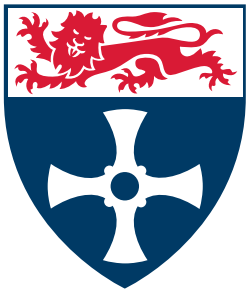
Democratising Data Science

Effective Use of Data by Communities for Civic Participation, Advocacy and Action



Aare Puussaar

Supervisors: Prof. P. James
Dr. K. Montague
Prof. P. C. Wright

School of Engineering
Newcastle University

This dissertation is submitted for the degree of
Doctor of Philosophy

October 2020

To my grandmother.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Aare Puussaar
October 2020

Acknowledgements

I have been fortunate to be guided by amazing tutors and supervisors over the years of my studies. Thanks to Prof. Dr. Patrick Olivier for giving me the opportunity, Prof. Dr. Thomas Plötz for guiding me in the early days of the PhD, Dr. Adrian Clear for helping me choose my academic path, Prof. Emer. Dr. Pete Wright for helping me think like an academic and for our stimulating discussions over the years, Prof. Phil James for pushing me to get stuff done and supporting me in my decisions, to Dr. Kyle Montague for being a never ending source of advice and for listening to me go off on tangents in the last phase of the thesis. Special thanks to Dr. Jan Smeddinck for all the support and guidance for my work.

I would love to thank all my colleagues (past and present) in Open Lab who helped and inspired me over the years. Thanks Andy G., Tom N., Rob A., Ed J., Seb M., Dave C., Shauna, Dalya, Samantha, Clara, Vasilis, Nils, Cassim, Chris E., Rob C., John V., Tom B., Clement, Dan J., and many others. Thanks to the amazing team at the Urban Observatory for keeping up with my parachute working. Thanks Jennine, Luke, Neil, Tom, Dave P., Richard and Rich, and Lindsey.

To my fellow CDTs and soon to be doctors – Reem, Helen R., Nataly, Jen, Janis, Delvin, Jay, Colin, Dan H., Dan R., Stuart N. – thank you for taking this journey with me, look at us now! Also, thanks to the extended CDT and PhD family – Ian, Emma, Sean P., Tom M., Lydia, Andy D., Zander, Tim, Euijin, Tagreed, Gerry and others.

Huge thanks to all my friends, whether close to home or across seas and mountains. Thanks Maarten for becoming my friend when I needed it the most and showing me life outside Estonia. Thanks to Kristjan, Raiko and Sven for the support from afar.

Thank you to my parents, especially my mum for putting a roof over our heads and helping me start off my education and adult life. Suur aitäh teile! To my big sister, thank you for your encouragement and for inspiring me to pursue a university degree. I hope I can repay the favour with this one. Also huge thanks to my new family members, Ann and Stuart, who have showed me love and support throughout this journey. Special thanks to my new sister Helen, for all the encouragement, belief and support.

Finally, most importantly I want to thank Antonia, my love, for all of your understanding and patience. Thanks for always being there for me and helping me overcome all the obstacles, self-doubt and fears. I would not have been able to do it without your never ending support and love.

Abstract

We live in an age of data, where it is being collected and archived in tremendous volumes and at great velocity. Smart cities are a good example of how we generate and use data with the aim of improving the lives of citizens. Cities adopting more technologies and embedding them in the physical fabric of the city will drastically change the way decisions are made in the city, in addition to the way citizens interact with the city. Research to date has predominantly focused on engineering agendas or has narrowly focused on citizens' participation as passive producers of data in the smart city. This thesis takes a more holistic approach by focusing on both the engineering problem-solving agenda and community problem-solving activities. Taking a participatory research approach, the thesis explores such a context through three case studies that involve the design, development and analysis of two Community Informatics (CI) systems. In addition to producing two open-source CI technologies (SenseMyStreet and Data:In Place) for active citizen participation, this study posits a Citizen Advocacy Framework and Community-Data Interaction (CDI) model as novel theoretical framings that enable researchers to discuss and design for the effective use of data by communities. Furthermore, this thesis provides a practical example of the use of CDI for supporting communities to take local action. This improved understanding of the relationship between data and communities demonstrates a better direction for future research and the design of CI technologies as they work towards democratising data science and enabling the effective use of data by communities for active civic participation, advocacy and action.

Research Outputs

Work undertaken during the PhD has resulted in multiple scientific publications and communications. Links to open source software that resulted from the work of this thesis are provided. Relevant publications are acknowledged at the beginning of each chapter. Other contributions might not be discussed in this thesis, but are added for the sake of completeness. They are indicated with an asterisk (*).

Publications in Peer-Reviewed Journals and Conferences

In chronological order:

- ***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 - CHI '17*, pages 6936–6942, Denver, Colorado, USA. ACM Press.
- *Vanhoof, M., Hendrickx, L., **Puussaar, A.**, Verstraeten, G., Ploetz, T., and Smoreda, Z. (2017). Exploring the use of mobile phone data for domestic tourism trip analysis. *Netcom. Réseaux, communication et territoires*, (31-3/4):335–372.
- Johnson, I. G., **Puussaar, A.**, Manuel, J., and Wright, P. (2018). Neighbourhood Data: Exploring the Role of Open Data in Locally Devolved Policymaking Processes. In *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):83:1–83:20.
- **Puussaar, A.**, Johnson, I. G., Montague, K., James, P., and Wright, P. (2018). Making Open Data Work for Civic Advocacy. In *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):143:1–143:20.

Book Chapters

Peacock, S., **Puussaar, A.**, and Crivellaro, C. (in submission). Sensing our Streets: Involving Children in Making People-centred Smart Cities.

Short Papers and Abstracts Contributed to International Conferences

In chronological order:

- ***Puussaar A.** (2017). Citizense Makers: Collective Sensemaking of Data. Data Publics (Lancaster, 2017).
- ***Puussaar A.** (2017). Citizense Makers. TICTeC (Flourence, 2017).

Guest Talks and Seminars

- *__Puussaar A.__ (2018). Democratizing Data Science: Effective Use of Data in Civic Participation and Advocacy. Guest lecture at the International Computer Science Institute (ICSI), UCL Berkeley (Berkely 2018).
- *__Puussaar A.__ (2018). SenseMyStreet: Sensor Commissioning Toolkit for Communities. Invited speaker at the annual Investigation of Air Pollution Standing Conference (IAPSC), (Telford 2018).
- *__Puussaar A.__ (2019). SenseMyStreet: Sensor Commissioning Toolkit for Communities. Seminar at the Computing & Communications Department, Open University (Milton Keynes 2019).

Other Dissemination

- Urban Observatory partnered with NOVAK Collective (2017). Google SPAN international design conference series. BALTIC Centre for Contemporary Art (Gateshead 2017).
- Evans Woolfe Media partnered with SenseMyStreet (2018). Using a congestion charge to control traffic pollution. Geography: The Big Issues, BBC Teach (2018).
- Evans Woolfe Media partnered with SenseMyStreet (2018). Walk to School. BBC Assemblies, BBC Teach (2018).

Open Source Software

- FidasFrogParser. Available at https://github.com/aarepuu/fidasfrogparser
- Data:In Place Platform. Available at https://github.com/aarepuu/data-in.place

Research Impacts

The results of the work undertaken in the PhD have been featured in mainstream media and recognised by others through awards and acknowledgements.

In Press

- Mike Kelly (2017). The smog on the Tyne is all mine - What Newcastle community is doing in fight against pollution. Chronicle Live https://www.chroniclelive.co.uk/news/north-east-news/smog-tyne-mine-what-newcastle-13610648 (Accessed 21/04/2020).
- Pollution Patrol (2017). Residents doing environmental monitoring using sensors from SenseMyStreet toolkit. BBC Look North News (Aired 14/09/2017).

Awards and Acknowledgements

- Bart Gorynski (2018). Rising European Stars: Eight Smart Cities to Watch in 2018. Bee Smart City, https://hub.beesmart.city/strategy/rising-european-stars-eight-smart-cities-to-watch-in-2018 (Accessed 21/04/2020).
- Newcastle City Council partnered with Urban Observatory and SenseMyStreet (2019) Winner of Smart City of The Year. Digital Leaders 100 Awards (London 2019).
- The Urban Observatory (Newcastle University, in collaboration with the universities of Birmingham, Bristol, Cranfield, Manchester and Sheffield) (2019), Shortlisted for Research Project of the Year: STEM. THE Awards (London 2019).

Table of Contents

List of Figures

List of Tables

Nomenclature

Acronyms / Abbreviations

AIT Automated Information Technologies

API Application Programming Interface

ASIK Applied Social Informatics of Knowledge

CAF Citizen Advocacy Framework

CC Community Conversational

CDI Community-Data Interaction

CI Community Informatics

citizens the people who inhabit, work and play in cities

CMS Content Management System

CoP Community of Practice

CRACIN Alliance for Community Innovation and Networking

CSCW Computer-Supported Collaborative Work

CWIS Cycling and Walking Investment Strategy

DEFRA Department for Environment, Food and Rural Affairs

DIKW Data-Information-Knowledge-Wisdom

FB Facebook

FOI Freedom of Information

GeoJSON Geographic JavaScript Object Notation

GIS Geographic Information System

GPS Global Positioning Service

HCI Human-Computer Interaction

ICT Information and Communication Technology

IMD Index of Multiple Deprivation

IoT Internet of Things

IS Information Systems

ISD Information System Design

LSOA Lower Super Output Area

MVP Minimum Viable Product

NHS National Health Service

NPO Non-profit Organisation

OA Output Area

ODK Open Data Kit

OGD Open Governmental Data

ONS Office of National Statistics

PADRE Participatory Action Design Research

POI Point of Interest

PPGIS Public Participation GIS

QFT Question Formulation Technique

RQI Right Question Institute

SI Social Informatics

SMS SenseMyStreet

SPACE Safe Pedestrian and Cycling Environment

Ubicomp Ubiquitous Computing

UKCRIC UK Collaboratorium for Research in Infrastructure & Cities

UO Urban Observatory

VGI Volunteered Geographic Information

WFS Web Feature Service

WMS Web Map Service

Introduction

I don't know where I'm going from here, but I promise it won't be boring.

— David Bowie

1. Prelude

1.1. Data, the City and Citizens

In the era of the information age, libraries no longer hold the largest sources of data accessible to humans. In fact, computer systems store and help generate far more data than can be found in the world's largest library – the Library of Congress. As a result, *big data* – either 'too big' in terms of volume, 'too fast' in terms of generation or a limited processing window (i.e. persisting in state) or 'too complex' in terms of structure or the variety of sources – is becoming the centre of our research, communities, and world. The term *datum*, referring to a singular unit of data, is rarely used any longer and seems to be gradually disappearing from our vocabulary. The plethora of data that can now be obtained by using a new generation of mining methods is becoming increasingly essential in decision-making processes in organisations and governance. Industry, an early embracer of big data, has successfully been using data-driven approaches for many years now in product development and market forecasting to make strategic decisions. In recent years, governments have also begun looking into new methods and sources of data to accompany the traditional ways of polling their citizens, i.e. by conducting censuses, to inform policies, for example, investing in *smart city* infrastructure or new techniques and methods for utilising mobile phone data – which permits spatial mobility analysis on a national scale – to extract early statistical indicators about the population. However, there is a debate to be had about whether these methods are scientifically rigorousness and whether they are actually working towards eradicating inequalities.

For algorithmic smart cities, citizens – the people who inhabit, work and play in cities – are seen as another source of data for management and control at scale. Citizens also participate in the smart city by contributing data – either actively or passively – using different services on their phones or computers. However, there may also be room for a more active participation in the smart city, where citizens are active users of data (i.e. smart citizens) who take part in decision-making processes. This surfeit of data and the instrumentation of technologies that support it has provided food for thought for scholars from a wide range of research disciplines – from technical domains like pervasive computing, datalogy or data science and urban studies

to human-computer interaction (HCI) and computer-supported collaborative work (CSCW). Whether it is concerned with techniques or methods for recording and performing mass analysis of human behaviour or looking at complex human interactions mediated by computer-supported systems, data always has a role to play. Indeed, the private and public sectors, research and the industry have all recognised the great potential of working with data to produce monetisable scientific or social value.

1.2. Democratisation of Data

Whether it is sparse data generated from a sensor attached to a lampost in the city or dense data obtained through a thoroughly planned collection process, its efficiency is determined by its effectiveness to produce value. The ultimate goal for data is to aid people in creating *actionable knowledge* or *wisdom* – often defined as *intelligence*, or in business lingo as *insights* – which will ultimately help people make better decisions. Having access to these vast amounts of data allows humans to create new knowledge and wisdom that can used in action. Although ubiquitous computing (UC) and other information technologies permit people to develop systems that collect, store and retrieve data and information in immense quantities and velocities, such systems will never be able to generate wisdom.

> *It may well be that wisdom – which is essential for the pursuit of ideals or ultimately valued ends – is the characteristic that differentiates man from machines* (Ackoff, 1999).

Ackoff (1989) proposes a Data-Information-Knowledge-Wisdom (DIKW) pyramid, a simplistic idea stating that in order to transform data into actionable knowledge (i.e. wisdom), one has to go through all the steps of the pyramid from the bottom up. In addition, Ackoff (1989) states that data is a raw material (Gitelman, 2013; Tuomi, 1999, cf.) that needs to be moulded into information, knowledge and then wisdom. According to his knowledge pyramid, there is always more data than information, more information than knowledge and more wisdom (i.e. knowledge that can be applied in action) than knowledge (Jennex and Bartczak, 2013, cf.). However, there have been different takes on the knowledge pyramid (Jennex and Bartczak, 2013; Tuomi, 1999). Tuomi (1999) states that data is not, in fact, the building block (or foundation) for the pyramid because it is not observed, collected or recorded in a vacuum, so it can therefore not be a raw material. Jennex and Bartczak (2013) argue that rather than being one or the other, it is both at the same time, in addition to also pointing out that in learning processes, the pyramid acts more like that of Ackoff (1989), whereas in organisational structures and knowledge-making, it acts like the top-down model of Tuomi (1999). These theories were born from fields such as business intelligence, systematic thinking, systems sciences and engineering, and organisational memory management. However, we are now working with data in a completely different context, where such rules of organisational structure do not apply.

