claiming to hold no data for the requesting participant, though participants found this implausible, which indicates either a problem with compliance, explanation or trust. The other named companies here did not respond at all, despite at least two follow-up emails being sent to them, and despite in some cases having initially acknowledged and promised to satisfy the request.

Companies producing responses with good coverage and good quality included Niantic, Nectar and Sunderland AFC as well as to a lesser extent Natural Cycles, Revolut, Spotify, Tesco and Amazon. Facebook and Google fared well for the breadth of data returned (due in part to their download dashboards), though the quality of Google's data was found lacking across multiple categories. Last.fm (owned by CBS) fared poorly overall due to poor category coverage, despite the limited data that it did return being of high quality.

Table ARI5.1 - Best and Worst Data Holders for GDPR, according to Participants' Judgements[a]

Category / Metric	Best Companies	Worst Companies
Availablity of Data / Breadth of Data Returned	Nectar, Niantic, Sunderland AFC, Natural Cycles, *Facebook, Google, Spotify, Revolut*	Sainsbury's, Freeprints, Tyne Tunnels, LinkedIn, Huawei, Bumble, LNER, Nexus, *Three, Philips Hue (Signify), Check My File*
Completeness of Returned Data	Niantic, *Nectar, Sunderland AFC*	Sainsbury's, Tyne Tunnels, Freeprints, Nexus, LinkedIn, Huawei, Revolut, Bumble, LNER, last.fm (CBS), *Google, Tesco*
Accuracy of Returned Data	Sunderland AFC, Niantic, *Tesco, Nectar, Amazon, Natural Cycles*	Direct Line, last.fm (CBS), *Google*
Understandability of Returned Data	Nectar, Spotify, Sunderland AFC, Niantic, Apple, *last.fm (CBS)*	AirBNB, Virgin Media, *Google, Instagram, Tesco*
Meaningfulness of Returned Data	Niantic, *Spotify, Sunderland AFC, Natural Cycles, last.fm (CBS)*	AirBNB, Credit Karma, Philips Hue (Signify), Direct Line
Usability of Returned Data	Amazon, last.fm (CBS), *Facebook*	AirBNB, Credit Karma, Virgin Media, Sunderland AFC, Huawei, Three, *Google*
Usefulness of Returned Data	Amazon, Facebook, Virgin Media, Spotify, Revolut, Niantic, last.fm (CBS)	AirBNB, Credit Karma, Nectar, Direct Line, Three, *Google*
OVERALL [a]	**Niantic, *Sunderland AFC, Facebook, Spotify***	**Sainsbury's, Freeprints, Tyne Tunnels, LinkedIn, Huawei, Bumble, LNER, *last.fm (CBS), Philips Hue (Signify), Nexus***

[a] Companies were ranked according to total number of all responses in that category for this company that were "Yes" .
[b] Company names in normal text are best/worst; names in italics are second best/second worst.

ARI7. Additional Reference Information for Chapter 7

ARI7.1 A Prototype Entity Extractor and Time-Event Extractor

An additional Figure from my time on Cornmarket [See Section IV Introduction] that was not featured in the main body of the thesis is shown in Figure ARI7.1 below. This shows a screenshot from a functional prototype I produced during a hack week that allows the user to upload data retreived via GDPR or download portal from a few specific companites. It proved the concept of programmatically identifying key entities 9.4.3 and identifying time-labelled events [2.2.2; Principle 8–10.1.8] for display as life information [7.4.1] to users as part of a life interface [9.4.1].

Figure ARI7.1: Prototype Entity Extractor and Time-Event Extractor

ARI9. Additional Reference Information for Chapter 9

ARI9.1 How I compelled Spotify to improve their GDPR return

In this section, I will provide additional details of my mini-case study where I was able to get Spotify to improve the quality of their GDPR returns, as referenced in Principle 9 [10.1.9] and within Section 9.3.1.

As an avid user for several years of the music streaming service Spotify who has built up a large library of playlists, I have made a number of GDPR requests to get copies of my personal data.

When I was first given a copy of my personal data, I was returned a basic ZIP file including 12 JSON files containing playlists, search queries, account information, my last 12 months of track play history, and inferences about my musical tastes. Spotify also make an extended data download available, including technical log data, and extended play history (which covers the lifetime of my account).

I requested this extended download and received a much larger dataset with 175 JSON files, including granular details of when I had used different interface features and the precise details of every song I had ever played.

Thinking that I would like to use this data to build a view of my listening history that was not tied to the Spotify platform (in line with the idea of increasing agency by separating one's data from the service that holds it [Chapter 8]), I examined the streaming history and playlist data with this purpose in mind. What I found was that individual songs were identified only by textual strings of the title, artist and album name. This information is insufficient for a programmer's use - there is no unique identifier or Uniform Resource Indicator (URI) to uniquely identify the specific version and release of a track played. Also without such an identifier, it would not be possible to generate a thumbnail image of the track, or build functionality such as a clickable link to 'play this track in Spotify'.

This highlights a common issue that occurs with data access requests, as highlighted in 5.5.3 - there is ambiguity over whether providers should identify data in a machine-readable way (useful for programming), or in a human-readable way (to optimise understanding). In my

case, I needed both. I e-mailed Spotify back and was provided with an alternative file set which contained only Spotify Track URIs, such as `spotify:track:4cOdK2wGLETKBW3PvgPWqT`.

These met the programmer need to uniquely identify the track, but not the human need–I had no idea which artist or track each of these URIs corresponded to, as there was no human-readable text accompanying each entry.

So, I e-mailed Spotify back, making the case that my GDPR rights had not been fully satisfied, because I needed *for each play history entry*, both machine-readable ID and human-readable track title and artist name.

I sent Spotify over 30 e-mails on this matter between October 2020 and May 2021. There is little continuity of conversation between support agents, and it was hard to be escalated to the correct staff with the technical or legal expertise to assist with such nuanced questions.

However, by persistently and politely repeating my questions and not accepting 'No' for an answer, I was able to achieve a notable outcome, Spotify **changed the format of their data returns**, not just for me but **for all future customers**. Now, each item in the playback history data you get back from Spotify, every item includes textual track and artist details AND a Spotify track URI. The data can now be understood by both human and machine.

The likely interpretation here is that I successfully able to persuade their Data Protection Officers (who handle GDPR requests) the importance of returning data that is both machine-readable and human-understandable. Perhaps they also recognised the amount of work they had invested in supporting my query, and wanted to **avoid having to do such work ever again** should I or any other customer make the same request in future. This was a tiny impact, but a lasting one, and it shows that the discovery-driven activism / civic hacking approach [9.2] can have an effect in improving HDR with a target organisation.

Appendices

Appendix A. The Pilot Study: Published CHI 2018 Paper

For additional context, see 1.3.1. The paper is on the following 13 pages.

Understanding the Family Perspective on the Storage, Sharing and Handling of Family Civic Data

Alex Bowyer[1], Kyle Montague[1], Stuart Wheater[2], Ruth McGovern[3], Raghu Lingam[3], Madeline Balaam[4]

Open Lab, Newcastle upon Tyne, UK {a.bowyer2, kyle.montague} @ncl.ac.uk

Arjuna Technologies, Newcastle upon Tyne, UK stuart.wheater @arjuna.com

Institute of Health & Society, Newcastle upon Tyne, UK initial.lastname @ncl.ac.uk

Media Technology and Interaction Design, KTH Royal Institute of Technology, Stockholm, Sweden balaam@kth.se

ABSTRACT

Across social care, healthcare and public policy, enabled by the "big data" revolution (which has normalized large-scale data-based decision-making), there are moves to "join up" citizen databases to provide care workers with holistic views of families they support. In this context, questions of personal data privacy, security, access, control and (dis-)empowerment are critical considerations for system designers and policy makers alike.

To explore the family perspective on this landscape of what we call *Family Civic Data*, we carried out ethnographic interviews with four North-East families. Our design-game-based interviews were effective for engaging both adults and children to talk about the impact of this dry, technical topic on their lives. Our findings, delivered in the form of design guidelines, show support for dynamic consent: families would feel most empowered if involved in an ongoing co-operative relationship with state welfare and civic authorities through shared interaction with their data.

Author Keywords

User-centered design; personal data; social care; healthcare; big data; civic data; data privacy; ubicomp; data security; data sharing; dynamic consent; design games; family design games; boundary objects; family research; ethnographic interviews; family civic data; citizen-state relations; GDPR

ACM Classification Keywords

H.5.2. User Interfaces: User-centered design

INTRODUCTION

Government and civic authorities in the UK, as in other countries, are seeking ways to better serve the most disadvantaged sections of society. Issues including youth crime, poverty, unemployment, homelessness, truancy and domestic abuse, bring misery and upset to British families as well as significant cost to the state [18]. Civic authorities, health trusts and welfare organizations hope to create a better quality of life for vulnerable or at-risk families by using more "joined-up" data-handling practices to ensure that public sector workers who support families directly are able to see the complete picture of a family's life [44–48]. However, the sharing of data about families holds the potential to significantly disempower them as citizens, by hiding individual facts or additional context and enabling group-based profiling [4,34]. Decisions made based upon shared but erroneous data could have profound knock-on impacts across all parts of a family's life as they interact with different services, a risk that the EU aims to specifically address through legislation [49]. Families could also be exposed to identity fraud, social problems or discrimination if their private data is shared too widely. In this context, it is critical that the family perspective about how their data will potentially be stored, handled and shared is understood, and the primary aim of our research is to address this gap in understanding.

In this paper, we contribute a new understanding that families would like their *Family Civic Data*[1] to be handled as a precious commodity that must be looked after in partnership with the state through ongoing conversations and a dynamic consent model. We describe three guidelines that can inform the design of future civic data-handling systems. Additionally, we share our methodology, *Family Design Games*, which we found highly effective at motivating intergenerational participation and contribution from both adults and children to a meaningful discussion of the topic of civic data, and explain the steps needed to adapt this technique to explore other tricky research topics.



This work is licensed under a Creative Commons Attribution 4.0 International License.

CHI 2018, April 21–26, 2018, Montreal, QC, Canada © 2018.
Copyright is held by the owner/author(s).
ACM ISBN 978-1-4503-5620-6/18/04.
https://doi.org/10.1145/3173574.3173710

[1] We introduce the term *Family Civic Data* as a useful shorthand phrase. See Methodology for definition.

The Troubled Families Program (TFP)

In 2012, the UK government implemented the Troubled Families Program [18,23], having found that £9 billion per annum was being spent through public services upon just 120,000 families, due to their reduced life chances, long-term health problems and reduced productivity at school and at work. The TFP funded new initiatives to connect social care, health and welfare services, schools and police so that social workers could see the complete set of issues affecting a family in one place, with an objective of breaking repeated generational cycles of problems through better-informed and targeted interventions. At its outset, the TFP focused on direct family engagement, but soon evolved to harness large-scale data-sharing technologies. The sharing of civic data between data-holding authorities has been shown to be useful in giving citizens a better quality of service [12], and a thriving international data linkage network now exists, especially in healthcare [21]. However, projects reliant upon sharing of sensitive family data must obtain a family's consent. The NHS's failed care.data project, which closed following widespread public distrust of the data collection and sharing it required, showed that if these systems are to work they must reflect what families and citizens want and need [50]. A good example of this is gov.uk [35], an e-government system designed from a user-centric 'citizen needs' perspective, which has been shown to promote accountability [42] and increase trust and confidence in government [38].