Regardless of the inability of computer systems to create wisdom (i.e. applied knowledge), they are still a crucial part of using data in action. However, a lot of the tools for generating,

accessing, interpreting and using data are still catered to professionals and are often driven by governmental initiatives or commercial endeavours, which means that the format of data and available tools for communities for using data in knowledge-making processes and in action are often constrained to the same organisational knowledge network (Hakken, 2003) they come from. In order to use such data and information systems (IS) to support knowledge-making in communities, the latter need to possess a similar knowledge base as the systems developers or data publishers. However, the *tacit knowledge* (Polanyi, 2009) that engineers possess and embed into expert systems is very difficult or nearly impossible to communicate to people who may benefit from using such systems, which means that citizens are currently in need of support from professionals in order to access, interpret and use these datasets. Furthermore, it is easy to think of knowledge as something individualistic in the head of an individual, but because many aspects of civic advocacy require collective action, there is a need to share knowledge and distribute information. In order for it to become *focal knowledge*, it needs to come from communities themselves by constructing and/or deconstructing data in context and in situ (Suchman, 1985) using tools and methods designed with people from those specific communities with their purposes in mind.

The value of data and its usefulness for citizens to influence governance and decision-making is still unclear, mainly because it is underutilised by communities due to the lack of access, understanding and resources to support data science practices. Such practices and skills are essential for knowledge creation and value generation in this data-saturated world. However, enabling access to new and improved tools will not be enough to solve the problem of inequality and democratise data science – it is also crucial for research to pursue deeper understandings regarding the role of data in these contexts. Working together with communities, researchers need to identify caps and barriers and help increase data literacy in order to empower communities to create knowledge they can put into action. The present thesis seeks to do exactly this by investigating how citizens can effectively use data in civic participation, advocacy and action.

2. Aims and Objectives

This research explores the use of data within civil society and examines how citizens can use data in processes that contribute to activities that result in local benefit. The main aim of the thesis is:

To explore tools and methods for the effective use of data by citizens for civic advocacy and action.

From this, the underlying objectives of this thesis are:

i. (*Conceptual objective*): Through the critical lens of engineering (i.e. information theory), epistemology and social studies, develop a conceptual model and understanding of 'what it means to make data usable for the purposes of civic advocacy and action'.

ii. (*Technical objective*): Design and build tools that enable and promote the use of data in civic participation and advocacy. This involves understanding the context, culture, values and practices of different community groups working with data in order to adequately respond to their needs and the issues they are tackling. This also includes establishing processes that will help build capacity within the community so as to sustain the support for community groups working with data.

iii. (*Pragmatic objective*): Provide a model of practice for interactions with data in knowledge networking at the community level that can be applied when designing and building tools for the effective use of data by citizens for civic participation, advocacy and action.

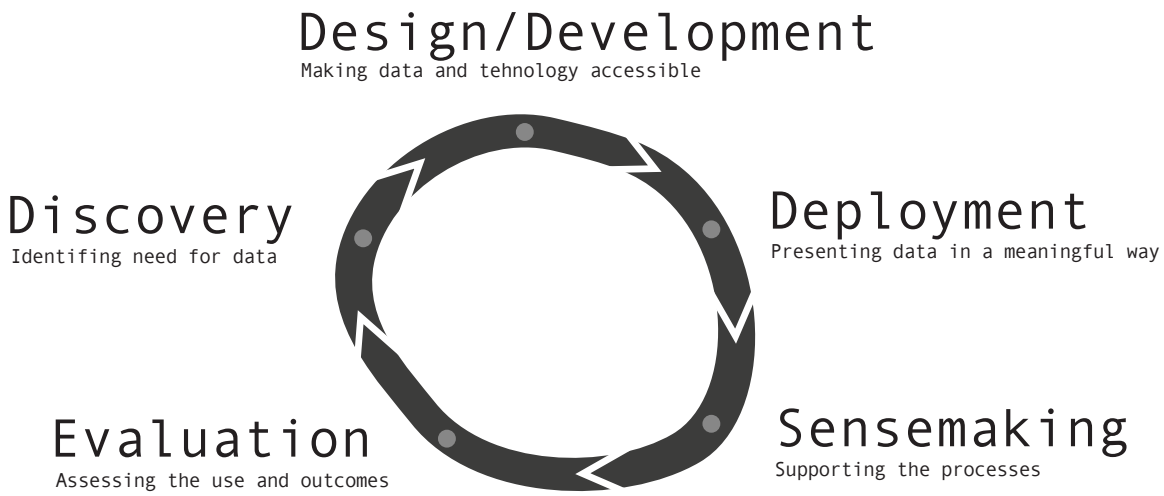
2.1. Research Questions

The following research questions have been developed to further study this area and address the aims and objectives of the thesis:

1. What is involved in citizens creating actionable knowledge using data?
2. How do we build and configure tools and processes to support knowledge-making and enable the effective use of data by citizens in civic advocacy and action?

2.2. Research Approach

The availability of different datasets and capabilities of information technologies helps us to rethink the role of data in civic participation and in complex socio-technical processes. Furthermore, it provides opportunities to focus on digital technology and mechanisms that support human-centred and alternative approaches. To answer the proposed research questions and to understand what makes data useful for communities, this thesis has adopted a case study approach to explore barriers and opportunities for the effective use of data by citizens for civic participation, advocacy and action. Hence, the level of participation that this thesis is aiming



Source: Author

Figure 1 Illustration of the five steps in the thesis' guiding methodology

to support is very active, even activist, in nature, wherein citizens become active producers, consumers and users of data for civic action.

The research conducted in this thesis was *transdisiplinary* (Ferreira et al., 2020; Madni, 2007) in nature, meaning that the study aimed to address both the engineering problem-solving agenda and community problem-solving activities (Stillman and Denison, 2014). In addition, the research was conducted through the lens of a Participatory Action Design Research (PADRE) (Haj-Bolouri et al., 2015, 2016) approach, which manifested itself in user-centred design processes wherein iterative design, and more engineering-focused design sprints, were applied. The research focused on the active involvement of people in the design processes so as to not only identify their particular needs and then respond with technical solutions but also encourage citizen-led inquiries and activities that would drive the research forward. After each iteration, there was a feedback loop from the community, which then dictated the next steps and refinement of the developed technologies. This process involved frequent field visits, ethnographic work, engagements, focus groups and workshops with communities, followed by design and development that resulted in technical tools, platforms and research methods. Throughout the case studies, the following methodology (Figure 1) was applied to study the role of data and technology in knowledge creation for meaningful civic participation and community decision-making:

- Identifying the need for data (*Discovery Phase*). Through conducting interviews and observations of existing communities or groups of individuals, to map out how they are using or not using data.

- Making data and technology accessible (*Design/Development Phase*). Undertaking a user-centred design process to create tools and platforms that respond to the needs of the

community – either for data access or production, starting with paper-based methods and then moving to digital tools where appropriate.

- Meaningful presentation of data (*Deployment Phase*). Ascertain what visualisations and formats of data are more accessible and useful for the sensemaking (i.e. interpretation and comprehension) processes in different contexts and with different communities.
- Taking collective action by purposing data (*Sensemaking Phase*). Supporting processes of sensemaking and knowledge creation by local communities to use data for decision-making and develop new thinking and solutions.
- Assessing the outcomes of data use (*Evaluation Phase*). Evaluating the use of data for improving civic participation and applying it to activities for local benefit.

The contexts and the communities with which the research was conducted differed in terms of proficiency of data; existing community networks, links and capacity; social, economic and cultural backgrounds; and social capital:

- local groups of active residents in Newcastle Upon Tyne (Newcastle, UK) that were leveraging environmental monitors to produce data for campaigning and local decision-making;
- a neighbourhood-planning group in North Tyneside (UK) that was exploring the use of data to inform the process of local policy-making; and
- two charity organisations in Newcastle (Charity A) and North Tyneside (Charity B) that were reliant upon data to focus their work on the needs of the community and to illustrate their impact in order to extend their funding.

Broadly, these research efforts can be linked to different modes of active participation though data: *making or producing data*; *finding or accessing data*; and lastly *interpreting and using data*. The design process and the developed technologies were directly derived from working with these communities and were translated into technical outputs by the researcher. Although the research was heavily dependent upon the engagement activities of the research team, the aim was to identify the processes and build the required technology that would sustain engagement without the lead from the researchers in longitudinal deployments. Throughout the research, there was a constant dissembling of the role of the professional data analyst (i.e. the thesis author), scoping the boundaries for how far the expertise of the professional could be removed, knowledge handed down to citizens, or the extent to which it could be replaced by intelligent systems. This included experimenting with processes of guiding issue formulation, eliminating barriers of data access and production, and enabling communities to take ownership of data and self-organise to take action around issues. Furthermore, having these technologies and processes implemented outside 'the lab' in longitudinal deployments and accessible online enabled study of their use and adoption by different communities in order to understand how to refine them

to remove barriers of engagement and work towards more inclusive technologies that support citizens in taking action on local matters.

This thesis embraced the messiness of society and conducting research outside the laboratory environment (i.e. 'in the wild') with a wide range of stakeholders, which provides no guarantee of successful outcomes because of the multiplicity of uncontrolled variables for success. Hence, the lens through which the case studies have been analysed and reflected upon is mainly empirical and qualitative in nature, focused on observations within the communities and consisting of both communities' and the researcher's reflections on the impact of created tools and platforms with regard to helping empower people use data to take action.

The Issue of The 'Black Box'

The thesis, and the research therein, was written and conducted from the perspective of an information engineer who trained and practised as a data scientist working on big data analytics for a decade. Working with big data over the years led the author to question the ethical preconditions for conducting such studies on passive datasets from the population. Furthermore, a sympathetic reading of works from the social and human sciences and conducting research with real people 'in the wild' helped to inform new understandings of data and people that extend across disciplines. Although no surveys of researchers were conducted, it appears that many of the people working in these social contexts were from design or humanities backgrounds, which often means that there is a lack of in-depth understanding about how technology works, so much so that it is often then seen as a 'black box'. Such a situation can foster resistance to technology and may be seen as giving away control from the human to the machine. However, within the field of HCI, we are seeing more and more people from mixed disciplines crossing over and working together to develop new thinking around these issues. Similarly, the unique skill set of the author enabled a full understanding of the technology and ways that it could be used or reconfigured to help citizens achieve more control over local matters.

3. The Case Studies

In order to explore the research questions and address the aims and objectives of the thesis in a systematic way, a series of case studies were conducted, as outlined below:

Case Study I (in Chapters 2 and 3) engaged people with smart city technology and gave them the power to use the technology on their own terms to produce data. The case study supported communities in collecting and commissioning the collection of environmental data from scientific-grade sensors by establishing a toolkit called SenseMyStreet (SMS). The toolkit consists of a GIS system for issue mapping, a scheduling system of environmental monitor loans and deployments, a mobile application for additional citizen data collection, and a collaborative proposals platform to commission environmental sensor deployments and data. It was designed through a user-centred approach, enabling people to generate their own data sources to raise awareness about local issues and use it in civic advocacy to help influence local policy-making processes. Enabling citizens to take control of the technology in the smart city has increased the usage of the data collected from the city and has helped with the wider engagement of the public in civic issues. This case study illustrates the importance of active citizen participation in the smart city, in addition to identifying processes and resources needed for producing and effectively using that data. Furthermore, the case study evaluation also helped to uncover new understandings of citizens taking action in these contexts, which is presented though a Citizen Advocacy Framework (CAF). SMS continues to be deployed and used by communities – while also being expanded to open up different routes for civic engagement, education and advocacy – with processes put in place to achieve a sustained model of engagement. Having routes for citizens to actively contribute to decision-making processes is changing the power relationships between citizens and decision-makers, thus moving towards a truly inclusive smart city.

Case Study II (in Chapter 4) looked at the opportunities and challenges of accessing and using Open Governmental Data (OGD) in community settings and for informing local decision-making processes. Although the number of published datasets are increasing by the day, the usage of that data remains low or is contained within an expert group of users. Looking at the current state of available tools for accessing and interrogating open data, issues of paywalls and a lack of support for non-expert users were identified. Engaging with the local community and taking a user-centred design approach, an online system called Data:In Place was built to utilise a visual map-based querying system so as to help define requests for official data sources connecting to their expert systems and retrieving data to be visualised on the map. The case study aimed to address the information engineering problem of removing technical barriers to data access and interpretation and to make interrogation of open data possible for non-experts; however, it approached this issue from a community problem-solving perspective.

Case Study III (in Chapter 6) used the new model of Community-Data Interaction (CDI) derived from the findings of the previous two case studies and applied it in action. It used a dual approach

that leveraged the capabilities of the technologies built in Case Study II and developed new training and guiding methods to help transfer some of the skills of data professionals to citizens. With this case study, the aim was to help citizens increase social capital and effectively use data for the local benefit.

4. Summary of Contributions

The research undertaken in this thesis resulted in five contributions to the fields of HCI, IS engineering and Community Informatics (CI). In addition to published work in peer-reviewed scientific journals and conferences (see Publications), the research also produced open-source tools (the SenseMyStreet and Data:In Place platforms) and processes that continue to be used and make a lasting impact on communities. The list of scientific contributions from the thesis are as follows:

- Establishing an understanding of what makes data usable (i.e. effective use) for non-expert citizens and communities, contributing new understanding of the relationships between data, communities and civic advocacy and action. This includes a deeper understanding of the role that expert professionals play in all this.

- *SenseMyStreet (SMS)*[1] – establishing a first sensor commissioning toolkit for communities that enables people to use scientific-grade environmental monitors to produce data about their neighbourhoods: (i) the design and development of the SMS toolkit (a GIS system for issue identification, a mobile application, an automatic scheduling system, user tutorials, and a collaboration platform for sensor commissioning) (Section 2.4 in Chapter 2); (ii) evaluating the toolkit and its tools through multiple deployments 'in the wild' to determine the required commissioning services for data production and the utility of that data to support local decision-making; (iii) presenting a novel CAF to help understand and plan activities for civic advocacy and action using data; and (iv) establishing a sustainable model for sensor commissioning.