Data Sharing, Data Handling and Consent

Consent to share data has been most thoroughly examined in the medical profession, where it offers clear, uncontroversial benefits that are easy for families to see [5]. *Informed consent* (where a patient is informed about intended data usage then given an "opt-in" or "out-out" choice), has become the norm and increases trust [36]. Patients are happy to share their data for research use [41]. However, patient data is routinely used for purposes beyond those for which consent was given, suggesting that a simple Yes/No view of consent is insufficient to cope with the rapidly changing informational needs of health and social care professionals and patients in the big data era [27,43].

New and competing models for consent have emerged. In *open consent*, patients provide broad, unrestricted consent to sharing data for unspecified purposes [20]. Here, the sharing of truths is deemed important for the good of public health, more so than individual privacy. A competing model is that of *dynamic consent,* which recognizes the need to provide protection for individual interests in the complex and evolving landscape of data sharing [14,15,40]. Here, consent is viewed as an ongoing relationship, where the patient is kept informed of how their data is being used and is free to express different preferences at any time.

Informed consent has been widely adopted in research and in software design, and is often considered adequate to satisfy the legal obligations upon data-holders [11]. In HCI, research has shown that software designs which ask users to give informed consent to standard end-user-license agreements (EULAs) during installation are ineffective, with 74% of people rarely or never browsing them [13]. Patrick and Kenny suggest four design principles for a more privacy-aware interface design: comprehension, consciousness, control and consent [29]. Consent is paramount because it is "the primary means for individuals to exercise their autonomy and protect their privacy" [8]. Morrison et. al. argue that we need to increase end-user understanding of the consent question, and propose a delayed intervention approach, where the user is asked for consent in the context of their task using a visualized, personalized representation of relevant data to make the question more meaningful [24]. This "just-in-time" approach has been endorsed in the US Federal Trade Commission's best practice guidelines for mobile app developers [9], and Facebook also follows this model.

Such approaches may be more effective, but still do not address a key failing of informed consent – that users must be able to meaningfully and knowingly revoke consent as easily as they give it. Consent routinely becomes a 'point of severance', beyond which users surrender all agency and control over their personal data [19]. The field of Ubiquitous Computing (ubicomp) has paid particular attention to consent in recent years, due to the complexities which arise when citizens use multiple devices in many contexts; a consent given in one context may no longer apply in another. Luger and Rodden suggest harnessing Friedman [10]'s idea of 'actively embedding human values' in the systems we design and recasting consent as a social process [19], where communication replaces 'informing' and users are given regular and comprehensible feedback to support an ongoing relationship with their data – a design that exemplifies the dynamic model of consent.

Family Civic Data

We introduce the term *Family Civic Data* to refer to the data about families held by any local or national authority - any and all data that might be stored about a family. By 'civic' we refer to the administration of society and citizen relations as a whole, not just of the local region, as a preferable term to 'social', which carries online and leisure connotations. Our working definition of 'data' is 'information that has been stored' (typically on a computer).' It is important to consider the meaning of 'family'. Using any single factor such as biology, co-residence, or parenting, is insufficient for defining family [26]. We took a pragmatic approach of including any self-declared family group, including single parents. Cornford et. al. [6] have identified that this ambiguity around the concept of family causes great difficulty for the state, especially when attempting to construct a holistic view of a family, as it is difficult to identify which individuals to include and whose records to link together. They conclude that a simple database is insufficient for encapsulating a family's details, and that any system handling Family Civic

Data must be agile and configurable by both professionals and citizens, and should support aggregation, negotiation, personal development, planning and more complex capabilities.

As part of our grey-literature review of the field, we collected examples of different types of Family Civic Data, in order to develop a taxonomy of terms and data types, as shown in Table 1.

S	Type	Examples / Details
1	Personal details	Date of birth, address, telephone number.
	Relationships	Marital status, ex's, step-parents.
	Children	Parentage, adoption, fostering, childcare.
2	School Records	Attendance (truancy), special needs.
	Academic Results	SATs, reports, exam failures, training courses
3	Social Support	Social worker visits & notes, details of family crises, interventions, allegations.
	Welfare Benefits	Jobseeker's Allowance, child support, Disability Living Allowance, tax credits
4	Family Finances	Salary, savings, credit cards, spending, debt
	Employment	Job history, periods of unemployment, performance at work, NI, PAYE, pensions
5	Housing	Council house provision, eligibility criteria.
	Legal documents	Birth/marriage/death certificates, citizenship/immigration status, work permits
6	Criminal records	Arrests, cautions, offenders' registers, prison time, speeding tickets, spent convictions
	Court orders	Restraining orders, lawsuits, custody, ASBOs
	Domestic Violence	Allegations made, medical records, social/legal interventions, victim support
7	GP records	GP's notes, prescriptions, tests, referrals
	Hospital records	Operations, hospital stays, emergency care
	Medical conditions	Diagnoses, diseases, allergies, blood type
	Mental health	PTSD, breakdowns, depression, sectioning
	Addictions	Substance abuse, gambling, rehab, crime
8	Library Usage	Books/CDs borrowed, computer access
	Sports & Health	Gym usage, class attendance
	Shopping Habits†	Loyalty cards, store & online purchases
	Transport Data†	Buses used, ANPR tracking, walking patterns

Table 1. Example Categories of Family Civic Data.

Super-categories (S) are: family (1), education (2), welfare (3), money/work data (4), civil (5), crime (6), medical (7), leisure (8).

In this paper, we augment existing research, which frames the storage, sharing and combining of Family Civic Data as a problem for the state best solved through technology [44–48], with a meaningful understanding of the family perspective. In doing so, we present a contrary perspective that families require more, not less, human interaction. We show that while it is possible to design new, holistic systems for presenting and interacting with Family Civic Data, these will only be effective at helping families if such systems are designed to meet both citizen and practitioner needs and facilitate use by both parties in close collaboration as part of a human interaction.

† Some leisure categories were included that are not strictly civic data, as these are useful for exploring issues of ethics and appreciating the "big data" benefits of data linking.

METHODOLOGY

In order to address our main research aim of understanding how families would prefer their data to be handled, shared and stored, we set ourselves two objectives. The first was to design methods of interacting that would allow families to meaningfully converse about the dry, abstract topic of civic data. Our second was to motivate and engage participation from the whole family, including children. Our method, *Family Design Games*, which was ethically approved by our University's IRB, consists of a home-based ethnographic interview [22], structured as a series of games designed to make the topic interesting and to encourage intergenerational conversation about Family Civic Data. Our approach draws from the *design games* of Brandt and Messeter [2,3], but also from Sarah Pink's sensory ethnography techniques [30]. Interviews-in-place can offer deeper insights than a traditional interview; as Dawn Mannay observes, *life happens around you* [22]. Home visits enabled us to interact with the whole family at once, grounding our design understanding in their lived experience. Projects such as VOME [51] have used mixed methods to research attitudes to privacy and consent, but much human-data interaction research has focused upon the individual, so our focus on engaging the whole family together through design games in the family home brings a fresh perspective to this under-researched field.

Our interview schedule, showing the questions posed by each activity, is included in the supplemental materials. Each session began by inviting families to consider who might store information about a family. This established a shared understanding of Family Civic Data, while also encouraging broader thinking about data. Then, in our *Family Facts* game, we presented participants with a series of six *Petals* (writable paper tags to be hooked over poles) containing fictional 'facts' about a family, for example 'Mum Sarah (54) buys three packets of cigarettes every week.' These scenarios were designed to provoke consideration of information ownership. The family was invited to discuss each Fact and categorize it as "Not Data", "Our Data" (meaning, this belongs to the family) or "Their Data" (meaning, this belongs to the state or data-holding authority) by putting it over the appropriate pole (see Figure 2). Participants then created their own Facts, drawing inspiration from their own lives, and discussed and categorized them in the same way. The objective was to



Figure 1. Example Data Cards created as "things-to-think-with" – visually embodying and exemplifying Family Civic Data.

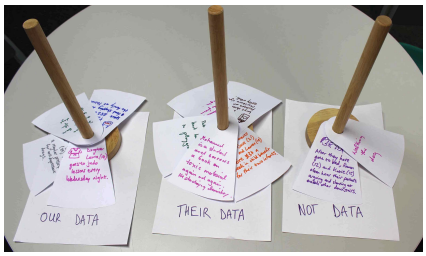


Figure 2. The *Family Facts* sensitization exercise: placing family facts onto poles according to whose data it is.

endow an appreciation of the inherent conflicts of interest between family and state on the storage and sharing of Family Civic Data, and to prime participants to be able to consider who should control it.

Card Sorting

Inspired by Brandt and Messeter's observation that game pieces can be used to create common ground and as "things-to-think-with" [3,28], we created a set of *Data Cards*, that would serve as a visual and tangible representation of Family Civic Data, and as boundary objects [1,17] to bring researcher and participants' worlds closer together. The development of the cards and associated lexicon made accessible the topic of data, which, as we discovered during recruitment, many families feel unqualified to talk about. A Data Card was created for each category in Table 1, including a summary and meaningful examples, so that the cards would be easy to digest, yet still contain sufficient detail to stimulate thinking. Keeping child-friendliness in mind, bright colors were a key element of the design. The cards (see Figure 1) were printed on high-quality, thick card with a glossy finish using a business card printing service.

For our main exercise, families were invited to perform a card sorting task using these Data Cards. We labelled the extremities of a corkboard so as to give meaning to the space on the board, creating two notional graph axes – a

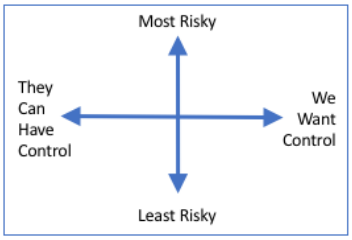


Figure 3. Notional axes for corkboard in card sorting exercise.



Figure 4. Doing the card sorting exercise with a family.

'control' axis, and a 'risk' axis (see Figure 3). A family member was invited to pick a Data Card from the deck, read out the card's data type, summary and example data, and then discuss with their family where to pin the card on the board, considering:

i. how much control they think the family ('we') should have of that data versus the authority holding the data ('they'),
ii. how risky it would be if this type of data was lost, stored with mistakes in it, or shared with the wrong people, and
iii. how it should be placed relative to other cards already pinned on the board.

This was repeated around a dozen times, with a different family member picking the card each time. Figure 4 shows a family carrying out this activity.

We then conducted a follow-on exercise (designed to be fun for all ages) to consider how each data type should be handled. A set of custom stickers were produced, containing statements that we thought a family member might say about a specific type of data, for example, *"this makes me uncomfortable"* or *"I don't care about that."* Plain white stickers were also available so that participants could write their own where no suitable sticker was provided for the idea they wished to express.

Reviewing and Reflecting on a Mock-up Interface

We then interviewed families about the requirements for an interface to allow access to and control of a family's data. We contextualised this discussion by providing printouts from spreadsheets of (fictional) data of arrests, benefits and domestic violence cases (which is how Family Civic Data is currently shared between agencies), as well as some news articles illustrating what can go wrong when citizens' data is leaked, or contains errors or omissions. Then, we presented two single-page printed mock-up interfaces for TFP social workers (see Figure 5), explaining how a social worker might view and use a family's data. Finally, we invited comment on the design and asked participants what they considered the most important features or attributes of a computer system that handles Family Civic Data.