- *Data:In Place*[2] – an open-source web platform that supports citizens in accessing, interpreting and using open data for the purposes of civic advocacy and action. The main contributions of Data:In Place are as follows: (i) the iterative design process of Data:In Place from the community problem-solving perspective; (ii) an online GIS tool with a novel map-based querying system for accessing official statistics about a place from expert systems, helping to interrogate the data and enabling the mapping of additional data sources and creating data visualisations (Section 4.6 in Chapter 4); (iii) reflections on the Data:In Place system, its utility and current limitations as a tool for civic advocacy; and (iv) design recommendations to inform future research of data systems for community knowledge-making and civic advocacy.

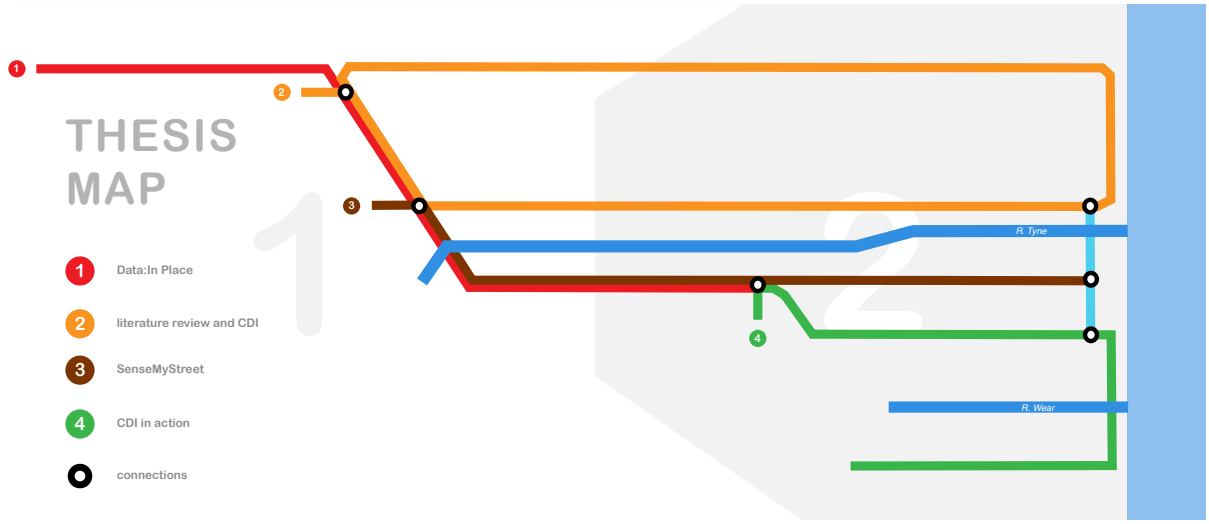
- Extending the field of CI by presenting of a model of CDI to understand the process of community knowledge-making for civic action through the effective use of data by citizens, in addition to providing a vocabulary and ontology that help identify the gaps, needed resources and capacities to support data science practices within civil society.

[1] https://sensemystreet.uk
[2] https://data-in.place

- Empirical findings regarding the practical use of CDI translated into technical implementations and research methods to support community knowledge creation using data for the purposes of civic advocacy and action.

5. Thesis Structure and Outline

The remainder of this thesis consists of eight chapters that report on three case studies with the community groups outlined in Section 2.2. Figure 2 presents a map and timeline of the thesis, which should be read from left to right. Because of the nature of conducting research 'in the wild' and with communities, a lot of the time the different research activities often started simultaneously, ran concurrently, overlapped, fed into each other or stranded off. In this sense, the figure illustrates how the different case studies were conducted within a timeframe of three years. In this sense, we can see that engagements for (1) Data:In Place (Case Study II in Chapter 4) started before (3) SenseMyStreet (Case Study I in Chapters 2 and 3); however, they are reported in this thesis in reverse order for the purpose of a clearer narrative. Furthermore, (2) the literature review was constantly being expanded upon and analysed in light of new findings from the case studies. Reflections on the iterative design processes, development and deployment of the two actively engaging data systems (SenseMyStreet in Case Study I and Data:In Place in Case Study II) provided insight to conclude with the development of a new model of CDI for the effective use of data by citizens for civic participation, advocacy and action. The CDI model was then trialled (4) in action in Case Study III (Chapter 6), thus presenting empirical evidence for its value with regard to the purposeful use of data and sustained engagement with data around civic concerns. Additionally, the first half of the research was more focused on answering the first research question, whereas the second half dealt more with the second research question.



Source: Author (adapted from RaviC under CC BY-SA 4.0)

Figure 2 Illustration of the thesis structure, from left to right

Chapter 1 (*Literature Review*) explores the relevant literature for this study. It defines a smart city and the role of citizens in it as active and passive data users. The purpose of this chapter is to determine the additional level of participation for citizens in the smart city that enables them to become active users of data for participating in decision-making processes. Additionally, the chapter enables an understanding of what makes data usable for citizens as active participants.

Chapter 2 (*Case Study I: Democratising Data Production*) presents the iterative design, development and deployment of SMS, a sensor commissioning toolkit for communities that enables people to use scientific-grade environmental monitors to collect data about their neighbourhoods. In this regard, the chapter describes the evolution of the toolkit and the analysis of lessons learned from setting it up.

Chapter 3 (*Case Study I: Data Use in Advocacy*) presents an evaluation of the use of the sensor commissioning toolkit by a group called SPACEforHeaton[3], where citizens were using the data collected to evidence issues, raise awareness in their community and influence policies linked to their local areas. Based on the analysis of the findings, a novel theoretical model – the CAF – is presented to enable discussion and plan for civic advocacy and action by citizens.

Chapter 4 (*Case Study II: Democratising Data Access, Interpretation and Use*) focuses on the access and interpretation of already existing sources of data for community problem-solving. It presents Case Study II, the design and iterative development process of Data:In Place, an online GIS system for accessing and interpreting open data from governmental sources to support local decision-making and civic action.

Chapter 5 (*CDI: A Model of Effective Use of Data*) reflects on the first two case studies and goes deeper into the barriers and issues around data use and appropriation for local benefit. It involves a description of the researcher's process and the difficulties faced when aiming to make data accessible, usable and useful for communities. From this analysis, a new model for Community-Data Interaction is presented that supports future research in CI relating to the effective uses of data by citizens for civic participation, advocacy and action.

Chapter 6 (*Case Study III: CDI in Action*) puts the new CDI model into action through a dual approach involving the use of the platform developed in Case Study II (i.e. Data:In Place), additional technologies for deliberation (i.e. Ambit[4]) and social facilitation methods that aim to increase social capital and help transfer some of the skills of data professionals to local communities so as to enable them to effectively use data in action.

Chapter 7 (*Discussion*) discusses and assesses the success of the research through the primary aims and objectives and outlines its contributions to the wider field of research. It reflects on the challenges and limitations of the built tools and processes, proposing routes for future research and suggestions for the sustainability of these developed solutions.

Finally, Chapter 8 (*Conclusions*) presents a summary of the contributions to the multiple fields of HCI, IS engineering and CI and concludes by positing responses to the thesis' research questions.

[3] https://www.spaceforheaton.com
[4] https://myambit.co.uk

Chapter 1. From Smart City to Smart Citizens: The Role of Data in Supporting Civic Participation

Thinking is hard. Words must make it as easy as possible.

— Juhan Liiv

This chapter outlines the concept of a smart city, which, through the use of information and communication technology (ICT) and data, is changing the way cities operate. Under this model of city governance, the people who inhabit these cities are often perceived as another passive source of data for measurement, control, optimisation and management at scale. The literature review in this chapter explores different modes of civic participation and examines what other roles citizens could have in the smart city. How do we move from smart cities to smart citizens? What is the role of active citizens and the role of data in supporting activities and engagement with decision-making that determines the future of their city? As the creation and use of data in civic life and within communities span a wide range of disciplines, the literature from technical fields such as pervasive computing and information engineering is overlapped with social science, HCI and CSCW. Drawing on a range of critiques, this chapter identifies criteria for a more active idea of the citizen as a user of data (i.e. an active or a smart citizen), which will be the main focus for the research presented in this thesis. Thought the literature review, the research gaps are identified, providing the basis for the thesis' case studies.

1.1. Participation in the Smart City

1.1.1. Defining the Smart City

The vision of a smart city, where information technology is woven into the city infrastructure to help optimise resources and improve the lives of citizens (Chourabi et al., 2012; Ratti and Claudel, 2016), has been the focus of heated debate over the past five years[1]. Perhaps considered fictional at the time, such ideas were echoed in concepts like the *'third wave'* of computing or '*ubiquitous computing'*, as envisioned by Weiser (1991, 1993) in the early 1990s. For many years now, commercial companies – IBM, Cisco and Siemens, to name but a few – have been 'hooking up' the city with sensors and other information technologies in order to gather all the data about the city in the hopes of solving some of the old and new 'wicked' problems of society. For example, intelligent street lighting is in place in many cities to improve efficiency and reduce energy costs; CCTV and traffic counters are used to manage the flow of traffic through traffic signals; and pedestrian counters are utilised to understand the flow of people through the city. In this regard, there seems to be a push for the use of data in all aspects of life in order to optimise and organise resources (Batty, 2013; Calvillo et al., 2016), reduce our carbon footprint (Li, 2015; Pierangeli et al., 2020; Zubelzu and Álvarez, 2015) and improve sustainability (Vilajosana et al., 2013), in addition to also facilitating novel modes of participation (Tenney and Sieber, 2016), democracy and social interaction (Baack, 2015; Tan et al., 2013). Although there are plenty of 'smart' applications available, discussion around what makes a city smart is still ongoing (Dameri and Cocchia, 2013; Foth, 2009; Klein and Kaefer, 2008; Nam and Pardo, 2011; Su et al., 2011). However, with the increase of urbanisation and the expansion of cities, this trend will likely continue to increase, and more businesses are developing solutions to supposedly '*compute away local problems*' (Townsend, 2013).

Smart cities live off data – traditionally acquired through a rigorous collection process by government (i.e. conducting censuses) Ð that is increasingly captured through the use of sensor networks located on lampposts, roadsides, trees and rooftops (Su et al., 2011; Zanella et al., 2014) from humans inhabiting the city (e.g. mobile phone data, passive mobile positioning and 'citizens as sensors') (Ahas et al., 2010; Berntzen et al., 2018; Rein Ahas, 2010), as well as from the web and social media platforms (Batty, 2010; Puiu et al., 2016). The vast amount of data accumulated through the smart city, coupled with ideas of 'data-driven everything', has resulted in claims that cities are becoming 'sentient' in terms of developing an awareness of everything that is going on in the city (Thrift, 2014). Furthermore, interest in data-driven city governance is growing as more technologies are being developed and deployed in the city. In turn, this requires a new skill set relating to *urban informatics* (Thakuriah et al., 2017) from not only city managers but also from private and non-profit organisations (NPOs) that work to improve the life of people in the city (Foth, 2009; Kontokosta, 2017).

A popular way of showing and using all the accumulated data is through real-time *'data dashboards'* (Kitchin, 2014; Kitchin et al., 2015). Almost all aspiring smart cities have at least

[1] https://trends.google.com/trends/explore?date=all&q=smart%20city

one dashboard at hand to create and visualise statistical indicators, benchmark the city against others and drive city governance (Kitchin et al., 2015). According to Kitchin et al. (2015), the utility of these dashboards can be separated into two broad categories, i.e. *managerial* and *contextual*, with the first more linked to the assessment and performance measurement of city governance and the second prioritising the use of indicators for informing policy. Developing statistical indicators and visualising them through dashboards is a necessary and valid approach in city management, especially when trying to condense all that messy data into useful chunks that can be represented and used by decision-makers. However, Kitchin et al. (2015, p. 24) points out that these initiatives are often presented as a mirrored model, or an exact *'digital twin'* (Bolton et al., 2018), of the city through numbers – *'visualised facts that reflect the truth about the world'*. Instead, what these dashboards actually are is a representation of abstractions that make up a view of the city. Unfortunately, in many cases, these dashboards are representations of the view of the city from the perspective of management and governance, which means that problems and data collection are framed by institutional actors to fit the governance processes of the city, often without engagement from the relevant publics (Asad et al., 2017). Similar concerns are raised by Greenfield (2013) and Townsend (2013), who argue that the power of making educated decisions based on that data is currently in the hands of the big tech giants, whose software works in controlled centralised locations. In addition to concerns relating to transparency and ethics surrounding surveillance and control (Tufekci, 2014b; Van Dijck, 2014), this lack of inclusivity in smart cities can lead to a narrow representation of views and social desires in governance (McMillan et al., 2016; Tufekci, 2014a).

1.1.2. Citizens in the Smart City

In the smart city literature, the people inhabiting the city are referred to by a variety of generic terms, such as 'subjects', 'consumers', 'users', 'stakeholders' and 'citizens'. In this thesis, the term *'citizens'* is used for those who inhabit, work and play in cities.

Berntzen et al. (2018) argue that in addition to political participation – direct and indirect participation in political decision-making – the role of the citizen in the smart city is also to participate in helping solve problems and improve services in the city (also (Townsend, 2013)). However, this requires the city to have the necessary infrastructure and processes to support communication and multiple types of participation (Berntzen and Karamagioli, 2010). New technological infrastructure and sources of data definitely broaden channels of participation. In modern democratic societies, citizens should be actively participating in all matters of civic life (Chourabi et al., 2012), including political discourse (Hoffman, 2012), through civic engagement activities run by the state (Berntzen et al., 2018) as well as citizen initiatives. On the other hand, from the algorithmic smart city perspective, citizens can also be viewed as additional sources of data that could be tapped into to improve the efficiency and functionality of the smart city.

There is enormous potential for technology to enable new ways of participation, engagement and service delivery; however, this comes at a time when the UK, among other countries in the EU, is subjected to acute economic austerity, which has had a significant effect on local

governance, resulting in public service delivery having severely suffered and in some cases even being outsourced or privatised. Many of the solutions provided by institutional actors are not relevant to the real issues in the community (Boehner and DiSalvo, 2016), and people living in communities often have a much keener sense of what the real problems are (DiSalvo et al., 2008).

With the age of the internet and information technologies becoming more accessible outside of institutional boundaries, digital civic approaches (Olivier and Wright, 2015) have been emerging from HCI-related disciplines, which promote active participation in civic life and engagement in service creation and delivery, with the hope of increasing civic professionalism (Boyte and Fretz, 2010) in society and increasing dialogue between citizens and the government Boyte (2004). Examples of these kinds of study areas are community commissioning (Balaam et al., 2015; Garbett et al., 2016; Taylor et al., 2009; Taylor and Cheverst, 2010), democratic discourse (Crivellaro et al., 2015; Johnson et al., 2017, 2016; Valkanova et al., 2014), situated voting (Behrens et al., 2014; Golsteijn et al., 2016; Hespanhol et al., 2015; Taylor et al., 2012; Vlachokyriakos et al., 2014) and community infographics (Claes and Vande Moere, 2013; Koeman et al., 2014; Lindley et al., 2017). Common to all these studies is that they are moving out from the research facilities and laboratories to conduct investigations in the real world. Rogers (2011) calls this approach research *'in the wild'* – carrying out in situ studies and, rather than conducting observations of existing practices, focusing on local issues and designing disruptive technologies to address the concerns and needs of a specific population. Depending on the configuration of participation (Vines et al., 2013), this could also mean including citizens as co-designers in the process of design.

Although these processes are often messy, unpredictable and uncontrollable, they provide a better insight and understanding of how technology should be designed and how it is used in the real world (Suchman, 1985, 2007). However, these approaches often stay within HCI- and social science-related disciplines, meaning that the actual technology and systems implemented and supported in the smart city do not extend to community problem-solving activites (Stillman and Linger, 2009). As machines and algorithms are getting better at making decisions and acting on behalf of humans, gaining more 'non-human agency', citizens cannot be disconnected from the decision-making processes that affect their lives (Graham, 2014). Instead, these approaches should overlap with engineering, geography, planning and spacial modelling to tackle inequalities in the smart city[2] and respond to the UN's Sustainable Development Goals[3].

Indeed, many have recognised (Rudolf Giffinge et al., 2007; Saunders and Baeck, 2015) that using technology for participation is the key to a successful model of a smart city. However, the type of participation enabled by the smart city is determined by how citizens can engage with technological solutions and what roles they take in the use, design, creation and ownership of these solutions.

[2] https://www.turing.ac.uk/research/research-projects/spatial-inequality-and-smart-city
[3] https://sustainabledevelopment.un.org/?menu=1300

1.2. Data-Driven Participation

Citizens are not only moving around in this sensed environment (i.e. smart city) but are also contributing some of the data to it by using smart devices in their pockets, i.e. smartphones, and other portable personal computing devices. Townsend (2013) calls them *'smart-city construction kits'* or *'digital Swiss Army knifes'*, which are increasingly becoming sources of data for studying human behaviour and platforms for new bottom-up civic movements (Eltantawy and Wiest, 2011; Obar et al., 2012).

However, not all participation is created equal. In the literature, two distinct modes of data-driven participation can be identified: *passive* and *active*. Passive participation refers to the use of some service or automated system (e.g. a data logger or a service on a personal computing device), or as the result of a 'data exhaust', that sends data off to be analysed by algorithms implemented on automated systems. This means that citizens are contributing data either voluntarily by knowingly using a particular application or program or anonymously in the background by using some type of service on their computers or mobile devices. Passive participation is the classical model of smart city participation, where citizens are contributing another dataset for the smart city to use.

On the other hand, active participation is referred to in the literature as the type of participation where citizens take a more active role in data production, with the aim of helping to solve the city's problems. This can manifest itself, for example, in citizens actively contributing data by reporting issues and problems through their personal computing devices or using citizen science devices to collect data about issues concerning them. This model of data-driven participation turns people into data producers and in some cases consumers, which is a more active way of participating in the smart city. Furthermore, some citizens with skills could even become prosumers of data.

Both modes – *passive* and *active* – refer to the ways citizens supply data to the smart city and its algorithms, which in turn inform decision-making. However, there is another type of participation – *the active data user* – that has generally been neglected by researchers, designers and city officials. This type of participation is the focus of the present thesis, i.e. participation where citizens can *access*, *create*, *understand*, *use* and *make use* of data to actively participate in decision-making processes in the smart city.

1.2.1. The Citizen as Passive Data User

Historically, people were principally connecting by trade, post, telephone or face-to-face interactions. In the digital age, it is becoming more commonplace to use computerised systems and data to mediate communication of ideas and opinions. The internet enables us to establish enormously complex networks, connecting us with different people and helping to engage in numerous activities, both individual and collaborative. Although not always explicit, all participation in cyberspace is driven through data sharing and use, which can manifest itself, for example, in participation on social media platforms like Twitter and Facebook (FB) (Crivellaro et al., 2014),

using personal informatics tools[4] for self-reflection or improvement (Epstein, 2015; Epstein et al., 2015; Li et al., 2010), or engaging with collaborative platforms to work towards a common goal (Salovaara et al., 2006). Popular examples of the latter are Wikipedia, a crowd-powered free encyclopedia, and OpenStreetMap, an open-source geographical database compiled by a community of volunteer mappers.

The availability of portable smart devices in almost every citizen's pocket is producing excellent sources of data for the algorithmic smart city to understand how humans use the city. For example, there is a growing body of research into the use of mobile phone call logs for developing near real-time models of spatial interactions (Ahas et al., 2010; Vanhoof et al., 2017) and deriving early statistical indicators to optimise planning and transport (Sonntag et al., 2014). Furthermore, usage of mobile phone data has also shown to be beneficial for developing epidemiological models (Finger et al., 2016; Panigutti et al., 2017). In addition to using passive mobile positioning, the 'smart city' is tapping into the data generated by social media in order to gather opinions and model large-scale human behaviour. There is a growing body of literature around the usage of social media platforms to conduct large-scale content analysis for semantics (Bontcheva and Rout, 2014; Cranshaw et al., 2012; Golder and Macy, 2011; Procter et al., 2013; Sizov, 2010). Other applications include the use of social media for crisis and emergency watch (Kamel Boulos et al., 2011; Liu, 2014), developing predictive models for public health interventions (Chourabi et al., 2012) and indicating political sentiment or even predicting the outcomes of political elections (Tumasjan et al., 2010). Social media can be a powerful tool for mediating participation, but it can also be used to influence the outcomes of democratic elections (Bond et al., 2012). Although data from this type of passive participation of citizens can be useful for the smart city, there is definitely a discussion to be had around the ethical and privacy concerns surrounding it (Lanier, 2014; Van Dijck, 2014).

In addition to using the 'data exhaust' of people's phones or activities on social media, this type of passive data participation can also be voluntarily. Certainly, civic engagement is one domain where such technologies and methods are being introduced, with the aim of increasing participation in democracy through data. In particular, these new approaches have caught the interest of city planners, who see them as an excellent opportunity for large-scale participation through the use of data and pervasive technology, i.e. trying to harness the power of crowds through data science (Tenney and Sieber, 2016). In big geography and urban planning, this is often referred to as volunteered geographic information (VGI), which has been argued to increase participation and the representativness of data (Goodchild, 2007). Examples of these types of systems include projects like MetaPos[5], Habits[6] and Sunset[7], each of which has a mobile application that collects anonymous location data to be used in spatial modelling to help with the organisation of infrastructure and/or traffic in the city. Although these applications have elements of feedback in terms of people seeing their analysed mobility patterns or being able

[4] http://www.personalinformatics.org
[5] https://www.gsa.europa.eu/r-d/gnss-project-portfolio
[6] https://habitsdata.org
[7] http://sunset-project.eu

to share travel tips with others on social media, they are still contained within the frame of the research or agendas of city authorities. Furthermore, some of these agendas can also be linked to behavioural change (not necessarily bad), which is administered through incentives[8]. This *'citizens as sensors'* (Berntzen et al., 2018) form of participation is often preferred by corporation, busy citizens and city managers because of its low impact on citizens' everyday lives and the reduced workload on city officials to actively engage their citizenry (Tenney and Sieber, 2016).

There is a lot of potential for crowdsourced data collection approaches to broaden participation and create alternative sources of data for public use in multiple domains. Salim and Haque (2015) developed a taxonomy of interaction and participation to assess different types of urban computing solutions. Their paper surveyed a variety of applications in mobile crowdsensing, urban informatics, urban probes and interventions, and Internet of Things (IoT), categorising them in terms of *'the technologies used, the level of participation they stimulate, the participation scale they support, the manipulation and effects mode they enable, [and] the interaction mode and scale they enable'* (Salim and Haque, 2015, p. 44). Although urban computing systems may vary in shapes and sizes, they all aim for maximum participation from the public. Through reflecting on projects, Salim and Haque (2015) compiled a list of strategies that put emphasis on identifying needs and tensions through active dialogue with stakeholders and providing open and accessible tools that help citizens construct their own rationales for participating. It seems that the key for successful engagement is not only to provide the best technology for participation (i.e. only working on engineering problem solving) but also to prioritise the social interactions surrounding it (i.e. work on community problem-solving activities). There are new opportunities emerging around designing tools that successfully facilitate this kind of data collection, which tries to harness crowds and capitalise on the capabilities of big data. However, there is an ongoing discourse around the use of crowdsourced geospatial (not exclusively) data to mediate civic interaction, both in term of its validity for participation in democracy (Le Dantec et al., 2015) and as a tool for decision-making (Brabham, 2009; Misra et al., 2014; Tenney and Sieber, 2016). Caution needs to be taken not to fall into the same pitfalls by taking data at face value. As Le Dantec et al. (2015) and Kitchin et al. (2015) observe, it is often appealing for city managers to base their decisions solely on what the numbers are showing and the maps are identifying rather than on engaging directly with citizens to also consider what the people on the ground are experiencing, including not considering that technology sometimes excludes groups of people who are struggling with low-tech literacy or just do not have the resources to access them. People need to be reminded that big data can only show the 'what is' but cannot reveal the 'how', 'why' and, most importantly, 'what is not' (Mayer-Schonderger and Cukier, 2013).

1.2.2. The Citizen as Data Maker and Producer

Active data-driven participation also leverages the capabilities of modern-day information technologies and the internet but also tries to actively involve and engage in the problem solving, design and reorganisation of urban environments. These active participation approaches, such

[8] https://www.betterpoints.ltd

as Public Participation Geographic Information Systems (PPGIS), are perceived to be more grassroots, enabling citizens to actively contribute to decision-making processes (Jankowski, 2009; Sieber, 2006). Examples of such systems include FixMyStreet[9], popular in the UK, and its Norwegian clone FiksGataMi[10]; SeeClickFix[11], PublicStuff[12] and Street Bump[13] in the US. Although PPGIS, at its core, is a *'bottom-up'* approach, its technological realisation is often more '*top-down'*, serving governmental interests (Tenney and Sieber, 2016). However, the scope for these technologies more often originates from institutional actors who frame the issues from their perspectives. Although helpful for citizens, this can result in a technical solution that generates data specific to the formulation of governmental concerns. In addition, the 'datasets' are usually collected by non-expert citizens without any formal training, thus resulting in questionable accuracy, which makes city managers sceptical of the value of the data and opinions coming from non-expert citizens (Tenney and Sieber, 2016). Furthermore, there are issues around the interpretation of citizen concerns from data to actual tangible solutions in the city and in sufficient feedback loops to measure the successes of those solutions.

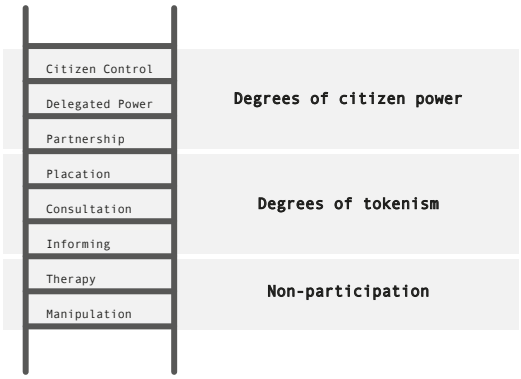
Citizen Sensing and Citizen Science

In addition to sensors in the city, i.e. passive and active data-driven participation that supports the algorithmic smart city agenda, there is also citizen sensing, an approach that usually involves (albeit not exclusively) citizens using a sensing technology (or mobile phones) to voluntarily gather data. Citizen sensing is not exactly the same as crowdsourcing, where people are seen as workers or data collectors who complete tasks or collect data for the work 'requester', often getting paid for their participation. Furthermore, citizen sensing is not necessarily exclusively linked to citizen science initiatives, in which non-scientist members of the public are recruited to take part in science projects, volunteering their time for data collection and/or analysis Cohn (2008). In the same way, simply taking part in a study – for example, participating in a sleep study or filling in a survey Ð should not be confused with citizen science. In this sense, citizen sensing can be used to refer to any activity where citizens are engaging in voluntarily data collection using sensing technologies, and it can be linked, although not exclusively, to crowdsourcing, participating in a study or a citizen science project.