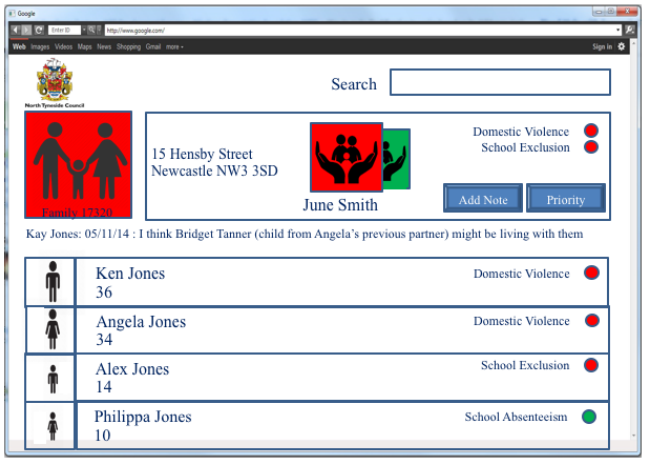


Figure 5. Mock-up interface printout shown to families.

Participants

Working with a local community center as well as through word-of-mouth, we recruited four families from the region. Two hours was allowed for each family visit, and families were rewarded with a gift voucher for a family day out. Families were reassured that they would only need to talk about families in general, and that sharing specifics from their own lives would be entirely optional. All sessions were audio recorded and field notes were written later the same day after each session.

Family 1 was an academic couple in their twenties (with an infant, not present), who took up the option of meeting at our University rather than in their home; the remaining sessions were all home visits. Family 2 was a twentysomething mother along with her 5- and 12-year-old children and her fiftysomething mother. Family 3 was a couple in their thirties, whose children were not present. Family 4 was a single mother in her forties, and her four youngest children, aged between 6 and 15.

FINDINGS

The primary purpose of our activities was to produce a good conversation which we would then be able to analyze. We photographed the physical outputs, during and after each session, but their sole purpose was to add contextual understanding to conversation transcripts. Transcripts and field notes were coded and organized through reductive data display cycles, as per the Miles and Huberman framework [32]. The themes uncovered by our analysis are presented below under three main questions: what does this data mean to the family; when should it be in control of the authorities; and how should it be handled? In addition to these main qualitative findings, we use a quantitative analysis to show how families collectively think different types of data should be handled. Finally, ethnographic and methodological observations are recorded separately.

What does Family Civic Data mean to the family?

Data (and information) is private

In the Family Facts exercise, families talked about everyday information about their lives as being no-one else's business – *"just personal* [private] *stuff that no-one has written down"* (F2[2]) – that only becomes data when it is reported or someone wants to know about it. Families expressed a desire to keep control of their *"delicate"*, *"sensitive"* (F1) information, which "*should be kept private until we say otherwise."* (F4). Relationship data regularly came up as an example of *"inherently personal"* (F1) data that *"no-one has a right to know"* (F2).

Data handling requires trust

Because information can *"reveal a lot about a person"* (F1), it entrusts a great deal of power to the data-holder. Families desire to keep control of their data was often motivated by fear; they expressed a desire to keep control when they perceived risks of it being shared or misused, fearing the consequences of mishandling. Often implicit in this was a distrust of authorities. For example, (F3), who we discovered had been heavily involved with the criminal and welfare system as a result of addiction-related issues, felt that authorities often shared, withheld or misrepresented their data against the family interests, and thus wanted full control of all the data about their situation.

[2] In this report, we use the shorthand (F1) to refer to 'a member of Family 1,' or in some cases to the whole family.

Data is risky

All Family Civic Data was perceived as risky. More severe perceived consequences of data mishandling led to families wanting more control. Families perceived a variety of risks – criminal, medical, welfare, social and psychological:

Hacking and identity fraud was often mentioned; (F2) noted that *"when you think of the forms you fill in to apply for a course, we give them everything you need to clone us."* (F1) were afraid to complain to police about a neighbor for fear of retribution if their anonymity were compromised, while (F2) feared that knowledge of their salary information would make them a target for burglary. Information about parentage was identified as carrying risk of significant psychological damage, *"He thinks that is his dad and he doesn't know any different. But obviously with him having social and emotional problems, if he found out, it would mess with him."* (F3) Families were also concerned by the potential for social stigmatization if information became public, especially in the form of gossip from neighbors. (F2) recalled how she and her sister *"wouldn't have any of the free school meals, because you had to line up for a dinner ticket, then everyone knew you were different."* However, families could see the benefits of certain risky data being stored and appropriately shared, for example the medical risks if their medical histories were not made available to doctors and hospitals. (F3) recounted an incident where a probation worker's refusal to disclose his whereabouts to his wife put his own welfare at risk.

Data is misleading

A common problem the families observed about data stored about them, is that there is no absolute truth; a lot of the data is recorded by partisan observers meaning it is at best subjective and at worst purely an opinion. While looking at the mock-up interfaces that invited social workers to provide numerical scores for troubled families' progress, (F3) highlighted that these are just opinions, adding *"It's like they're giving you marks... it shouldn't be like that"*. (F2) talked about the subjectivity that schools and playgroups would exert when deciding whether to (i) record and (ii) share details of an injury that a child had sustained while in their care. (F3) said that their social worker recorded details of the couple's arguments, and threatened to report these to social services. (F2) talked about being wrongly classified as a troubled family because of their postcode. Ultimately, the families' observations show that data can disempower a family, especially when they are labelled with categories such as 'domestic violence' or 'troubled family' and then judged solely by that label. In response to the Family Facts activity, all families felt that the scenario regarding a mother who had had a breakdown ten years earlier would lead to unfair judgement. The families noted that anyone could fall foul of such unfair judgements, even if they have done nothing wrong – *"False allegations can be very damaging,"* (F1) and *"anyone can be harmed by data sharing."* (F3)

When does Family Civic Data belong to the state?

When "they" need to know

Families acknowledged that data is routinely generated and stored about them through interactions with state services, and that sometimes police, schools, medical professionals, care workers, and employers need that data to operate and perform their job effectively. The question of 'Do they need to know?' is the major determinant in deciding whether data should be shared, controlled or accessed by public sector service providers. As (F2) put it *"If they need the information, that's a reason for them to have control* [of it]*."* During our card sorting activity (F2) said they didn't need the "not for police eyes" sticker because *"the police should be allowed to see everything."* (F4) said they wouldn't mind care workers having access to family data, but "*Just them. Me and the people that really need it. Say, like, doctors, social services. Not just any randomers.*" (F3) noted that sometimes organizations want, rather than need, your information – *"When you're getting Sky, when you're getting a house, everyone wants to know if you're working or on benefits."*

When "they" are responsible

In some cases, families felt that authorities having control of data would be beneficial. For example, F2, following an observation that she'd lost her old exam certificates, said "*Most things I'd prefer other people to have control of* […] *then I know it's somewhere where it won't get lost."* Keeping data is a responsibility, and not having that responsibility removes the burden of worry. (F2) observed a similar 'hassle-removing' benefit from medical data sharing, saying she would opt-in to such sharing because *"it's a right nightmare when I've got to go back and forwards to explain everything."* Other examples where the data-holder was deemed to have a clear responsibility to control the data were medical prescriptions, adoption data (F2), and sensitive data about addictions (F3). Coupled with responsibility – and in line with literature – was a sense of trust, with both (F3) and (F4) noting the trust they put in their social workers and doctors not to share their data.

When the data can serve the public interest

Families observed that in certain circumstances, data should be shared with relevant authorities or, sometimes, with the public at large. (F2) said that criminal records should be shared if there is a danger to the public, and (F1) talked about the importance of sharing information that will protect a child from potential abuse. (F3), from his experience of working in a substance abuse charity, observed that part of the social worker's responsibility is to interpret data and spot patterns, and in this context, it is good to have access to as much data as possible, giving, as he put it, *"many hands, many eyes, many ears, many indicators."* He felt that not enough data sharing happens between public sector agencies, which (F1) also observed from her voluntary experience with Sure Start, a governmental child support initiative. (F4) said people should be able to know who lived in their house previously:

"it could have been a dodgy person and then [...] *you could be getting people coming in, and getting their drama in your life."* Despite having viewed tracking people's journeys as *"scary and stalkery"*. (F2) did say that there is a wider public benefit to large scale data collection: *"If they're tracking where people are going and there's not a bus going to that place, they know to put a bus on."*

When the family has nothing to lose

In several cases families were happy for the authority to control their data when they felt there was no risk. (F2) were unconcerned about shopping data and routine medical data being outside of their control because *"if somebody got hold of that information, what could they do?"* (F4) made the same observation about shopping data and a similar observation about exam results. (F2) said they did not care about library data because they had nothing to hide.

When data is gathered by default, or actively sought

One area where families would prefer state authorities not to hold the data they do, is when excessive data is collected as part of enrolment processes, and kept when not really needed. There was a general sense that too much data is being collected - *"It's like they need to know every single thing going on."* (F4). (F2) observed that *"data's going to be saved somewhere no matter what you say."* Families felt they have no choice about data held by others about them: *"They're all things in databases somewhere that we have no control over"* (F2). *"Sometimes there's no way of avoiding it"* (F1). Each family identified types of data that should not be stored – commercial (F1, F3), leisure (F4), and relationships (F2). (F1) felt that when asking about illnesses, employers were asking for data they had no right to ask for, and observed that private health insurers would love to obtain data about people's smoking. (F1) said that health data should never be sold, citing a recent incident where Google obtained NHS data. However, (F2) acknowledged that making data available to companies can be a trade-off: *"I wouldn't get my money off if I didn't have* [a loyalty card that records my purchases]*."* In this sense, personal data becomes a commodity [33]. Families do not have the right to total control of their data, though: (F2) said while trying to think of a good example for the Family Facts game, that *"it's hard to think of something that is only our data, that they* [the authorities] *have no claim to."*

When data sticks around

Most families expressed a view that authorities are routinely keeping data around for a long time, longer than is needed. Speaking of a breakdown in her past, (F2) said *"I would rather people not know* […] *Obviously, it's somewhere on the doctor's computer, but it's more your data than his, isn't it?"* (F3) talked about how the "drug addiction" label had stuck to her family so that they are permanently in that category and treated differently by social and medical care professionals. (F1) suggested that the problem was the lack of any legal obligation for authorities to delete families' information.

How should Family Civic Data be handled in the future?

A set of basic rights - to be informed, involved and accurately represented

As families considered who should control different types of Family Civic Data, and what a future computer system should do, a phrase that came up often was 'the right to…' The families would like the right to be informed about and involved with the handling of their data. (F1) said while looking at the TFP interface mock-up, *"Families need the right, without a long and costly procedure, without a lot of bureaucracy, to see the information social workers have about them."* (F2) and (F3) also envisaged a *"right to know"* and a default *"right to see."* The mock TFP interfaces were thought to be practitioner-centric and clearly not supportive of these rights. It emerged that access is not enough; families must be actively informed about what data exists and how it is being used. This is in line with Human-Data Interaction (HDI)'s concept of *legibility* [25]. (F2) described the proposed social worker interface as *"a load of crap",* and (F1) observed that it *"treats families like subjects, not as individuals with rights like privacy,"* a view that provides evidence for Cornford's premise that we need to 'move beyond using information systems that treat [families] as objects to be administered.' [6] A commonly occurring theme was the idea that information should be fair and accurate; (F1) added a sticker for this, and (F3) talked about the damage caused by labelling or categorizing a family. (F1) and (F4) want (similar to HDI's concept of *agency* [25]) the *right to choose* how data is used, and when to share or disclose it, especially when telling children about adoption. Families also are concerned about incorrect or incomplete data, and need associated rights to deal with that. (F1) described this as the *"right of veto"* and also wished for a *"right of deletion";* (F2) talked about needing the *right to explain* or add comments when the data does not tell the full picture. The manner of engagement was also discussed – (F4) talked about wanting to go and see someone in person to talk about her data, reminding us that families' interaction with data should be thought of not just as a technological interface design but as a human interaction, a relationship with representatives of the state. Families want their representation in data to be complete and true. A row in a spreadsheet, or a set of numerical values, is not enough to adequately represent the complexities of families' situations: *"You need to be able to write and explain, because every family is different."* (F2).