According to the definition of the network of the National Contact Points for Science with and for Society[14], citizen science projects are *'projects where citizens contribute to actual science research'* (Eitzel et al., 2017), examples of which are BioBlitz, a biological surveying project, and British Bird Watching, a bird surveying project. It seems that citizen science has multiple definitions, and it is often down to the 'scientists' to decide how to frame it; however, it is not often clear who 'the scientist' is (Haklay, 2013) in these projects. Haklay (2013, p.11) proposes a framework to classify citizen science projects that focuses on the level of

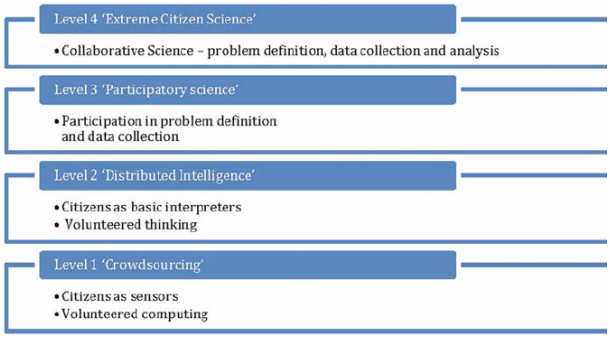
[9] https://www.fixmystreet.com
[10] http://www.fiksgatami.no
[11] https://seeclickfix.com
[12] http://www.publicstuff.com
[13] http://www.streetbump.org
[14] https://www.sisnetwork.eu/sis.net-project

engagement and involvement of participants in science activities (Figure 1.1b). This is somewhat similar to Arnstein (1969)'s *'ladder of participation'* (Figure 1.1a); however, instead of solely focusing on the power relationships between citizens and decision-makers on a fixed ladder, Haklay (2013) also points out the processes of obtaining new understandings that foster curiosity, promote knowledge creation and are likely to move people up the ladder of participation. In addition to finding an appropriate and sustainable engagement level for the project, there is also the question of 'who knows best?' Although there have been successful citizen science projects like Neighbourhood Nest Watch (Evans et al., 2005) or volunteer computing projects like climateprediction.net and Rosetta@home, there is ongoing debate about whether the 'science' should be left to the 'scientists' or whether we should fully accept citizen contributions to science activities (Cohn, 2008; Gabrys and Pritchard, 2018).



Source: Author

(a) Ladder of participation (Arnstein, 1969)



Source: Haklay (2013, p. 11)

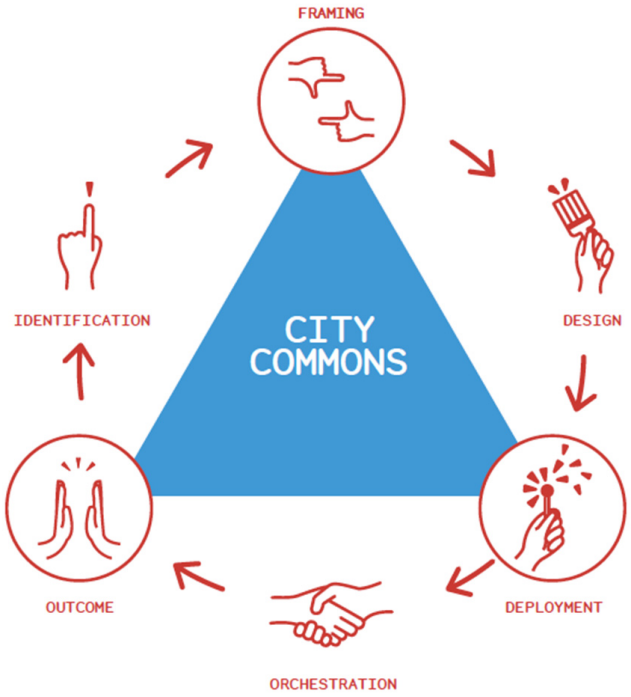
(b) Levels of participation in citizen science (Haklay, 2013)

Figure 1.1 Ladder of participation and levels of participation in citizen science

However, there can be other purposes for conducting citizen data collection than advancing science or providing data to the smart city, e.g. for self-knowledge and personalised health interventions (Huck et al., 2014; Larkin and Hystad, 2017; Swan, 2012), to raise public awareness around an issue (Loreto et al., 2017, ch. 16) (Hasenfratz et al., 2012; Kuznetsov and Paulos, 2010), to evidence an issue in the community (DiSalvo et al., 2009) or to push for change (Balestrini et al., 2017). The latter approaches involve community-led investigations, whereby citizens are actively involved in the definition of the issue, as well as the collection or acquiring of data and its analysis. In relation to the typologies of the ladder of Arnstein (1969) and the four levels of Haklay (2013), these approaches come at the very top in terms of the level of active engagement. Balestrini et al. (2017) examines this type of participation through a *City-Commons Framework for Citizen Sensing* (Figure 1.2), which utilises a bottom-up approach to orchestrate citizen sensing and community engagement. The types of citizen sensing projects that explore issues related to the environment and place often involve some sort of sensing technologies (Thompson, 2016) and PPGIS or mapping coupled with HCI for geospatial technologies, examples of which are Aircasting[15] and SmartCitizen[16] (Chapter 2), both of which consist of lightweight sensing

[15] http://aircasting.org
[16] https://smartcitizen.me

equipment and GIS applications for exploring the data. Citizen sensing provides an opportunity to rethink the role of data in civic participation and focus on digital technology and mechanisms that could support such grassroots approaches. More projects like this are now being developed around the world, with community volunteers using sensing equipment and collecting data with the aim of promoting dialogue on environmental issues and getting closer to the next level of participation through data.



Source: Balestrini et al. (2017, p. 2284)

Figure 1.2 City-Commons Framework (Balestrini et al., 2017)

1.2.3. The Citizen as an Active User of Data

Although citizens are actively involved in data production with active data-driven participation, and they are also more involved in problem definition with citizen sensing approaches, they are not necessarily using the data for activism or changing their communities. Another level (i.e. the third level) of data-driven participation is where citizens not only supply data, either passively or actively, but also have access to such data in a way that they can make use of it. This type of participation is referred to in this thesis as the citizen as *active data user* – equipped with the tools for accessing and understanding data and sensing technology to create and use data and knowledge so as promote dialogue on issues and push for positive change in the community.

This shifts the focus from the smart city to smart citizens who are making the city smart. Active citizens not only produce data but also get access to the data and can even commission the kinds of data they need, opening up another level of data-driven participation and empowerment through data. The ‘new‘ smart city, which is creating data with the people for the people and enabling them to commission and use data, then has to think about what structural, digital and

legal infrastructures are required to support these new forms of data use, which go beyond simply data consumption or enabling the smart city to consume it. This could be done by opening up processes of the smart city to citizens and moving away from the technocratic governance and corporatisation of city management (Kitchin, 2014). Kitchin et al. (2015) encourages smart city initiatives to openly recognise their flaws, uncertainty, messiness and levels of error and accept the fact that they do not reflect the world as it actually is but actively frame and produce the world through measuring and observations it. In this sense, there is no one true image of the city, and everything depends on the perspective of the observer. While there is great potential in using data in such contexts, and an active effort is being undertaken to integrate sensors into our spaces, meaning that a lot of technology is already in place, it has to be configured and shaped in a way that it serves the needs of citizens and not the technology itself.

Civic Advocacy

This type of stance is also activist in nature, promoting activities of civic advocacy and action. Engaging in practices linked to advocacy is not a new concept in democratic societies. For many people, lobbying and grassroots lobbying are considered the main forms of advocacy; however, advocacy is not limited only to lobbying. It utilises protesting, petitions, citizen campaigning and a variety of other information resources to influence decisions within the socioeconomic and political systems that shape the lives of people (Christoffel, 2000) as legitimate means of democratic participation that should be part of a healthy society. Advocacy efforts can come from individuals, but they are more likely to be associated with some sort of organisation or self-organised group (Leroux, 2007; Obar et al., 2012; Prakash and Gugerty, 2010; Suárez, 2009). Advocacy can also be categorised into two strategies: insider and outsider strategies (Asad and Le Dantec, 2017), with the former using more traditional approaches to directly affect policy systems and the latter focusing on informal processes by contributing resources or skills to organisations in order to help address issues (Erete et al., 2016; Hackler and Saxton, 2007; Marshall et al., 2016). Additionally, the advocates are often not the ones with the personal attachments to the issues and are acting as proxies to voice community concerns (Johnson et al., 2017; Leroux, 2007; Voida et al., 2012).

ICT affordances and capabilities are enabling new forms of digital advocacy (Brady et al., 2015), where people are taking advantage of digital tools and platforms to affect change in the community (Karpf, 2010), e.g. using platforms such as Change.org[17] and TheyWorkForYou[18] to digitalise traditional forms of advocacy (e.g. signing petitions and sending letters to representatives). Research in HCI has started looking into ways that technologies and data can facilitate these new roles of active citizen engagement. For example, Garbett et al. (2016) examined a model for the community commissioning of technologies to help communities gather data about local concerns; Balestrini et al. (2014) explored how low-cost sensing equipment can be used to empower citizens tackling specific sustainability issues; and (Kuznetsov and Paulos, 2010)

[17] https://www.change.org
[18] https://www.theyworkforyou.com

looked at using low-cost sensors for measuring the outdoor air quality of diverse communities and explored the uses of data for public awareness and activism, in addition to projects where citizens used their mobile phones to collect data about transportation issues to advocate for changes in cycling infrastructure (Dantec, 2014; Dantec et al., 2016; Maskell et al., 2018) or other types of sensor data to bring issues to the public attention and advocate for more accessible city infrastructure (Rodger et al., 2016). All of these examples are part of the model of citizens as active data users. At the other end of the spectrum, citizens themselves have taken control and repurposed technologies to aid in their advocacy efforts. For example, we have already witnessed disruptive social movements through the use of social media platforms (Eltantawy and Wiest, 2011; Howard et al., 2011) and smaller movements aimed at hyper-local concerns Crivellaro et al. (2014) or the coordination of humanitarian action (Starbird and Palen, 2011; Starbird et al., 2010). A significant benefit of online tools is that they do not put geographical restrictions on participation, allowing for participation from other places and also enabling instant access to shared information.

Digital advocacy efforts are similar to the participation of citizens as data producers; however, data-driven advocacy (Asad and Le Dantec, 2017; Gray et al., 2016) is taking a more active stance, wherein citizens become the creators of data. While this also involves people using digital technologies to collect data, they then use that data to inform or evidence an issue or advocate for change. In addition, as a more active form of participation, it has greater emphasis on the relational interactions between civic stakeholders (Asad and Le Dantec, 2017; Olivier and Wright, 2015; Vlachokyriakos et al., 2016). Although there might be a multiplicity of stakeholders from different organisation involved, issues take precedence and different communities, or *publics* (Section 1.5.1), form around them. Projects like The Bristol Approach[19] (Figure 1.2) provide examples of issue-led investigations where people often co-design the technologies used in advocating. Furthermore, it has been pointed out that the technology and how it is used in advocating for issues play an active role in forming these publics (Dantec and DiSalvo, 2013; Jenkins et al., 2016). When designing or repurposing digital tools for active citizen (i.e. active data user) engagement, there is an obligation from designers and developers to think about how they support different groups and promote civic participation and advocacy (i.e. infrastructuring) among all citizens (Asad and Le Dantec, 2017). It seems that the way forward is to open up and democratise data and technology that is already used in decision-making to help integrate a multiplicity of views and data into the smart city.

In his book *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*, Townsend (2013) states that the answer to the question of what is the smartest city is *'[...] the one you live in'*. People make the city smart and not the technology, although technology can certainly lend a helpful hand in defining and informing people of some of the issues in the city or helping to communicate ideas and knowledge. However, citizens also need to take a more active role in matters of local concern and actively engage in issue exploration and data production and

[19] http://www.bristolapproach.org

analysis and use that to develop the smart city. In addition to being passive sources of data, the producers and users of data citizens are going to become commissioners of data. However, with these new modes of participating, it might change the nature of what we mean by data and how we use it in knowledge-making processes in such contexts. Hence, there is a need for a better understanding of the implications for data or for what we understand by data. Obviously, as a necessity, data needs to be open for anyone to use it, but the ways it needs to be open and what it needs to be doing for it to be used in a citizen-driven smarter city that support these new modes of participation need further exploration.

1.3. Open Movements

Just as all ideas and opinions were historically shared through face-to-face encounters, all data, information and knowledge was also exchanged orally and spread freely. However, with the rise of the printing press in the 1400s, knowledge started to be seen as individual rather than collective property. Soon, multiple copyright laws were put in place and enforced by governments across the world to protect the property rights of publishers. The coming of the internet, which made sharing and copying information even easier and broader, resulted in copyright laws becoming stricter and more unified around the globe. In response to the prevalence of copyright laws, a new political and social movement – the *Open Movement* – was brought into being, which seeks to promote the sharing of content and resources for the good of society. In this regard, the idea of open movement operates in a spirit of transparency, collaboration, reuse and free access. It has been realised in some regards through specific groups, institutions and individuals around the world embracing the benefits of releasing digitised text, data and multi-media content online free of charge, often collaboratively, and free of most copyright and licensing restrictions. Open Movement is seen as an umbrella term that encompasses factors such as open access, open data, open government, open development, open science, open education, open source, open licensing and open content. However, open can mean many things, and what the research in this thesis is examining is how these movements are contributing to making data and knowledge open so as to support active citizen engagement in new modes of participation in the smart city.

1.3.1. Open Knowledge

Open movements aim to promote participatory processes through the sharing of data, knowledge and outputs as well as the source code of technical implementations. The openness of these projects and the resources they provide is not always defined the same way. Open data and knowledge is commonly defined by Open Definition[20] as follows: *'Open data and content can be freely used, modified, and shared by anyone for any purpose'*. The most successful example of open knowledge sharing is Wikipedia, which was founded in 2001, the same year as Creative Commons, which aims to facilitate the sharing of creative works in the public domain and is used by Wikimedia to publish all of its content. Wikipedia is an online encyclopedia where

[20] http://opendefinition.org

anyone can become a contributor of content that is interlinked and publicly shared. According to internet rankings, it is in the top five most visited sites on the web[21]. Although just about anyone can edit and add content to Wikipedia, it has been highlighted that the shared information has value and quality like any other professional source of knowledge (Giles, 2005). Other notable projects are Khan Academy, for publishing free educational resources, and the Open Knowledge Foundation, which promotes and shares free information. Open knowledge also extends to publishing scientific research, where people advocate for open access to publications, especially if the research is publicly funded (García-Peñalvo et al., 2010).This also includes publishing of all the data and tools to reproduce the results in the benefit of public knowledge (Molloy, 2011).

1.3.2. Open Source

However, with software, it is not always clear what is freely usable and to what extent there is a difference between 'open' and 'free' software. The difference between open and free is that open is software people can access the source code of, fork (i.e. download and replicate) and change to work within their contexts. The initial idea of *open source* software was proposed by the members of the Free Software Foundation[22], who promoted sharing software and design documents for reuse and repurposing in other projects. Open-source software are published using open-source licences[23], which dictates the specific rules for how people can use and modify the source code. Perhaps the most well known and successful open-source project is Linux, created by Linus Tonvalds in 1991, the kernel code for which has been used in multiple popular operating systems, including the UNIX-like GNU operating system initiated by Richard Stallman. Nowadays, sharing and collaborating on software development results in geographically dispersed programmers working together on projects through online collaboration interfaces (West and O'Mahony, 2005) such as GitHub[24], Bitbuket[25] and SourceForge[26]. Nowadays, models of free and open-source software are becoming more popular, to the point that even commercial companies are publishing their software on these platforms, potentially sharing their intellectual property with competitors (West and Gallagher, 2006). This extends to the idea that there is 'strength in numbers' or, as Linus's Law states (also proven by the development of Linux Kernel), *'given enough eyeballs, all bugs are shallow'* (Raymond, 2001), meaning that if there are more people working on a project, it will ultimately produce higher quality software. In addition to better quality software, people working on these projects have also reported that they are motivated by the participation and sense of belonging to a collective and the sharing of knowledge and skills (Lakhani and Wolf, 2003). People working in the spirit of transparency and openness, sharing skills and knowledge on these new virtual platforms, enables trust and opens up new opportunities for these systems to promote new ways of using them for social good. Open-source software could be a beneficial resource to tap into in order to support active

[21] https://www.alexa.com/topsites
[22] https://www.fsf.org/
[23] https://opensource.org/licenses
[24] https://github.com
[25] https://bitbucket.org
[26] https://sourceforge.net/

citizen engagement in the smart city; however, working with such software requires software development skills and knowledge.

1.4. Open Data

Then, potentially the most transformative branch of the Open Movement is open data, i.e. data that *'can be freely accessed, used, modified and shared by anyone for any purpose – subject only, at most, to requirements to provide attribution and/or share-alike'*[27]. As defined by Open Definition, it has to be open both legally (licensed under open licence) and technically (without attached costs) in bulk and machine-readable format. Among such data is OGD (Ubaldi, 2013), one branch of open government alongside other open governmental initiatives like Open Budgeting, data that is collected or whose collection is commissioned by governmental entities and then published using open licences. The main aims for publishing this data are to serve the transparency and accountability of government, unlock the potential of the social and commercial reuse of the data, and also to promote new ways of democratic participation (Jetzek et al., 2014). Governments, including the UK's, are publishing data about land use, the environment, governmental services, citizenry and their health and public spending and expenditure. However, the breadth and depth of the data published differs from country to country and is also constantly changing as new legislation, data collection methods and rules are implemented. Open Movement advocates and organisations are keeping their hand on the pulse by releasing yearly reports about the current state of open governments, the most well known of which are Open Data Barometer[28], published by the World Wide Web Foundation, and Global Open Data Index[29], published by Open Knowledge International, which aim to document outcomes, share ideas and knowledge, and spark discussion around open data in governments. Despite the popularity of the open data movement, there are still a lot of unanswered questions relating to what it means for active citizens wanting to participate in the smart city and how open data is serving the needs of new modes of participation.

1.4.1. Finding and Accessing Open Data

One way to make data open is to make it available for others to use. The preferred method of making open data available is usually through open data portals. Leigh Dodds, chair of BathHacked[30] and Director of Advisory at the Open Data Institude[31] defines a most basic open data portal as *'a list of datasets with pointers to how those datasets can be accessed'* (Dodds, 2015). While this definition can be seen as the *minimum viable product (MVP)* for an open data portal, when a data portal needs to cater to a wide range of audiences (Davies, 2010), it is worth looking into a more sophisticated solution. The usual functions of open data portals can be categorised into two groups: (1) discoverability - functions that help index and organise

[27] http://opendatahandbook.org/glossary/en/terms/open-data
[28] https://opendatabarometer.org
[29] https://index.okfn.org/
[30] https://www.bathhacked.org
[31] https://theodi.org

the metadata in order to make data discoverable and (2) interoperability - functions that help interlink the data from different sources and make it programmatically accessible through an Application Programming Interface (API). Table 1.1 lists the most popular software systems with their features and example usages.

Portal	Maintainer	Discoverability	Interoperability	Openess	Example
CKAN	Open Knowledge	Metadata API	RPC API	Open Source	daten.berlin.de
OPENDATASOFT	OpenDataSoft	Catalogue	RESTful API	Proprietary	datos.cdmx.gob.mx
PUBLISHMYDATA	Swirrl	Linked Data	SPARQL	Proprietary	statistics.data.gov.uk
DATAPRESS	Datapress	Catalogue	RESTful API	Proprietary	datamillnorth.org
DATASHARE	RedBridge	Catalogue	RESTful API	Proprietary	data.redbridge.gov.uk
JUNAR	Junar	Catalogue	RESTful API	Proprietary	data.cityofpaloalto.org
SOCRATA	Tyler Technologies	Metadata API	RESTful API	Proprietary	data.cityofchicago.org
INSTANTATLAS	Esri	Catalogue	Map Creator	Proprietary	suffolkobservatory.info
CUSTOM	Newcastle University	Metadata API	RESTful API	Open Source*	urbanobservatory.ac.uk

Source: Author

Table 1.1 List of popular data portals, their features and example usages

Using these technologies, but not exclusively, governments have spin-off portals, such as open.canada.ca and data.gov.uk for Canada and the UK, respectively, with many others countries following their footsteps (Open Data Barometer), enabling people to search and download governmental open data. With the 'father of the World Wide Web' Tim Berners-Lee, among others, calling for *'raw data now'*, publishing open data has increased massively over the past ten years, and there are over 170 open data portals in the EU alone, according to the Open Data Monitor[32]. Although an abundance of data has already been published, there are still datasets not being disclosed by governments. One way that citizens can gain access to such datasets, provided it does not violate any restrictions, is by going through a Freedom of Information (FOI) request. However, processes linked to submitting a FOI request can be complex, lengthy and hard to document for reuse. To address these issues, MySociety, a UK-based charity, developed a simple online tool called WhatDoTheyKnow[33] to aid people in submitting FOI requests and sharing the process and results with everyone. Using this approach, people have made over half a million requests, resulting in new datasets being released on the web.

> *'[,...] if people put data onto the web [,...] it will be used by other people to do wonderful things'* (Berners-Lee, 2010).

It is important that data is out there because it promotes transparency and encourages discussion, but it has turned into a 'numbers game' with little context. In this regard, there is a lot of ambiguity surrounding what makes a government open and what the purposes of open govermental data are – is it merely a political statement or a call for technologies that make that data usable (Renee E. Sieber and Johnson, 2015). People have critiqued that the *'measure of openness'* for governments is flawed as it only takes into account the number datasets and repositories published online (Davies, 2010) and that there are more nuances to the effective reuse of open data than simply the ability to access (while not even the case for everyone) the generated data files (Gurstein, 2003, 2011). Similarly, Open Data Barometer's latest report points out that there is a lack of work

[32] https://opendatamonitor.eu
[33] https://www.whatdotheyknow.com

being undertaken around inclusivity surrounding open data (World Wide Web Foundation, 2018). Clearly, making data available is not the same as making it accessible and usable for everybody, especially for citizens who want to engage in active forms of civic participation through data. One of these necessities for what data needs to become in order to make this vision of fully engaged citizens or data engaged citizens a reality is to get data from *available* to *usable*. A similar point is made by Gurstein (2011), who proposes a model of the *'effective use of data'*, as opposed to simply open data, that includes far more steps than just making data available in order to tackle the 'digital divide' around data.

Open data from governments and private companies could potentially provide valuable input for citizens to leverage so as to participate in democracy and help improve their communities. Working with data, however, requires specialist know-how, tools and understanding of data science techniques – skills that community groups and citizens often lack. Even with tools like WhatDoTheyKnow, which simplify requests for data, there may be problems with not knowing what data would be useful in particular contexts, especially when the request consists of having to explain what data one is looking for in a 'free text' field, with limited guidance and ways for pulling in context.

Perhaps the biggest misconception around open data is that it comes in a 'raw' state. In fact, data is always *'influenced by potential uses, expectations, contexts, and theoretical constructs'* (Tuomi, 1999). Heidegger (1976) and Polanyi (2009), among other philosophers of knowledge, have criticised the empirical model of epistemology, arguing that an object's being and our interpretation of it exist in an entwinded state, meaning that we can not observe an object without it being part of its current interpretation. Hence, it is impossible for humans to take data as raw material and obtain knowledge from data through empirical observations.

In the form that open data is made available to citizens, it does not fully lend itself to being used in supporting new forms of civic participation. Citizens are more likely to gain from data exploration tools that wrap data in some narrative that positions data in context and helps to tell a story. Taylor et al. (2015) have emphasised the importance of data production and use being bounded to a place with geographical and social dimensions, giving it context and making it usable. There are already examples of systems such as MapIT[34] and TheyWorkForYou[35], which enable people to access open data linked to a specific postcode, or Know My Neighbourhood[36] and Local Insight[37], which enable community organisations to link different open governmental datasets to an area of interest in order to aid local decision-making or improve service delivery (Chapter 4).

Openly sharing data and knowledge, whether originating from official sources or being shared by Wikipedians on the web, takes it closer to becoming knowledge that is understood and accessible by everyone. In addition, if knowledge is shared, it can be questioned, confirmed or

[34] https://mapit.mysociety.org
[35] https://www.theyworkforyou.com
[36] https://neighbourhood.knowmyarea.org
[37] https://local.communityinsight.org

contested – either through public critique or systematic reviews and meta-analyses – ultimately leading to an understanding of different perspectives and a better quality of focal knowledge. Not openly sharing data and knowledge or deliberately making access obscure and difficult for reasons such as the data being *too complex to understand* or *too difficult to interpret*, or that it *might be misunderstood*, are not valid reasons for not putting it out there. The issue is not that the data is complicated but that the presentation of data is complicated or missing crucial context to make sense of it, meaning that there is a representation issue rather than a data issue, which is more related to the information-systems problem-solving agenda; however, this can be approached from the community problem-solving angle (Stillman and Linger, 2009). Furthermore, if something is hidden from people, it cannot be discussed, and its representations cannot be worked on.

1.4.2. Working with Open Data

In order for people to start making sense of data and working with data, it needs to be condensed into something that is comprehensible to the human senses. This is where the visual communication of data, or data visualisation, comes into play, aiming to communicate the information contained in the data. There is often an assumption that data visualising is just reflecting the information the data holds. However, as there is no such thing as 'raw data', there is also no such thing as direct or 'raw visualisation'. Data visualisations, although based on data, are already analysed and edited forms of data projections that convey the message of the data analyst or editor. The aim of data visualisation is to support the analytical process and help with information discovery and knowledge creation, which means it is also important to identify the type of audiences that will use the data representation in order to better contextualise and present information and aid them in knowledge creation (Chi and Riedl, 1998).

People might theoretically have access to the same data, but not everyone has the know-how, skills and resources to turn data into something useful. Manipulation of data in a spreadsheet can be manageable for the vast majority of people; however, when data requires some cleaning, merging or reducing, more specialised tools such as Matlab[38], Python[39] or R[40] are needed. Although still considered a professional trade, advances in information technologies and the open-source movement have produced multiple visualisation libraries such as Matplotlib[41], ggplot2[42], and D3.js[43], thus enabling more people to experiment with data visualisation. However, data visualisation tools often require software development skills in order to manipulate the data and produce visual outputs. With this in mind, software like Tableau and Gephi are trying to break down such barriers and use graphical user interfaces (GUI) to help users visualise data. Advances in web technologies also permit data manipulation on web platforms in an easy 'drag

[38] https://www.mathworks.com/products/matlab.html
[39] https://www.python.org
[40] https://www.r-project.org
[41] https://matplotlib.org
[42] https://ggplot2.tidyverse.org
[43] https://d3js.org

and drop' manner. Examples of such tools are Carto[44], for manipulating geographic data and making interactive maps, and Charted (Sall, 2014) and RAWGraphs (Mauri et al., 2017), for automatic chart creation from files, online spreadsheets, web APIs or other data sources, which can be easily shared with others. Although there are open-source web tools for making graphs, people still have to choose the right visualisation for particular data to convey the information. DataVizProject[45] from Ferdio makes good headway towards helping people make the right choice for their data by having simple visual representations of different types of charts, categorised by family, input, functionality and shape.