One safe place for data, with sensible access controls

When discussing current practices, families felt that emailing spreadsheets, and even social workers retrieving search results of family data as needed, represent a data handling approach which is too scattered, too uncontrolled, and too risky. It was proposed that Family Civic Data should only exist in one place. (F2) said that *"you should go to the data rather than have it sent to you"* and suggested accessing your data like requesting a record in the library. (F4), viewing herself as not computer-literate,

argued for a place you could physically visit to see, and talk to someone about, your data. (F1) said *"data should never leave its source"* and that families and practitioners should *"go to a place to get the data"* in order to keep the data safe. (F3) said "*All data should be protected* […] *from whoever* [we] *choose.*" A young child in (F4) described how important it was to *"keep the lock on it"* and have different accounts for each person allowed to access the data, highlighting the importance of being able to tell who has accessed information after the fact. (F1) imagined a sole organization for citizen data management to make families' relationship with their civic data really work. Interestingly, contrary to our expectations, all the families think systems should not allow them to directly edit their data. (F2) said she lies about her smoking habits whenever the doctor asks her, and observed that families lie to get themselves out of trouble or to deflect official attention. (F3) also cited smoking as something they would lie about, and talked about how people under suspicion can *"blag* [the authorities] *to death and they go away thinking, 'They're doing amazing.'"* In general, families' requirements originate from a desire to ensure their Family Civic Data is always fair, accurate and representative – consistent with Cornford's premise that families need to be *observable through data* [6]. The requirement to keep data safe and secure will soon become a legal requirement, through the EU's General Data Protection Regulations [16].

To work together with the state in a positive relationship

One of the most interesting themes that emerged was that the families are willing, and indeed actively want to, work together with authorities to get the best help and support they can. This began with suggestions of discussing, querying and explaining their data, but went further into families wanting to *"have a conversation"* with the state authorities (F2), and wanting authorities to get to know them as a family with its own unique story – which they are only too happy to tell – not just as a database record. (F3) recognized the importance of having a good relationship with data-holders, and offered a positive vision for the future, suggesting that with better data sharing, *"There would be no loopholes, no uncovered stones. People would be working together. To protect something, to move something on, you can't do it all on your own, you have to work as a collective."* The idea of families being involved in the state's recording of their data is a radical and beneficial departure from current approaches. (F1) observed that if you invite families to view and comment on their data, you *"get them involved, and they think, 'Oh we are not* [just] *subjects anymore.'"* He further suggested that engaging families like this on a human level could inspire passive families that have "given up" to seize more control of their lives; *"Maybe that changes the whole family."*

Quantitative Analysis: Who should own families' data and what are the risks?

While the sample size was not large enough to be statistically representative, it was possible to analyze the data from the card sorting exercise in a quantitative manner, in order to identify trends in opinion for specific super categories of data. This was done by assigning numerical scores for control and risk as x and y values according to the placement of the Data Cards on the corkboard.

On average (Figure 6), we see that families thought that those types of data that concern society as a whole (Civil, Money/Work, Crime and Welfare) should be more in the control of the state, whereas those data types closer to the individual (Family Information, Education, and Leisure) should be more in family control. One exception was Health data, thought best left to the professionals. For risk, we see a different pattern, with Money/Work data being thought risky, and Crime data especially so. The more "individual" types of data such as Health, Family Information and Education were thought moderately risky. Civil and Leisure data were considered the lowest risk. Looking at variance and standard deviations, the areas where families expressed most varied opinions were around Health, Family Information and Leisure. Crime, Civil, Money/Work and Welfare data were most agreed upon.

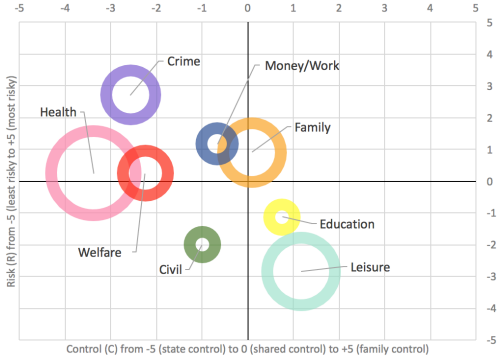


Figure 6. Mean average control and risk per data type. Size of circles denote relative variance (not to same scale as axes).

In general, families thought Family Civic Data should be slightly more in the data-holder's control (-1.1C, σ=3.2), and that it carries an above average risk (0.5R, σ=3.3).

Ethnographic and Methodological Observations

In all four sessions, adults and, contrary to expectations, the children present (especially those aged between 5 and 12) engaged enthusiastically with the activities, rather than being passive observers to their parents' participation. All participants seemed to have fun and be stimulated and interested in the activities, which managed to keep the atmosphere light and playful – as seen when (F1) drew her third purple (Crime) card in a row and exclaimed, *"Oh no, Alex, not another purple one!"*. The designs of our cards, stickers and activities had been designed to be child-friendly (in the hope that this would encourage both adult and child participation), and indeed the children were most interested in the game-like aspects – turn-taking, card-picking, pinning cards and placing stickers; even the researcher's camera was of interest. The drawing and writing elements of the Family Facts task were embraced

by younger participants – e.g. a five-year old boy created a Family Fact by drawing a picture of his late pet tortoise, and labelled it with help from his mother, despite his undeveloped writing skills (F2). In this family setting, literacy was not essential for participation: both parents and researchers took the time to explain things or assist the children, which gave the sessions an unplanned educational aspect. The child-friendly designs did not "dumb down" the topic, and also helped adults to engage easily with the topic.

Although the sessions were structured as games, the activities played out mostly as discussions between family members about where to place Data Cards or Petals, or what should happen to certain data types. The researcher's role was mainly that of facilitator – occasionally clarifying or sense-checking an idea, or offering an alternative point of view to stimulate the discussion, but largely the researcher was able to step back and observe the family talking with itself. It was these discussions that provided the majority of the findings detailed above. In our role as ethnographic observers, the researchers were able to infer a lot about the lifestyles, relationships and living patterns of the families, even the absent family members, e.g. from photographs or domestic interruptions, that did not occur in the removed setting of (F1)'s interview. Many interesting dynamics between family members were observed – children vying for control of the game, parents 'talking over' each other, parental realism contradicting childhood optimism, children revealing past parental conflicts, parents making data decisions for their children, jokes and banter between couples about keeping spending secret from each other or exasperated-but-playful expressions of despair at children's interruptions. These observations are simply the substance of normal family life; clearly the familiar setting of the home put families at ease, while also revealing hidden aspects of family life and providing researchers with an unexpectedly rich window into families' lived experiences.

DISCUSSION

Through our lexicon of Family Civic Data and the use of Family Design Games in family homes, we have successfully conducted meaningful conversations with families. Our analysis has revealed a new understanding of how families would like their data to be handled. While recognizing that Family Civic Data is *private* and *risky*, they are generally happy for their Family Civic Data to be in state control, but are concerned about *losing their influence*, especially where the stored data may be *misleading*. The families share a belief that that numerical metrics, problematized labels like 'domestic violence' or 'troubled family', and rows in spreadsheets, cannot adequately or fairly convey the reality of their lives. This seems to be a systemic problem, as beyond the aforementioned "point of severance" [19] that a simple opt-in/opt-out model of consent [41] creates, the data becomes embedded in the organizations that hold it – it becomes *data-in-place* [37]. Privacy researchers agree that meaningful consent must be more than a moment-in-time decision [31,39]. Families want to be *involved* in the stewardship of their data to ensure that those making decisions that affect their lives can always see a full, fair and accurate story of their lives – in this, we see strong evidence that families' needs can only be met by the dynamic consent model [14,15,40], where management of Family Civic Data becomes an ongoing conversation between family and state.

Implementing a dynamic consent model will be a daunting task for any data-holding organization; it will require cultural change and significant redesign of both analogue processes and data management systems, as outlined by Kaye et al. [40]. It may be hard for a system designer to know where to start. Building upon the ideas and insights gained from our participants, we propose three precepts which can serve as guiding principles to inform the design of a new dynamic-consent-based data-handling system:

Families must be given continuing rights, control and visibility over their Family Civic Data and its handling, to keep data accurate and meaningful.

A core principle of dynamic consent is that the citizens about which the data is held must be able to see what data is held about them and how it is being used, at any time, and change their consent. This is not just a matter of displaying the data in an interpretable way, but giving *negotiability* [25] to families so that they can discuss, challenge and communicate over it and give or revoke consent. Patrick and Kenny's aforementioned design principles could be helpful here [29]. It is also important to note that this requirement is not just about "doing the right thing for families" – without their ongoing involvement, the gap between the family's reality and that suggested by their stored data will grow, increasing the risk of the family being mismanaged as a result. Having the family involved serves as a quality control check for the data records.

Family Civic Data must never be allowed to serve as a proxy for the family's involvement in decision-making, to avoid disempowerment.

Our analysis of families' views shows that in isolation, data about a family stored in an official database will never be enough to adequately capture the complexity of their lives – if you want to tell the whole story of a family, they have to be involved. Similarly, families' accounts of their lives cannot be assumed as fact, since as our participants told us, families will lie or bend the truth in pursuit of their own interests. As such, to find the objective truth requires an ongoing interpretation of both the state's data and the family's statements. The family and the state must work in partnership, and where there is disagreement, there must be discussion and collaboration. The implications of this shift to user 'control' of organizations' data is further explored by Whitley et al. [39].

Families must be able to have a relationship with their Family Civic Data, not just its holders. This requires a mindset shift for data-holding organizations, who must co-operate with each other in a family-centric way.

Through the lens of our research, we can see that Family Civic Data exists in the world, not as a simple description of a family or just as a side-effect of family-state relations, but as a distinct entity with an ongoing lifespan, which can encapsulate facts, opinions and possibly incomplete views of a family's life. As such, designers should develop interfaces for family involvement that support Human-Data Interaction [25] – where the Family Civic Data is an artifact co-operatively managed by both family and state. The relationship between family and state will need to have this data at its core, rather than in the background, out of the family's view or influence. The reality is that Family Civic Data is scattered throughout many, often disconnected, databases and systems, and so what is needed is not a unifying technological solution, but rather a paradigm shift in the way that civic organizations think about Family Civic Data. The family's data should become a living, breathing thing that both the family and the organization(s) will nurture together. There must be no point of severance. This shift in focus will begin to address the families' need for safe and controlled data handling, while the shift in focus towards co-operative family-state collaboration will improve data quality, by encouraging reconciliation of discrepancies and the establishment of shared and accepted truths. This will leave organizations better placed to meet the requirements of forthcoming GDPR legislation [49].

Reflection: Developing the 'Family Design Games' approach into a re-usable method for engaging families

As well as our insights into families, we believe our evolving approach of *Family Design Games* can be valuable to other researchers who wish to uncover family perspectives about other dry or complex topics. It can be adapted to other study topics by following these steps:

1. Gather Examples – Find examples of the type of data or thing you wish to explore: in our case, we found real-world examples of Family Civic Data.