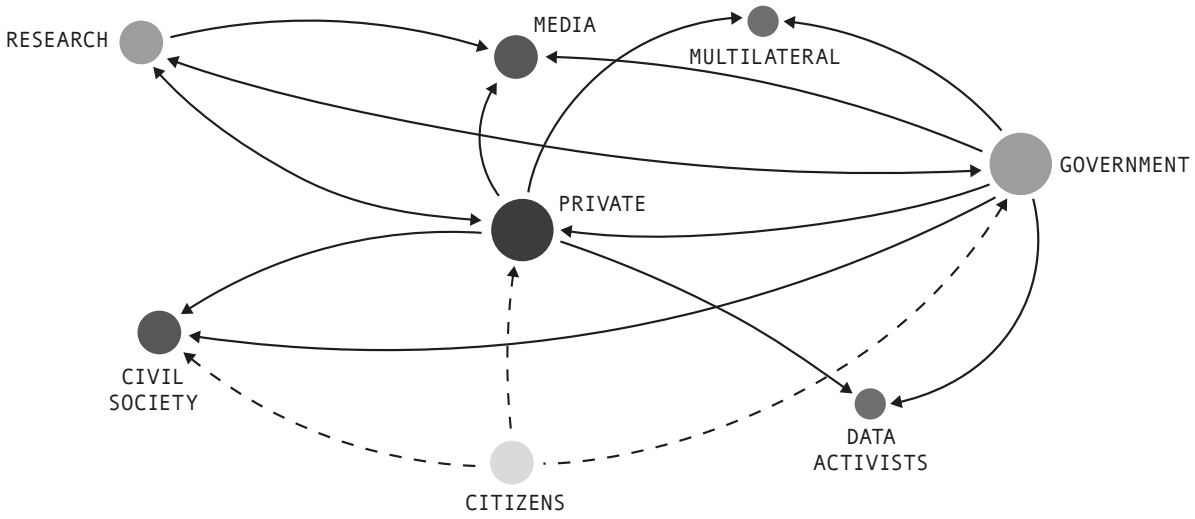
Even if data is made accessible and visible and is formatted nicely, or put into 'pretty' graphs, it does not necessarily mean that the average person (or citizen) can use it. Presenting data effectively requires both artistic and technical skills for identifying the message or insight it needs to convey and choosing the correct representation guided by visualisation techniques (Chi, 2000). Working with data is a complex task that requires both technical and cognitive skills that people spend years training to achieve, working on asking the right questions; figuring out where to look for data and how to access it; improving their visualisation craft; experimenting and figuring out the best practices for different types of data that suit different contexts; and ultimately becoming professional data analysts and visualisers. Sites like InformationIsBeautiful[46], Visualising Data[47] and Flowing Data[48] are good examples of people achieving that level of professionalism. Furthermore, the data in question, the data that the smart city uses and citizens produce, is data at scale and complexity (i.e. big data) – involving quite complex constructs such as demographics, geospacial patterns and trends – and is inherently complex to process, analyse and understand, which makes it far more challenging to make it work for ordinary citizens.

Indeed, it appears that, at the moment, citizens need professional analysts or information engineers to mediate the processes that make data work for citizens actively participating through data. Going forward, there needs to be a better understanding of the role of professional data analysts in these processes. Furthermore, perhaps there is a process that makes citizens smarter and able to participate directly in decision-making that involves data at the city, country or even global level. To make it truly democratic, the challenge is how data science skills can become transferable to ordinary citizens, so they can do it for themselves, or built into IS that can automate those processes for them. Whether or not professional expertise can be replaced or whether the solution is a mixture of professionals, automation, and smarter citizens are challenges that the research in this thesis explores and addresses.

[44] https://carto.com/
[45] https://datavizproject.com
[46] https://informationisbeautiful.net
[47] http://www.visualisingdata.com/
[48] https://flowingdata.com

1.4.3. Stakeholder Groups

From smart city management to citizen sensing, data is at the epicentre of our research, communities, and world. Despite the pervasiveness of the term, there is still little critical understanding of what data might mean for different people in different contexts (e.g. smart cities, business, research and the community). Whether the term may be used to refer to 'big data' analytics and statistics or to the much broader collection of diverse 'things' like images, texts, and audio, interpretations of data and the meaning people draw from it is intrinsically liked to the motivations, contexts and situations people find themselves in. Such diversity and ambiguity raises tensions, particularly in contexts in which professionals and non-professionals are required to collaborate around data to communicate ideas and produce knowledge. From a Foucauldian perspective, there is a need to identify a common language, understandings and representations of data to promote discourse (Foucault, 1982a,b), which could aid people in knowledge production and help balance the power relationships between citizens and decision-makers, moving towards citizens becoming active users of data and being part of the decision-making processes in the smart city.



Source: Author

Figure 1.3 Stakeholder group network around data

Stakeholder	Roles
Government	data generator, data publisher, data user
Business/Private Sector	data generator, data publisher, data user, (intermediary)
Research	data generator, data publisher, data user
Journalists/Media	data user, (data publisher)
Civil Society	data user, (data publisher)
Multilateral Organisations	data user
Data activists	data user, (data publisher)
Citizens	data generator

Source: Author

Table 1.2 Stakeholder group roles in a data network

When designing civic technology that supports communities and institutions working together with data, it is important to understand different stakeholder groups and their roles and

motivations regarding generating, providing and using data. This classification of stakeholder groups is drawing on work carried out by the State of Open Data (Davies et al., 2019) project and its contributors. Figure 1.3 illustrates the relationships between different stakeholder groups around data. The number of connections define the size of the stakeholder and the arrows dictate the direction of the relationship between the different stakeholders. The more inward connections a group has, the darker the colour of the circle. The dashed line indicates a purely data generation link. Links between different stakeholders clearly illustrate the transactional relationships that the current state of data enforces between those in power and those who are subjects of it. Furthermore, 1.2 lists the different stakeholders under examination in this thesis and their roles in the data network. Some of the roles are in brackets, which means that they are secondary roles for a particular stakeholder. Understanding relationships and the roles in this network helps to identify barriers and opportunities to reconfiguring these roles and to making data more inclusive by increasing participation through the use of data, with the aim of supporting the vision of citizens as active users of data.

Government

From ancient Egypt and China to modern day countries, governments have held censuses to gather information about their citizenry in order to inform decisions and understand or avoid risks. Historically, censuses were carried out to count citizens for tax purposes and to inform the government about eligible young men for military service. After collecting the data, it took about a year to add it up and provide the statistics. Although the techniques have changed and efficiency has increased when moving to the digital age of computers, the aims of a census have remained pretty much the same. Governments are well established data collectors, and the processes in place for polling their citizens have been perfected over the years. As mentioned previously, opening up that data is becoming a common practice for governments (local and national) and is increasingly being integrated into the way they operate. In addition to data generated through censuses, governments are also releasing data about how they operate and spend public money (Marshall et al., 2016). Introducing open data initiatives puts pressure on governments to 'reform' current processes of governance in order to guarantee the continuance of these initiatives as public services go through a digital transformation. Data portals, currently the main channels for publishing OGD, are still supply-driven and lack engagement and tools for the effective use of data to help reach all stakeholder groups (Davies et al., 2019, sec. 3).

Business/Private Sector

Data is probably the biggest business for private sector companies operating in ICT. In addition to the data that private sector companies generate themselves, having the tools and know-how to tap into open data, they also benefit from the data published by other institution. It has been suggested that open data has the potential to help generate $3 to $5 trillion of economic

value annually in the US[49]. Alongside transparency and accountability, this huge potential for economic growth has also become a driver for some governments to open up their datasets for reuse. The private sector has, for many years, been an avid user of open data, highlighting its value to help understand demand, be used in strategic planning and produce useful products and services (Open Data 500[50] and Open Data Impact Map[51]). The value of open data for the private sector is clearly illustrated by the Open Data Impact Map, which provides a searchable catalogue of open data projects, as two-thirds of the organisations on the list are for-profit companies (Gurin et al., 2019).

In addition to private sector leveraging open data to improve their products and services, a number of companies have also become 'data intermediaries' (Gurin et al., 2019) by helping governments publish their data or repackaging existing data for public and business use. There are companies that provde data portal software for governmental and other institutions (Section 1.4.1) based on a data-as-a-service model to help publish their data, but recently more companies are coming out with data-broker tools catering for consumers and the general public, such as LG Inform Plus[52] from the Local Government Association, which combines multiple sources of open data and makes them available through APIs and file downloads. There are also companies that package data and try to visualise data better by using GIS reporting and graphs, including companies like OCSI and HACT, which provide software for understanding trends in housing market[53] through OGD, Porism[54], which packages OGD for informing neighbourhood plans, and MySociety[55], which has developed multiple tools for making open data more accessible and promoting new routes for democratic participation from the public.

Although private sector organisations are mainly data users and generators, they are also increasingly becoming data publishers by opening up anonymised user datasets or accumulated data sources for researchers and developers alike. Companies like AirBnB, Uber, Strava and Zoopla have recognised the potential benefits of opening up data and have published some of their data in order to improve the quality of or find new business value for their data, but also to promote collaborations to help understand and tackle some societal problems. This is enabling companies to collaborate with one another and with research facilities. GovLab's Data Collaboratives Explorer[56] categorises data collaboratives into six types: *Trusted Intermediaries*, *Prizes and Challenges*, *Research Partnership*, *Intelligence Products*, *API*, and *Data Pooling*. The Alan Turing Institute[57] in London is a good example of such a collaborative approach, where resources are often shared between private companies, governments and research institutions. Another example of this is Positium LBS[58], a university spin-off company that collaborates

[49] https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information
[50] https://www.opendata500.com
[51] https://opendataimpactmap.org
[52] https://about.esd.org.uk
[53] https://ocsi.uk/community-insight
[54] https://porism.com
[55] https://www.mysociety.org
[56] http://datacollaboratives.org/explorer.html
[57] https://www.turing.ac.uk
[58] https://www.positium.com

with telecommunication companies, government and research facilities to provide statistical indicators for public institutions about the population derived from big mobile data.

Journalists/Media

The role of journalists has always been to keep the public informed about governmental accountability, business developments and academic and scientific research. In the digital era, *Data-Driven Journalism* (Ausserhofer et al., 2017) is a term that is becoming widely recognised as part of the journalistic agenda, i.e. using data as a resource to investigate and/or shed light on matters of public interest. Becoming an effective journalist nowadays requires the skills of a data scientist, which has resulted in traditional media outlets dying out and being replaced by online ones. Computer-assisted reporting is becoming mainstream in journalism, and major news outlets have dedicated data visualisation and data science departments focusing solely on storytelling with data (e.g. The New York Times and The Guardian). These new 'programmer-journalists' (Parasie and Dagiral, 2013) are equipped with new skills and tools to obtain, analyse, visualise and even open up some of the data for citizens, e.g. initiatives such as Fatal Force, where journalists have created a database of unreported killings by the police to highlight racial injustices, or making governments responsible by releasing data about chemical attacks and human rights violations in Syria[59].

Through automated tools developed by 'programmer-journalists', who are often also part of open movements, some of the data can potentially be reported automatically when new data is collected or released, making way to 'automated journalism' (Davies et al., 2019, ch. 27). The interdisciplinary nature of data journalism promotes collaboration between different people and organisations to help address social issues. However, it has been reported that the working groups are usually small and enclosed, focusing on providing descriptive statistics or visualisations to exploratory questions (Choi and Tausczik, 2017). In addition, the majority of the stories in data journalism are focused on political rather than social issues and mainly use existing open data (Loosen et al., 2017). With the increasing popularity of big social media platforms becoming the new news outlets, and politicians creating 'fake news' (Carson, 2017), it is even more difficult to spark collaborative efforts between diverse groups of journalists, data scientists and communities of practice to use data for social good.

Multilateral Organisations

Multilateral organisations (Davies, 2010, ch. 28) like the United Nations (UN), World Bank, World Trade Organization and the Organization for Economic Cooperation and Development (OECD) have a role to play as promoters of open data and intermediaries to connect different stakeholders to work together and use data to improve people's lives. Having reach across different countries and continents, these organisations also have a key role to play in disseminating knowledge through consortiums and conferences. However, a critical lens needs to

[59] https://www.nytimes.com/interactive/2018/04/13/world/middleeast/syria-chemical-attacks-maps-history.html?mtrref=www.nytimes.com

be used to assess whether the successes of open government and open data can be applied in developing countries to achieve the desired social impact (Gigler et al., 2011). Additionally, these organisations also publish data about world economics [60] and demographics [61]. Multilaterals often have the ability to tap into resources that are far greater than any one government and use new techniques and methods to require, process and report on data. In addition to the world organisations, there are smaller non-profits that promote and use data, aiming to create better standards and more cohesive data ecosystems. Bodies such as the Open Data Institute, which helps develop and promote open data tools and initiatives, and Open Knowledge International, which advocates for knowledge sharing, are also responsible for keeping their finger on the pulse of developments around OGD through publishing Global Open Data Index reports [62].

Civil Society

Civil society – community organisations or groups of citizens with common interests that often mediate data to the rest of the community – is perhaps the most important link in realising the effective use of data within communities. Civil society is predominantly a data user, through conducting research with data and often publishing findings in order to shed light on different matters. Although citizens are mainly data generators, there is a particular set of enthusiasts among civil society called data activists, 'hacktivists' or civic hackers, who use civic data to bring concerns to light and contest decisions through enquiries and data visualisations (Schrock, 2016; Taylor, 2005). They often make strong political statements and act as watchdogs of governments and big corporations. In addition to individuals, there are also groups of people who operate through Meetups[63] and collaborative projects, consisting mostly of professionals working with data or researchers who want to contribute their time and skills for the good of society. Examples of such groups are Open Oakland[64] in San Francisco, US and NE Data[65] in Newcastle, UK. There are also networks such as OpenGov[66] in the UK that combine active citizens and civil society organisations to help make government work better for people through transparency, participation and accountability.

However, when it comes to pushing the open data agenda, civil society has far more obligations to uphold. In addition to being the bridge between governmental institutions and community groups, it also helps raise awareness around open data, sets the standards and course for data publishing, consults with and trains people on how to access and use open data, publishes tools to realise open data needs and advocates for new datasets to be released as open data (Davies, 2010, ch. 24). Members of civil society often even help gather data and openly publish it to promote advocacy and push for change. Being the driving force behind pushing the use of open data for social good, civil society has the biggest responsibility to citizens but also the most challenging

[60] https://data.worldbank.org/
[61] http://data.un.org/
[62] https://index.okfn.org
[63] https://www.meetup.com
[64] https://openoakland.org
[65] https://www.nedata.uk/
[66] https://www.opengovernment.org.uk/

role to play. However, it is not always clear where civil society ends and other stakeholders begin, in addition to there being an overlap with the research community, data donors and private sector companies (Davies, 2010, sec. 3).

Although civil society is perceived as being a powerful force that can achieve anything it sets out to do, the reality is that it often does not possess the resources it needs to help communities with their data agendas, especially smaller groups who are struggling with a lack of financial support or other resources to further their causes. Moreover, the unwelcome competition between civil society groups trying to do the same thing and competing for the same pots of funding is also problematic (Johnson et al., 2018) as this makes it more difficult for funding bodies to figure out which group would potentially have the greatest impact on society, thus often ending up with uneven or unfairly distributed resources. At the same time, the gap between the people who are already engaged and making use of data and those who may gain from it the most but do not have any routes to engagement widens. Instead of competing, non-profits could collaborate with each other, be more transparent and use data and technology to secure more funding (Goecks et al., 2008; Johnson et al., 2018). However great the potential of data to aid community organisations in their agendas, it is yet to be unlocked by everyone.