2. Categorize – Organize your examples into groups, and name them according to what they represent; this process gave us our lexicon and taxonomy of Family Civic Data.

3. Create Physical Representations – Design a physical object to represent each category that can serve as a "thing-to-think-with" [3,28]. In our case, our Data Cards represented the different types of Family Civic Data.

4. Design Sensitization Activities – Design conversation starter activities that will allow the participants to think about their own life experience in terms of the questions you want to ask. Our *Family Facts* activity helped the family to think about their own life as data, and to be better equipped to consider questions of data control and risks.

5. Design Activities that ask Questions – Decide what questions you want to explore, and design activities that invite participants to do something with the physical objects that forces them to consider these questions. In our card sorting task, placing each card required consideration of the questions of data control and risk.

6. Conduct the Interviews – Run your design-game-based interviews with your participants, to give a rich transcript that can be analyzed to find their answers to your questions.

A key element of the success of our *Family Design Games* was our Data Cards, which made the boring, abstract topic of Family Civic Data tangible and relatable. In Tim Coughlan et. al.'s work studying technology use in the home [7], the research team used a dollhouse as a 'stand-in' to embody a representation of 'the home' in a way that is both relatable to the participants' own home and life, yet also dissociated enough to allow participants to relax and not feel like they are being examined. Our Data Cards functioned just like Coughlan's dollhouse, as a focal point that allowed families to talk freely about their own lives and views without feeling personally interrogated. By putting something before participants that is both tangible and separated from real life, you can facilitate a more revealing and personal conversation than a traditional interview.

CONCLUSION

Using our novel *Family Design Games* technique, we have shown that families – both adults and children – can be engaged in meaningful conversation about the complex and esoteric topic of *Family Civic Data*. Through these home-based activities we have gained a rich qualitative understanding of families' requirements for the handling of their data, and derived guidelines that can inform the design of future data-handling systems and associated organizational processes: state organizations must support a dynamic consent model of data handling, and plan for a new paradigm of co-operative, data-based relationships with families, one where meaningful, representative data is nurtured for mutual benefit and families remain involved throughout.

ACKNOWLEDGEMENTS

Thanks to our colleagues at Newcastle University and in local communities across the North-East for their valuable support. This work was funded by the EPSRC CDT in Digital Civics (EP/L016176/1) and partially funded by the Department of Health and Social Care's Connected Health Cities (North East & North Cumbria) initiative (SILVER sub-project [48]). Data supporting this publication is not openly available due to confidentiality considerations. Access may be possible under appropriate agreement. Additional metadata record at http://dx.doi.org/10.17634/154300-65. Please contact Newcastle Research Data Service at rdm@ncl.ac.uk for further information or access requests.

REFERENCES

1. Geoffrey C. Bowker, Stefan Timmermans, Adele E. Clarke, and Ellen Balka. 2016. *Boundary Objects and Beyond: Working with Leigh Star*. MIT Press.
2. Eva Brandt. 2006. Designing Exploratory Design Games: A Framework for Participation in Participatory Design? In *Proceedings of the Ninth Conference on Participatory Design: Expanding Boundaries in Design - Volume 1* (PDC '06), 57–66. https://doi.org/10.1145/1147261.1147271
3. Eva Brandt and Jörn Messeter. 2004. Facilitating Collaboration Through Design Games. In *Proceedings of the Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices - Volume 1* (PDC 04), 121–131. https://doi.org/10.1145/1011870.1011885
4. Jeff Chester. 2012. Cookie Wars: How New Data Profiling and Targeting Techniques Threaten Citizens and Consumers in the "Big Data" Era. In *European Data Protection: In Good Health?* Springer, Dordrecht, 53–77. https://doi.org/10.1007/978-94-007-2903-2_4
5. Connected Health Cities. 2016. *Liverpool Care Record*. Retrieved from https://vimeo.com/190984383
6. James Cornford, Susan Baines, and Rob Wilson. 2013. Representing the family: how does the state "think family"? *Policy & Politics* 41, 1: 1–18. https://doi.org/10.1332/030557312X645838
7. Tim Coughlan, Kerstin Leder Mackley, Michael Brown, Sarah Martindale, Stephan Schlögl, Becky Mallaband, John Arnott, Jettie Hoonhout, Dalila Szostak, Robin Brewer, Erika Poole, Antti Pirhonen, Val Mitchell, Sarah Pink, and Nicholas Hine. 2013. Current Issues and Future Directions in Methods for Studying Technology in the Home. *PsychNology Journal* 11, 2: 159–184.
8. Liam Curren and Jane Kaye. 2010. Revoking consent: A 'blind spot' in data protection law? *Computer Law & Security Review* 26, 3: 273–283. https://doi.org/10.1016/j.clsr.2010.03.001
9. Federal Trade Commission. 2013. *Mobile privacy disclosures: Building trust through transparency.* Federal Trade Commission.
10. Batya Friedman. 1996. Value-sensitive Design. *interactions* 3, 6: 16–23. https://doi.org/10.1145/242485.242493
11. Batya Friedman, Peyina Lin, and Jessica K Miller. 2005. Informed consent by design. *Security and Usability*, 2001: 503–530.
12. Pat Gannon-Leary, Sue Baines, and Rob Wilson. 2006. Collaboration and partnership: A review and reflections on a national project to join up local services in England. *Journal of Interprofessional Care* 20, 6: 665–674. https://doi.org/10.1080/13561820600890235
13. Nathaniel S. Good, Jens Grossklags, Deirdre K. Mulligan, and Joseph A. Konstan. 2007. Noticing Notice: A Large-scale Experiment on the Timing of Software License Agreements. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07), 607–616. https://doi.org/10.1145/1240624.1240720
14. Jane Kaye, Edgar A. Whitley, Nadja Kanellopoulou, Sadie Creese, Kay J. Hughes, and David Lund. 2011. Dynamic consent: a solution to a perennial problem? *BMC* 343, nov01: d6900–d6900.
15. Jane Kaye, Edgar A Whitley, David Lund, Michael Morrison, Harriet Teare, and Karen Melham. 2015. Dynamic consent: a patient interface for twenty-first century research networks. *European Journal of Human Genetics* 23, 2: 141–146. https://doi.org/10.1038/ejhg.2014.71
16. Christopher Kuner. 2012. *The European Commission's Proposed Data Protection Regulation: A Copernican Revolution in European Data Protection Law*. Social Science Research Network, Rochester, NY. Retrieved August 24, 2017 from https://papers.ssrn.com/abstract=2162781
17. Susan Leigh Star. 2010. This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology, & Human Values* 35, 5: 601–617. https://doi.org/10.1177/0162243910377624
18. Ruth Levitas. 2012. There may be 'trouble'ahead: what we know about those 120,000 'troubled'families. *Policy response series*, 3: 21.
19. Ewa Luger and Tom Rodden. 2013. An Informed View on Consent for UbiComp. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '13), 529–538. https://doi.org/10.1145/2493432.2493446
20. Jeantine E. Lunshof, Ruth Chadwick, Daniel B. Vorhaus, and George M. Church. 2008. From genetic privacy to open consent. *Nature Reviews. Genetics; London* 9, 5: 406–11. http://dx.doi.org/10.1038/nrg2360
21. Ronan A. Lyons, David V. Ford, Laurence Moore, and Sarah E. Rodgers. 2014. Use of data linkage to measure the population health effect of non-health-care interventions. *The Lancet* 383, 9927: 1517–1519. https://doi.org/10.1016/S0140-6736(13)61750-X
22. Dawn Mannay and Melanie Morgan. 2015. Doing ethnography or applying a qualitative technique? Reflections from the 'waiting field.' *Qualitative Research* 15, 2: 166–182. https://doi.org/10.1177/1468794113517391
23. Kate Morris. 2013. Troubled families: vulnerable families' experiences of multiple service use. *Child & Family Social Work* 18, 2: 198–206. https://doi.org/10.1111/j.1365-2206.2011.00822.x
24. Alistair Morrison, Donald McMillan, and Matthew Chalmers. 2014. Improving Consent in Large Scale Mobile HCI Through Personalised Representations of Data. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast,*

Foundational (NordiCHI '14), 471–480. https://doi.org/10.1145/2639189.2639239
25. Richard Mortier, Hamed Haddadi, Tristan Henderson, Derek McAuley, and Jon Crowcroft. 2014. *Human-Data Interaction: The Human Face of the Data-Driven Society*. Social Science Research Network, Rochester, NY. Retrieved December 11, 2016 from https://papers.ssrn.com/abstract=2508051
26. Carman Neustaedter, Svetlana Yarosh, and AJ Brush. 2009. Designing for families. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, 2735–2738.
27. Aisling A. O'Kane, Helena M. Mentis, and Eno Thereska. 2013. Non-static Nature of Patient Consent: Shifting Privacy Perspectives in Health Information Sharing. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (CSCW '13), 553–562. https://doi.org/10.1145/2441776.2441838
28. Seymour Papert. 1980. *Mindstorms: Children, computers, and powerful ideas*. Basic Books, Inc.
29. Andrew Patrick and Steve Kenny. 2003. From privacy legislation to interface design: Implementing information privacy in human-computer interactions. In *Privacy Enhancing Technologies*, 107–124.
30. Sarah Pink, Kerstin Leder Mackley, Val Mitchell, Marcus Hanratty, Carolina Escobar-Tello, Tracy Bhamra, and Roxana Morosanu. 2008. Applying the Lens of Sensory Ethnography to Sustainable HCI. *ACM Trans. Comput.-Hum. Interact.* 20, 4: 25:1–25:18. https://doi.org/10.1145/2494261
31. Henrich Pöhls. 2008. Verifiable and revocable expression of consent to processing of aggregated personal data. *Information and Communications Security*: 279–293.
32. Keith F. Punch. 2013. *Introduction to Social Research: Quantitative and Qualitative Approaches*. SAGE.
33. Beate Rössler. 2015. Should personal data be a tradable good? On the moral limits of markets in privacy. *Social dimensions of privacy: Interdisciplinary perspectives*: 141–161.
34. Bart W. Schermer. 2011. The limits of privacy in automated profiling and data mining. *Computer Law & Security Review* 27, 1: 45–52. https://doi.org/10.1016/j.clsr.2010.11.009
35. N. Shadbolt, K. O'Hara, T. Berners-Lee, N. Gibbins, H. Glaser, W. Hall, and m c schraefel. 2012. Linked Open Government Data: Lessons from Data.gov.uk. *IEEE Intelligent Systems* 27, 3: 16–24. https://doi.org/10.1109/MIS.2012.23
36. Therése Sullivan, William L Martin, and Mitchell M Handelsman. 1993. Practical benefits of an informed-consent procedure: An empirical investigation. *Professional Psychology: Research and Practice* 24, 2: 160.
37. Alex S. Taylor, Siân Lindley, Tim Regan, David Sweeney, Vasillis Vlachokyriakos, Lillie Grainger, and Jessica Lingel. 2015. Data-in-Place: Thinking Through the Relations Between Data and Community. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 2863–2872. https://doi.org/10.1145/2702123.2702558
38. Caroline J. Tolbert and Karen Mossberger. 2006. The Effects of E-Government on Trust and Confidence in Government. *Public Administration Review* 66, 3: 354–369. https://doi.org/10.1111/j.1540-6210.2006.00594.x
39. Edgar A. Whitley. 2009. Informational privacy, consent and the "control" of personal data. *Information Security Technical Report* 14, 3: 154–159. https://doi.org/10.1016/j.istr.2009.10.001
40. Hawys Williams, Karen Spencer, Caroline Sanders, David Lund, Edgar A Whitley, Jane Kaye, and William G Dixon. 2015. Dynamic Consent: A Possible Solution to Improve Patient Confidence and Trust in How Electronic Patient Records Are Used in Medical Research. *JMIR Medical Informatics* 3, 1. https://doi.org/10.2196/medinform.3525
41. Donald J. Willison, Karim Keshavjee, Kalpana Nair, Charlie Goldsmith, and Anne M. Holbrook. 2003. Patients' consent preferences for research uses of information in electronic medical records: interview and survey data. *BMJ* 326, 7385: 373. https://doi.org/10.1136/bmj.326.7385.373
42. Wilson Wong and Eric Welch. 2004. Does E-Government Promote Accountability? A Comparative Analysis of Website Openness and Government Accountability. *Governance* 17, 2: 275–297. https://doi.org/10.1111/j.1468-0491.2004.00246.x
43. 2009. Eyes wide open: the personal genome project, citizen science and veracity in informed consent. *Personalized Medicine* 6, 6: 691–699. https://doi.org/10.2217/pme.09.48
44. Leeds Care Record - Shared information. Better care for you. *Leeds Care Record*. Retrieved August 24, 2017 from https://www.leedscarerecord.org/
45. Great North Care Record. Retrieved August 24, 2017 from https://www.greatnorthcarerecord.org.uk/
46. Troubled Families Programme | Newcastle City Council. Retrieved August 24, 2017 from https://www.newcastle.gov.uk/social-care-and-health/safeguarding-and-abuse/troubled-families-programme
47. TFs. *Centre of Excellence for Information Sharing*. Retrieved August 24, 2017 from http://informationsharing.org.uk/policy-areas/families/
48. SILVER: Smart Interventions for Local Vulnerable Families. *Connected Health Cities*. Retrieved August 24, 2017 from https://www.connectedhealthcities.org/research-projects/troubled-families/
49. Regulation 71, General Data Protection Regulations (EU) 2016/679 - EUR-Lex - Europa EU. Retrieved September 18, 2017 from http://eur-lex.europa.eu/eli/reg/2016/679/oj

50. Controversial database of medical records is scrapped over security concerns | The BMJ. Retrieved August 24, 2017 from http://www.bmj.com/content/354/bmj.i3804

51. About. *VOME*. Retrieved September 18, 2017 from http://vome.org.uk/about/

Appendix B. Ethics Approvals

The following 8 pages include three different ethical approvals:

- the original ethics approval focused on Case Study One,
- the original ethics approval for Case Study Two, and
- the ethics-reapproval of Case Study Two for COVID-19 mitigations.

Subject: Ethics Form Completed for Project: Useful Computing: Empowering Citizens Through Smarter Data Interfaces

Date: Wednesday, 25 October 2017 at 09:55:36 British Summer Time

From: Policy & Information Team, Newcastle University

To: Alex Bowyer (PGR)

Ref: 1052/2017

Thank you for submitting the ethical approval form for the project 'Useful Computing: Empowering Citizens Through Smarter Data Interfaces' (Lead Investigator:Alex Bowyer). Expected to run from 01/11/2017 to 24/10/2021.

Based on your answers the University Ethics Committee grants its approval for your project to progress. Please be aware that if you make any significant changes to your project then you should complete this form again as further review may be required. If you have any queries please contact res.policy@ncl.ac.uk

Best wishes

Policy & Information Team, Newcastle University Research Office

res.policy@ncl.ac.uk

University Ethics Form Version 2.1.1

Date submitted
25/10/2017 10:55:36

Applicant Details

Is this approval for a:
Student Project [A2]

What type of degree programme is being studied?
Postgraduate Research (e.g. PhD) [A3]

Name of Principal Researcher:
Alex Bowyer

Please enter your email address
a.bowyer2@ncl.ac.uk

Please select your school / academic unit
School of Computing [A1]

Please enter the module code
N/A

Please enter your supervisors email:
kyle.montague@ncl.ac.uk

Please select your supervisor's school/unit:
School of Computing [A1]

Project Details

Project Title
Useful Computing: Empowering Citizens Through Smarter Data Interfaces

Project Synopsis
My PhD focusses upon the ways in which current designs for computing software and devices that handle our personal data fail to adequately cater for the reality that our digital lives are spread across many devices, services and systems. The goals of the research are to better understand the ways in which today's data interface designs are disempowering to users, and to design, prototype and trial different types of interfaces which could better empower people to take control of their digital lives through data. As part of the research I aim to conduct interviews, focus groups, design workshops, with a variety of people, and also to prototype and trial software designs upon participants.

Project start date
01/11/2017

Project end date
24/10/2021

Is the project externally funded?
No [A3]

Does your project involve collaborators outside of the University?
No [N]

Existing Ethics, Sponsorship & Responsibility

Has ethical approval to cover this proposal already been obtained?
No [N]

Will anyone be acting as sponsor under the NHS Research Governance Framework for Health and Social Care?
No [N]

Do you have a Newcastle upon Tyne Hospitals (NUTH) reference?
No [N]

Will someone other than you (the principal investigator) or your supervisor (for student projects) be responsible for the conduct, management and design of the research?
No [N]

Animals (I)

The Animals (Scientific Procedures) Act defines protected animals as: 'any living vertebrate other than man…in its foetal, larval or embryonic form……from the stage of its development when— (a)in the case of a mammal, bird or reptile, half the gestation or incubation period for the relevant species has elapsed; and (b)in any other case, it becomes capable of independent feeding'. In practice 'Protected' animals are all living vertebrates (other than man), including some immature forms, and cephalopods (e.g. octopus, squid, cuttlefish). Using this definition, does your research involve the observation, capture or manipulation of animals or their tissues?
No [N]

NHS, Health & Social Care: Facilities, Staff & Patients (I)

Will the study involve participants recruited by virtue of being NHS patients or service users, their dependents, their carers or human tissues or the use of NHS & Health/Social Care Facilities or otherwise require REC approval?
No [N]

Human Participants in a Non-Clinical Setting (I)

Does the research involve human participants e.g. use of questionnaires, focus groups, observation, surveys or lab-based studies involving human participants?
Yes [Y]

Does the study involve any of the following? [a. The study involves children or other vulnerable groups; as defined in Section 59 of the Safeguarding Vulnerable Adults Act 2006 as those who are relatively or absolutely incapable of protecting their own interests, or those in unequal relationships e.g. participants who are subordinate to the researcher(s) in a context outside the research?]

Does the study involve any of the following? [b. The study requires the co-operation of a gatekeeper (defined as someone who can exert undue influence) for initial access to the groups or individuals to be recruited e.g. students at school, members of a self-help group, or residents of a nursing home? NB. The IoN & School of Psychology volunteer pools are not considered gatekeepers in this case.]

Does the study involve any of the following? [c. It is necessary for participants to take part in the study without their knowledge and consent e.g. covert observation of people in non-public places?.]

Does the study involve any of the following? [d. Deliberately misleading participants in any way?]

Does the study involve any of the following? [e. Discussion of sensitive topics e.g. sexual activity or drug use?]

Does the study involve any of the following? [f. The administration of drugs, placebos or other substances (e.g. food substances, vitamins) to the study participants.]

Does the study involve any of the following? [g. Invasive, intrusive or potentially harmful procedures of any kind?*]

Does the study involve any of the following? [h. Obtaining blood or tissue samples?*]

Does the study involve any of the following? [i. Pain or more than mild discomfort?]

Does the study involve any of the following? [j. Psychological stress, anxiety, harm or negative consequences beyond that encountered in normal life?]

Does the study involve any of the following? [k. Prolonged or repetitive testing i.e. more than 4 hours commitment or attendance on more than two occasions?]

Does the study involve any of the following? [l. Financial inducements (other than reasonable expenses and compensation for time)?]

Data (I)

Does the research involve the viewing, usage or transfer of Sensitive Personal Data as defined by the Data Protection Act 1998 or data governed by statute such as the Official Secrets Act 1989 / Terrorism Act 2006, commercial contract or by convention e.g. client confidentiality? (If you are unsure please tick YES and complete the sub-questions).
No [N]

Environment (I)

Will the study cause direct or indirect damage to the environment or emissions outside permissible levels or be conducted in an Area of Special Scientific Interest or which is of cultural significance?
No [N]

International Projects (I)

Will the research be conducted outside of the European Economic Area (EEA) or will it involve international collaborators outside the EEA?
No [N]

Next Steps

Based on your responses your project has been categorised as (ethically) low risk and no further review is required before you start work. You will receive a formal approval email on submission of this form. Should your project change you may need to apply for new ethical approval.

Supporting Documentation

Please upload any documents (not uploaded elsewhere in the application) which you think are relevant to the consideration of your application.
consent_form.doc (83.456KB) Consent form - Information%20Sheet%20for%20Personal%20Data%20Interviews.docx (177.045KB) Information Sheet -

filecount - Please upload any documents (not uploaded elsewhere in the application) which you think are relevant to the consideration of your application.
2

Summary and Submission

Thank you for completing the University's Ethical Review Form. Based on your answers the University is satisfied that your project has met its ethical expectations and grants its ethical approval. Please be aware that if you make any significant changes to your project then you should complete this form again as further review may be required. Confirmation of this decision will be emailed to you. Please complete the declaration to submit your application.

Declaration

I certify that:

[the information contained within this application is accurate.]

Yes [Y]

Thank you for completing the University's Ethical Review Form. Based on your answers the University is satisfied that your project has met its ethical expectations and grants its ethical approval. Please be aware that if you make any significant changes to your project then you should complete this form again as further review may be required. Confirmation of this decision will be emailed to you. Please complete the declaration to submit your application.

Declaration

I certify that:

[the research will be undertaken in line with all appropriate, University, legal and local standards and regulations.]

Yes [Y]

Thank you for completing the University's Ethical Review Form. Based on your answers the University is satisfied that your project has met its ethical expectations and grants its ethical approval. Please be aware that if you make any significant changes to your project then you should complete this form again as further review may be required. Confirmation of this decision will be emailed to you. Please complete the declaration to submit your application.

Declaration

I certify that:

[I have attempted to identify the risks that may arise in conducting this research and acknowledge my obligation to (and rights of) any participants.]

Yes [Y]

Thank you for completing the University's Ethical Review Form. Based on your answers the University is satisfied that your project has met its ethical expectations and grants its ethical approval. Please be aware that if you make any significant changes to your project then you should complete this form again as further review may be required. Confirmation of this decision will be emailed to you. Please complete the declaration to submit your application.

Declaration

I certify that:

[no work will begin until all appropriate permissions are in place.]

Yes [Y]

Faculty of Science, Agriculture & Engineering

Newcastle University

CERTIFICATE OF ETHICAL APPROVAL

Project #: 19-BOW-047

Project Title: Exploring the Power of Personal Data Interaction

This certificate confirms that the application made by **Alex Bowyer (PGR student in Computing)** was **APPROVED** on 16/08/2019

*It is the responsibility of the applicant to ensure that any conditions of approval are fully met before proceeding with the research. Applicants are also required to notify the Faculty Ethics Committee (sage.ethics@ncl.ac.uk) if they wish to make any changes to the design/methods/participants of the study **before** commencing with any changes.*

Signed:  16th Aug 2019

Subject: FW: Ethical re-approval for Exploring the Power of Personal Data Interaction study
Date: Tuesday, 23 August 2022 at 18:18:42 British Summer Time
From: Alex Bowyer (PGR)

From: Alex Bowyer (PGR)
Date: Thursday, 23 April 2020 at 16:30
To: SAGE Ethics
Subject: Re: Ethical re-approval for Exploring the Power of Personal Data Interaction study

Thank you for the speedy response.

I agree to these conditions and am grateful to be able to note the ethical re-approval of my study.

Many thanks
Alex Bowyer

--
Alex Bowyer
Digital Civics PhD Researcher: Empowering People through Data
Open Lab, Newcastle-upon-Tyne

From: SAGE Ethics
Date: Thursday, 23 April 2020 at 15:46
To: Alex Bowyer (PGR)
Subject: Re: Ethical re-approval for Exploring the Power of Personal Data Interaction study

Dear Alex,
Thank you for updating us about the changes to your study; I hope that your studies haven't been too significantly affected by the changes you've had to make.

The amendments you have made to accommodate remote research all look to be in line with the guidance we have issued, therefore please accept this email as approval to continue with the methods you have outlined below.

The only extra condition we will stipulate is that you use passworded meetings when setting up your Zoom calls and ask participants to state an id number (Which you can keep in a separate encrypted file against their names) instead of any identifiable information.

Provided you are happy to accept this condition, please feel free to proceed as planned.

Best wishes,
Rachel

Dr Rachel Collum, Faculty Research Coordinator

From: Alex Bowyer (PGR)
Sent: 23 April 2020 14:56
To: SAGE Ethics
Subject: Ethical re-approval for Exploring the Power of Personal Data Interaction study

Hi,
I have just been forwarded the 'Research Ethics During Covid-19' University guidance and have just learned that my study needs to be re-approved for the adapted non-face-to-face version of the study. Proposed changes to adapt to COVID-19 are attached

Please could you provide ethics re-approval for these changes.

Many thanks,
Alex Bowyer

--
Alex Bowyer
Digital Civics PhD Researcher: Empowering People through Data
Open Lab, Newcastle-upon-Tyne

Index of Key Ideas, Insights and Contributions

This section serves as an index to easily locate key HDR-related concepts, ideas and contributions that this thesis contains.

The Six Wants in Data Relations

Direct Data Relations:

- Data Visibility - 6.1.1
- Data Understanding - 6.1.2
- Data Useability[11] - 6.1.3

Indirect Data Relations:

- Process Transparency - 6.2.1
- Individual Oversight - 6.2.2
- Involvement in Decision Making - 6.2.3

HDR Objectives

The objectives are introduced in 7.5. Their obstacles are explored in Chapter 8, and their opportunities in Chapter 9 and in 10.1.

- Data Awareness & Understanding - 8.1
- Data Useability[11] - 8.2
- Data Ecosystem Awareness & Understanding - 8.3
- Data Ecosystem Negotiability - 8.4
- Effective, Commercially Viable and Desirable HDR Systems - 8.5

HDR Obstacles

The obstacles are collectively shown in Figure 8.1 and are explained in the following subsections of Chapter 8:

- Closed, Insular and Introspective Practices - 8.5.2
- Complex and Invisible Personal Data Ecosystems - 8.3.1

- Diminishing Individual Agency - 8.4.2
- Data Holder Hegemony - 8.4.1
- Immobile Data - 8.2.1
- Inaccessible Data - 8.1.1, 8.2.1
- Insufficient Machine Understanding of Human Data – 8.5.5
- Intractable Data Self, the - 8.4.3
- Invisible Data - 8.1.1
- Lack of Metadata - 8.3.2
- Lack of Individual Demand - 8.5.1
- Lack of Interoperability - 8.5.4
- Lack of Provider Investment - 8.5.3
- Non-Interrogable Data - 8.2.1
- Personal Data Diaspora, the - 8.1.2
- Unmalleable Data - 8.2.1
- Unrelatable Data - 8.1.1

The HDR Principles for Generative Action

- Principle 1 - Life Information Makes Data Relatable. – 10.1.1
- Principle 2 - Data Needs to be United and Unified. – 10.1.2
- Principle 3 - Data Must be Transformed into a Versatile Material. – 10.1.3
- Principle 4 - Ecosystem Information is an Antidote to Digital Life Complexity. – 10.1.4
- Principle 5 - We Must Know Data's Provenance. – 10.1.5
- Principle 6 - Data Holders use Four Levers of Infrastructural Power. – 10.1.6
- Principle 7 - Human-centred Information Systems Must Serve Human Values, Relieve Pain and Deliver New Life Capabilities. – 10.1.7
- Principle 8 - We Need to Teach Computers to Understand Human Information. – 10.1.8
- Principle 9 - Individual GDPR Requests can Compel Companies to Change Data Practices. – 10.1.9

- Principle 10 - Collectives can Compare and Unify their Data and Use it to Demand Change. – 10.1.10
- Principle 11 - Automating the Identification of Entities can enhance Machine Understanding and Unburden Life Interface Users. – 10.1.11
- Principle 12 - The 'Seams' of Digital Services need to be identified, exploited and protected. – 10.1.12
- Principle 13 - It is Possible (and Necessary) to Demonstrate Business Benefits of Transparency and Human-centricity. – 10.1.13

HDR Approaches

The four approaches are collectively summarised in 9.7, and explained and illustrated as follows:

- Approach 1 - Discovery-Driven Activism – [9.3, Figure 9.2]
- Approach 2 - Building the Human-centric Future – [9.4, Figure 9.3]
- Approach 3 - Defending User Autonomy and Hacking the Information Landscape – 9.5; Figure 9.18]
- Approach 4 - Teaching, Championing and Selling the HDR Vision – [9.6, Figure 9.20]

Thesis Contributions

An Understanding of what People want from Personal Data

- 6.1 - An understanding of What People Want in Direct Data Relations
- 6.2 - An Understanding of What People Want in Indirect Data Relations
- 1.2.1 and Figure 4.1 - Enumeration of supported families' and support workers attitudes and needs around civic data
- 1.2.1 and Table 5.4 - An Understanding of What People Want in Data relations in a commercial / GDPR context
- ARI4.1 - A model to understand family civic data
- 5.4.4 - Evidence for the impact of knowledge about data handling practices on provider trust and perceived individual power
- Table 5.2 - A model to understand the five different origins of held personal data

Methodologies for Participatory Work around Personal Data

- Figure 3.6 and [Bowyer *et al.* (2018); Appendix A] - Data Cards and Family Design Games
- 1.2.2 and ARI4.3 - Storyboarding Action Cards
- 5.2.2 - A Methodology for Qualitative Interviews that Explore a User's Personal Data (and see also the digipower report (Bowyer, Pidoux, *et al.*, 2022))

Best Practices and Design Guidelines for Systems and Processes involving Personal Data

- 4.3.4 and 4.5.3 - *Shared Data Interaction* - A proposed model for more efficient and empowering social support relationships that embraces human-centricity
- 5.6 - Recommendations for Policymakers, Service Providers and Individuals on Data Use
- 1.2.3 and 9.6.1 - A Reframing of Data Literacy for the Sociotechnical Context

Principles for Generative Action towards Better Data Relations

See above.

A Detailed and Actionable Research Agenda and Strategy for Empowerment and Systemic Change

- Chapter 7 - The Synthesis and Formulation of the Research Agenda of Human Data Relations (HDR)
- 7.4 - A clear delineation of two primary motivators for individuals seeking better HDR
- Figure 8.1 - A map of the HDR landscape, identifying obstacles and insights
- 9.3, 9.4, 9.5 and 9.6 - Four identified trajectories for advancing Human Data Relations

Major Concepts of This Thesis

- **Auditing Data Holders** – 9.3.2
- **Categories of Family Civic Data** - ARI4.1
- **Categories of Personal Data** - Figure 3.7

- **Data Access & Understanding Services** – 9.3.4, 9.5
- **Data Literacy in an HDR Context** - see HDR Literacy
- **Data Cards** - Figure 3.6, 3.4.2, 4.3.1, 4.5.2, Bowyer *et al.* (2018)
- **Data Wants** - Chapter 6
- **Data Relations, Direct** - 6.1
- **Data Relations, Indirect** - 6.2
- **Digital Self Curation** - 4.4.3, 5.6.2, 6.3
- **Ecosystem Detection** - Principle 4 [10.1.4]
- **Ecosystem Information** - 7.4.2, Principle 4 [10.1.4]
- **Ecosystem Information Display** - Principle 4 [10.1.4]
- **Ecosystem Negotiability** - 7.5, 8.4
- **Empowerment (in the context of data wants)** - 6.3
- **Free Data Interfaces** - Bowyer (2018b)
- **Hacking the Seams** - Principle 12 [10.1.12]
- **Human Data Relations (HDR)** - 7.2
- **HDR Literacy** - 9.6.1
- **Human Information Operating System** - 8.2
- **Human Information** - see Life Information and Ecosystem Information
- **Inclusive Data Flows** - 4.5.1
- **Information Standards** - 5.6.1, 10.1.8
- **Landscape of HDR Opportunity** - Figure 8.1, Figure 9.22
- **Life Concepts** – 10.1.1
- **Life Information Utilisation** - 7.4.1, Figure 7.1
- **Life Information** - 7.4.1, Principle 1 [10.1.1]
- **Life Interface Design** - 9.4.1
- **Life Partitioning** - 9.4.2
- **Locus of Decision Making (LDM)** - 4.5.3
- **Meaning in Data** - Figure 2.1, 4.4.2, 5.5.2, 6.1.2, Principle 1 [10.1.1]
- **Perceived Individual Power** - 5.4.4

- **Personal Data Diaspora, the** - 8.1.2
- **Personal Data Ecosystem Control** - 7.4.2, Figure 7.1
- **Personal Data as a Proxy for Involvement** - 5.5.4, Bowyer *et al.* (2018)
- **Personal Data Stewardship** - 4.4.3, 5.7
- **Proxy Representations of Immobile Data** - Principle 4 [10.1.4]
- **Shared Data Interaction** - 4.3.4
- **Surface Information Injustices** - 9.6.1, Principle 12 [10.1.12]
- **Storyboarding Action Cards** - ARI4.3
- **Trust** - 4.4.4, 4.5.1, 5.4.4, 5.5.4, 6.2.1, Principle 13 [10.1.13]
- **Types of Personal Data (by origin)** - Table 5.2
- **Useability**[11] (as distinct from Usability) - 6.1.3

Glossary of Pre-Existing Abbreviations, Names and Terms

This section serves as a quick-reference glossary to explain abbreviations and existing terms used.

- **Action (stage of Personal Informatics)** - see SI
- **Action Research** - a mode of research where cycles of investigation shape future studies
- **Accessibility Tags (ARIA)** - tags within HTML code that screenreaders use and which can be exploited for seam hacking
- **Activism** - using vigorous campaigning to bring about political or social change
- **Agency** - the ability, described in HDI, to act for oneself in a system, see HDI
- **Barriers Cascade** - a series of obstacles in SI
- **BBC R&D** - the Research & Development division of the British Broadcasting Corporation, where I did a research internship
- **Boundary Objects** - tangible objects and representations that help different populations that may think in different terms to collaborate effeectively
- **Card Sorting** - a technique used in Participatory co-design where participants arrange cards to convey their thinking
- **CHC - Connected Health Cities** - government initiative behind the SILVER project, which I worked on
- **Civic Hacking** - technologists or enthusiasts working to reconfigure the way society works
- **Civil Libertarianism** - argues for the supremacy of individual rights and personal freedoms over imposition by authority
- **Collection (stage of Personal Informatics)** - see SI
- **Consent, Dynamic** - ongoing and changeable expression of one's preferences
- **Consent, Informed** - initial one-time expression of one's preferences
- **Constructivism** - a belief that new knowledge is formed by developing one's own mental models in order to explain new experiences

- **Conceptual Anchors** - the mental scaffolds which we use to organise our thinking and human information
- **Context-aware Computing** - designing systems that take account of the user's situation and varying needs
- **Co-experience** - bringing participants towards a shared perspective
- **Cornmarket** - codename for the BBC R&D PDS project I interned with
- **Critical Algorithm Studies** - research into making computer systems and their behaviour more understandable and accountable
- **Data Access Request** - see Subject Access Request
- **Data Brokers** - third parties that buy and sell user's data
- **Data Controller** - an organisation responsible for collecting and storing user data
- **Data Download Portal** - a website or service that allows users to access held data in a 'self-service' manner
- **Data Flow Auditing** - the use of apps or system services to monitor the communication and data sharing behaviours of consumer apps
- **Data Justice** - research into the relationship between datafication and social justice
- **Data Portability Request** - a particular kind of access request focussed on retrieving data in a machine-readable and useable format
- **Data Processor** - one who handles user data on behalf of a data controller
- **Data Provenance** - the history and origin of a piece of data
- **Data Self** - the representation of an individual in data through which state or commercial actors understand that person when making decisions
- **Data Subject** - the individual about whom data is stored
- **Data Transcendence** - the idea that data should not be tied to a single machine, but should move freely to the places it is needed
- **Data (general)** - digitally-encoded human information
- **Data, Civic / Family Civic** - the data stored about families by social care organisations such as Early Help programmes
- **Data, Acquired** - personal data that has been obtained from an official or public source or a third party

- **Data, Derived** - new data that has been extrapolated through interpretation of existing data
- **Data, Metadata** - data about the data itself, or about the incident recorded in data
- **Data, Observed** - data collected about individuals automatically, as a by-product of other actions or in the background
- **Data, Personal** - data about or related to identifiable individuals
- **Data, Volunteered** - personal data that has been knowingly shared by that individual with an organisation
- **Data, Trapped** - data that is hard to access due to technical, commercial or other restrictions
- **DERC** - Digital Economy Research Centre, funders of the Healthy Eating Web Augmentation project
- **Design, Co-** - researchers and participants collaboratively exploring problems and solutions
- **Design, Adversarial** - the design of systems or processes that challenge current norms
- **Design, Design After** - the ability to repurpose products or services for new objectives that might not have been initially considered
- **Design, Disrespectful** - design practices (often in user interfaces) that prioritise commercial needs over human convenience
- **Design, Experience-centred** - design that focusses on human psychology and lived experience of the situation
- **Design, Magical** - the presentation of technology offerings as powerful and mysterious, that need not be examined or understood
- **Design, User-centred Design** - design approach that builds up an understanding of user needs that is subsequently used by technical experts
- **Design, Value-centred Design** - designing in ways that focus on positive impacts on human life
- **Device Tenancy** - the idea that we are no longer owners of our devices, and lack control and autonomy over their use

- **Digital Civics** - a multi-disciplinary research field encompassing HCI, governance, education, planning, social science and computer science, practiced by - Open Lab, where I studied this PhD
- **Digital Self** - see Data Self
- **DIKW pyramid** - see Wisdom Curve
- **DPA** - Data Protection Authority** - the official authority for regulating data use in a given country
- **DPO** - Data Protection Officer** - the individual legally responsible for managing data use within a company and for handling access requests
- **EPSRC** - Engineering and Physical Sciences Research Council, funders of this PhD research
- **Early Help** - a set of programmes in UK local authorities designed to identify 'at risk' families and help them with targeted interventions
- **Effective Access** - ensuring that individuals have all necessary skills, systems and capabilities necessary to see and understand information
- **Embodied Interaction** - allowing users to create their own practices in information interaction
- **Empowerment in Use** - having freedom to use products and services in the way you want
- **Entities** - people, places, organisations, brands, topics or other identifiable 'things' that could be a stakeholder of, or related to, a piece of data
- **Entity Extraction** - the process of identifying real-world entities in data
- **Explainable AI** - algorithms whose decision making processes are described to system users
- **Faceted Search** - the ability to search information by its shared aspects
- **File Biography** - the lifetime of past actions on a computer file
- **Files, why they need to die** - article by myself (Bowyer, 2011)
- **GDPR** - General Data Protection Regulation - the EU's 2018 regulations that give users rights over the collection and use of their personal data
- **Gatekeeper** - One who controls the flow of data or information between an organisation and an individual

- **HCI - Human Computer Interaction** - research and practice that explores how people relate to and use computer systems
- **HDI - Human Data Interaction** - a sub-discipline of HCI that focuses on people's relationship with data, rather than with the system
- **Hestia.ai** - Swiss company working in the data access and understanding services space, which I currently work for
- **HII - Human Information Interaction** - a discipline in library sciences that considers how humans relate to information regardless of technology used
- **Humane Technology** - a movement focused on making technology that is more sensitive to people's lives and needs
- **ICO - Information Commissioner's Office** - the UK's Data Protection Authority
- **Ideation Deck** - a participatory design technique that uses 'ingredients' cards shuffled in a grid to generate new ideas
- **Individualism** - the pursuit of one's own objectives as a primary objective
- **Information (general)** - facts and assertions understood by interpreting data
- **Information, Human** - information about people that can be related to their lives or to their digital world
- **Information, Life** - information about people's lived experience that can be found within data
- **Information, Ecosystem** - information about people's data, where it is stored, and how it is used and shared
- **Information Landscape** - the general terrain of available information that a user can see and interact with through the services and apps they use
- **Infrastructural Power** - see Power, Infrastructural
- **Integration (stage of Personal Informatics)** - see Self Informatics
- **Interoperability** - getting systems to connect and exchange information through data standards or conversion
- **Interoperability, Adversarial** - making systems connect together in ways that were not intended by manufacturers
- **Legibility** - the ability, as defined in HDI, of being able to understand stored data

- **Lifelogging** - the practice of maximal data capture for personal SI benefit
- **Lenses** - different ways of focussing on some data or information according to the aspect of interest or the current role
- **Life Sketching** - a process of mapping out mental models of one's life on paper
- **MyData** - an organisation whose members pursue a human-centric change agenda
- **Negotiability** - the ability described in HDI to flexibly adapt and change one's preferences as the world or digital system changes
- **NER - Named Entity Recognition** - see Entity Extraction
- **Open Lab** - the research lab in Newcastle University in which I conducted this PhD research
- **Orienteering** - an associative process of information-finding
- **PDS - Personal Data Store** - See Personal Data Lockers
- **PIM - Personal Information Management** - the 1990s/2000s discipline that focused on new ways to manage and interact with data and information
- **PIM systems, contextual** - PIM systems that organise information according to what context it relates to
- **PIM systems, networked** - PIM systems that focus on the relationships between different pieces of information
- **PIM systems, semantic** - PIM systems that focus on the underlying meaning of the stored data
- **PIM systems, spatial** - PIM systems that focus on arranging data in a virtual space for easier management
- **PIM systems, subjective** - PIM systems that focus on the varied individual needs of users
- **PIM systems, temporal** - PIM systems that represent information using timelines or other visualisations that highlight change over time
- **PIMS** - Personal Information Management Services** - See Personal Data Lockers
- **PDE - Personal Data Economy** - the emergent marketplace of companies innovating and offering services relating to the management, self-exploitation or -harnessing of one's personal data

- **Personal Data Ecosystem** - the network of systems, accounts, files and digital information that constitutes an individual's digital life
- **Personal Data Lockers** - a place to store personal data so that it can be united, unified and interpreted by the data subject
- **Personal Data Vault** - See Personal Data Lockers
- **Personal Informatics** - see SI
- **Preparation (stage of Personal Informatics)** - see SI
- **Perspectives** - different presentations or aspects of information that support different mental models, focus or tasks
- **Point of Severance** - the point at which data is handed over, beyond which data subjects lose visibility, control and influence
- **Power - Behavioural Influence** - persuading others to carry out the desired behaviour
- **Power - Interpretative Influence** - determining how reality is externally represented
- **Power - Network Centrality** - becoming an indispensable hub of a wider ecosystem
- **Power, Authority** - ownership of technology or infrastructure
- **Power, Disciplinary** - using an influential position to affect others' mental models
- **Power, Infrastructural** - a model of understanding how providers exert power over their users, created as part of the digipower investigation
- **Power, Interpretive** - creating the internal representations of reality within an organisation
- **Power, (power to)** - an individual's ability to act (see Agency)
- **Power, (power over)** - a dominant actor's ability to limit or manipulate the actions of others
- **Power, Obscure** - where the subservient cannot tell when they are watched
- **Power, Pervasive** - where the one in power can see everything all the time
- **Power, Processual** - changing processes for competitive advantage
- **Power, Rational** - controlling decision-making processes
- **Power, Resource Control** - controlling the flow of resources

- **Power, Social** - power where the power holder attempts to influence the behaviour of individuals in pursuit their desired outcomes
- **Power, Socially-shaped** - influencing a wide audience to settle upon a preferred interpretation
- **Power, Systems/Structural** - see Infrastructural Power
- **Power, Zero Sum** - winning a battle for ownership/resource control at the other party's expense
- **Power Imbalance (over Personal Data)** - the established fact that data holders have more power in service relationships than data subjects.
- **Pragmatism** - an epistemology that believes knowledge is constantly renegotiated by individuals
- **QSM - Quantified Self Movement** - see SI
- **R&D** - Research & Development
- **Reflection (stage of Personal Informatics)** - see SI
- **Recursive Public** - a community of people who are attempting to reconfigure society for the better
- **SAR - Subject Access Request** - a request to a DPO of an organisation for a copy of held personal data
- **SI - Self Informatics** - an umbrella term for Personal Informatics and the Quantified Self Movement, where people track their activity in data and reflect - upon it, setting goals and tracking progress
- **SILVER** - the project working in the Early Help space that I worked with as a researcher, informing Case Study One, and as a developer, serving as an embedded project setting informing Section IV.
- **Sitra** - Finnish non-profit research organisation for which the digipower investigation was conducted
- **Scraping** - the process of programmatically extracting information from interfaces such as websites that were intended for human browsing.
- **Seams** - the 'edges' of products and services, at which service providers can exert restrictions and at which users can find new ways to adapt their product - usage and data access

- **Support Worker** - a specialist social worker who helps a family in an Early Help context
- **Supported Family** - a family participating in an Early Help social care programme
- **Text Mining** - the process of programmatically examining textual data to infer new facts and assertions from the data
- **ToC - Theories of Change** - a model for thinking about how to achieve change in society
- **Things to Think With** - the idea that tangible representations can be useful to aid discussions
- **Timelines** - visual representations of information anchored against points in time
- **TrackerControl** - see Data Flow Auditing.
- **Troubled Families** - historic term for those families targeted for help by programmes such as Early Help
- **VRM - Vendor Relationship Management** - a model where vendors are selected by customers in response to their published needs, instead of relying on - broadcast advertising to find customers
- **Web Augmentation** - the process of modifying a web page to provide new functionality or access data after it has been downloaded to a user's we browser
- **Web Extensions** - pieces of user code that are loaded into a web browser to modify or programmatically interrogate web pages
- **Wisdom Curve** - the process of converting data, to information, to knowledge, to wisdom
- **world2vec** - an internal system in Meta (Facebook) that attempts to understand the world through analysis of social media content, for commercial advantage