(Academic) Research

Researchers – from academic, policy and non-profit areas – have been engaging with open data from its early existence. Academic research has been turning to open data as a resource to understand populations (using data from censuses) and improve the scientific value of their research, in addition to also studying the impacts of open data on society (Hossain et al., 2016; Van Schalkwyk and Verhulst, 2017). Furthermore, initiatives such as ODDC and the Open Data Research Symposium are dedicated to examining open data research and its impact. However, research is still heavily focused on the measurable things in open data, the quantity of published datasets and how they are made available, often failing to look at the social impacts of data in research. Davies et al. (2019) stress that researchers need to go further to make sure that the work they do is relevant to the communities who have been identified as the potential beneficiaries of open data.

Fortunately, more and more researchers are engaging in not only looking at the impacts of open data to communities but also trying to help communities use open data to instrument social change, thus becoming data professionals that are trying to make the vision of actively engaged citizens or citizens as active users of data into a reality. For example, individual research agendas have sought to understand current open data practices amongst NGOs and design new tools that would unlock open data for these organisations (Erete et al., 2016; Marshall et al., 2016), often driven through larger initiatives that pair researchers up with civil society groups or local and transnational NGOs. Such initiatives, operating under the banner of *'Data Science for Social Good'*, are usually part of universities or research facilities that run fellowships and mentoring programmes to aid community organisations harness the power of data science on open data.

Examples of such programmes are the Data Science for Social Good Foundation[67] and Turing's DSSG programme[68]. In addition, there are also volunteering initiatives, such as DataKind, for experts who want to contribute their skills to the common good of humanity.

Researchers are also publishing the data accumulated from their research projects as open datasets. For example, the National Urban Observatory facility, which combines six urban observatories in the UK under one umbrella, has accumulated and openly published over a billion observations about the city environment. Furthermore, researchers are also helping citizens collect data through citizen sensing (Balestrini et al., 2014), new media and content creation tools (Bartindale et al., 2019; Crivellaro et al., 2016; Schofield et al., 2015; Taylor and Cheverst, 2010) and community commissioning technology (Garbett et al., 2016). All of these activities are contributing to the vision of citizens as active data users engaged in the production and use of data. Although researchers often work alongside communities and citizens to develop tools for access, use and production of data, the solutions they offer are often highly dependent on the researchers, thus lacking longevity and sustainability. The challenge here is to explore models that researchers could apply to support citizens actively participating in the smart city on a long-term basis.

Citizens

Davies et al. (2019) do not include citizens as one of the stakeholders (or it is meshed together with civil society) in the open data ecosystem. Perhaps the hope is that other stakeholders will do the work for citizens towards unlocking the potential of open data for everyone. However, for the focus of the research in this thesis, *citizens* as stakeholders deserve particular attention. In this regard, the aim is to explore ways to support citizens to become more active participants in the smart city using data, i.e. to become smart citizens.

There is often an assumption that opening up data makes it accessible for everyone (Berners-Lee, 2010); however, most of the population does not possess the skills or tools necessary to access and use data. In most cases, citizens are actually generating data or data is generated about them without them having any say in how such data should be used (Section 1.2.1). However, this is a rather impoverished notion of the citizens who live in cities; instead, we should think about citizens in regard to how they might be more actively engaged in decision-making, how they might be more than sources of information for the smart city, and how they can be transformed into the makers, producers and consumers of data. Furthermore, they might even want to use data for their own civic purposes, and they might even want to commission data from other people that they currently cannot get access to. Some people strongly believe that open data is not for the public and that *'raw data is for people who know how to use raw data'* (Whong, 2015). However, as mentioned before (Section 1.4.1), data is never raw – it always represents the bias of the collector and the observer (Gitelman, 2013; Tuomi, 1999). It is not contradictory that the idea of open data is not for the public per-se, it is just highlighting that the current

[67] https://dssg.uchicago.edu
[68] https://www.turing.ac.uk/collaborate-turing/data-science-social-good

representation of data – whether it is coined as 'raw data' or 'open data' – is not accessible or usable to everyone in this form. Data needs to go from open → accessible → comprehensible → usable in order to be used in active citizen engagement. There is no one-size-fits-all solution or representative format of data. When designing platforms to support communities and institutions working together with data to address problems in society, significant challenges – both technical and social – arise around the translation and representations of data that have yet to be addressed.

Considerations of what data is valuable, to whom, and in what form needs to be taken into account to better support communities collaborating around data. There is an obvious lack of resources in terms of funding and access to tools; however, if there are existing competencies in the community, they need to be linked up through developing discourse because discourse is essentially about the production of knowledge through language and through the sharing of practices (Foucault, 1982a). Developing discourse and relationships between stakeholders has the potential to increase social capital within the community. In a time of austerity (Lowndes and Pratchett, 2012), there is a great deal relying on local initiatives and the availability of social capital. This also includes human capital (Coleman, 1988) – combined knowledge and skills of individuals in the community – which is perhaps the most important resource for building community resilience. Exploring relationships and the roles of stakeholders with regard to maintaining, accumulating, sharing and using social capital – or *bonding*, *bridging* and *linking* (Granovetter, 1983) – could help understand the important variables that contribute to successful active citizen participation and action. There is empirical evidence that social capital is far more effective than outside support at times of crisis (Aldrich and Meyer, 2015; Morton and Lurie, 2013). Helping to establish connections and increase shared knowledge, local organisations and citizen groups are able to respond quicker and adapt to changing situations. However, this does not only apply when disaster strikes. At a time of economic disparity, a strong knowledge society and social capital are key for tackling inequalities (Lin, 2004, ch. 7) and moving towards active citizen participation in the smart city. There is a need for better understanding the impacts for the different actors involved and the social capital surrounding the active use of data by citizens. This also relates to the role of expert professionals and looking at the ways their expertise can be replaced, i.e. either handed down to citizens themselves, pushed into smart interfaces that do the intelligent processing for them, or finding a middle ground or balance between the two. More work is needed to understand how much can be done by computers and IS and how much training of citizens is required.

1.5. Knowledge Society

1.5.1. Communities, Publics and Networks

The usage of social capital – the exchange and sharing of skills and other resources in order to produce knowledge Coleman (1988) – is usually contained within a specific group or community that shares an interest or issue or are working towards a common goal. In the literature, such

communities have been identified differently, and there is confusion surrounding the topic. Dewey (1954, Orig. 1927) refers to these communities as *publics* and breaks that down by saying that the reason why publics form is because individuals are all affected in some way by an issue or social condition that makes them come together. However, these people may be complete strangers, and they also may be affected in different ways by an issue. Building on Dewey (1954) and on Warner (2005) ('publics' and 'counterpublics'), more modern takes on publics have manifested themselves, such as Mccarthy et al. (2015), and Le Dantec (2016). For Le Dantec (2016), the design of publics is looked at from a point of attachment to issues and linked to infrastructures (Le Dantec, 2016, ch. 2). According to Mccarthy et al. (2015), the conception of publics is more linked to *media publics* – strangers who come together and share things through different medias and do so because they have a common interest, even though they are quite diverse and different people. A *Community of Practice* (CoP), however, is a group of people who share a common set of ways of doing things, common practices and a common set of values. They might have different levels and types of capability and capacity, but they are all working together to address and achieve common goals. According to Lave and Wenger (1991), the important structural elements of CoPs are the *domain* of knowledge and the notion of a *community* and its *practice*. Domain is some form of common denominator that defines a group of people, the notion of a community relates to the connections and interactions that individuals share in the domain, and the practice is the focus and activities they engage in to produce, share and maintain the knowledge. A classic example of a CoP are healthcare professionals – hospital staff, nurses, doctors, etc. – who aim to work together to deliver a good healthcare service to people. On the other hand, a *community of interest* can simply be referred to as individuals who have a common interest in something. They can, however, be completely different people from different backgrounds and not share any common practices. In some cases, they do not even have to practice the thing they are interested in.

In the case of the research conducted in this thesis, the groups of people, *'communities or publics'*, are referred to as individuals who are connected by the proximity, interest, issue or any combination thereof. This can be individuals living in close proximity working together to improve the neighbourhood and tackle local issues, but it can also be a virtual community (Rheingold, 2000) that is present on online platforms such as FB, Reddit, GitHub or Quora or mediated through some other medium (DiSalvo et al., 2008). However, one could say that it could also be a combination of publics and CoPs because we have citizens, data professionals, charity organisations, and activists, people with different skill sets and capabilities that can be utilised and used as a resource, all working together to make use of data for civic participation. This combination or mix of different publics and CoPs is actually forming *community networks*. The emphasis here is on the use of technologies and IS to leverage data for community problem solving. It aligns with ideas of Carroll and Rosson (2003) and has the characteristics of the proposed *networked communities* of Clement et al. (2012, ch. 2). In a similar way, Varnelis (2008) uses the term *networked publics* to refer to how individuals and groups engage with each other and communicate ideas and knowledge using the capabilities of modern ICT and media.

The use of publics here emphasises a more engaged population as opposed to one that is passive and consumptive. Varnelis (2008) also points out that networked publics are far more complex in the way people are using the capabilities of ICT to communicate ideas, produce knowledge and take action. Furthermore, these networks are *'bottom-up, top-down, as well as side-to-side'* (Varnelis, 2008, p. 3), thus bridging the public, professional, institutional and personal.

1.5.2. CI and IS

IS and Information System Design (ISD) have strong theoretical and technical backgrounds more often rooted in engineering and computing, with a strong focus on business and commercial endeavours. The aim of ISD is to build a best solution that solves an engineering problem and results in an artefact – an information system. In this regard, the focus in on the design of an artefact to address a particular engineering issue. Other issues, as pointed out by many researchers (Carroll and Rosson, 2007; Hakken, 2003; Hirschheim et al., 1996; Leonardi and Barley, 2008; Stillman and Linger, 2009), such as user groups, interfaces, usability, contexts, choice of technology, and indented uses, become unimportant without the artefact. These issues, however, relate to the social aspects of IS, which scholars involved in ISD started calling *social informatics* (SI) (Kling, 2000; Star and Gasser, 1997). SI puts the users of the system in the centre of the design processes of IS and looks at how to address those non-engineering problems in an organisational context. For Lamb and Kling (2003), it is important to reconceptualise 'users as social actors' in these contexts.

However, working with communities, the main aim is to address the issues that people have or the things they want to change in the community rather than the problems of technology. Similarly, they are not bounded to organisational structures and uses of particular technologies in a specific context. As a result, scholars such as Gurstein (2000), Hakken (2003), Bradley (2006), and Carroll and Rosson (2007) have expanded on these ideas of SI to go beyond the context of organisations and users. A new field of CI Gurstein (2000) has emerged that is focused on the design and implementation of technologies to help local communities address their issues. Later definitions of CI have also taken into account other fields that influence CI in practice and expanded to factors such as equal opportunities, recognition of the importance of the 'lived physical community', user-focused underestandings of ICT, applied social leadership entrepreneurship, and creativity (Gurstein, 2007, p. 12). The overarching aim of CI is to overcome 'digital divides' through the design and implementation of ICT that enables and empowers citizens.

If the focus is on the uses of data and the information system, CI actually has a dual focus: *'first, the conduct of research about the relationship between the design of ICTs and local communities, and second, the implementation of ICT projects in local communities'* (Stillman and Linger, 2009, p. 255), pointing out the still present engineering agenda or *IS problem solving* on top of the *community problem solving* one. In the critical approach to IS provided by Stillman and Linger (2009), despite the differences between IS and CI, it can still be looked at as part of the 'fragmented adhocracy' of IS research, thus helping to provide CI with a theoretical basis.

DOMAINS	ORIENTATIONS			
	CONTROL		SENSE-MAKING	ARGUMENTATION
	INSTRUMENTAL	STRATEGIC	COMMUNICATIVE	DISCURSIVE
TECHNOLOGY	*Information Technology Systems* Hardware and telecommunication configurations; Program structures and modules; Database and file structures			
LANGUAGE	*Formalized Symbol Manipulation Systems* Data models and dictionaries; Data integrity mechanisms; Screen and form design; Model management systems	*Manipulative Communication Systems* Definition of terms and rules; Communication channels; Access rights; Data integrity	*Symbolic Interaction Systems* Speech acts; Intentions; Meanings; Metaphors	*Systems for Rational Argumentation* Arguments; Warrants; Breakdowns; Pragmatic inference
ORGANIZATION	*Mechanistic Social Systems* Tasks; Decision processes; Business processes; Organizational structures	*Political Systems* Power structures; Resource dependencies; Interest groups; Sources of authority; Indirect influence; Negotiated orders	*Cultural Social Systems* Values, beliefs; Myths, rituals; Negotiated meanings and practices	*Systems for Institutional Checks and Balances* Domination free discourse; Justification and minimization of power; Truth and justice; Due process

Source: Hirschheim et al. (1996, p. 17)

(a) ISD object system classes

	Orientations			
	I Control		II Sense-Making	III Argumentation
Domains	IA Instrumental	IB Strategic	Communicative	Discursive
Technology	Information Technology Systems **Systems Engineering**			
Language	Formalized Symbol Manipulation Systems **Comm. Eng'g**	Manipulative Communication Systems **Manipulative Communication Design** Access rights	Symbolic Interaction Systems: speech acts, intentions, meanings, metaphors **Consensual Communication Development**	Systems for Rational Argumentation **Rational Argumentation Design**
Organisation	Mechanistic Social Systems ('ritualized' tasks) **Org. Design**	Political Systems: systems to replicate decision making hierarchies, non-hierarchies surveillance, control, security **Political Organisation Design**	Cultural Social Systems; values, beliefs, myths, rituals, negotiated meanings and practices Social and work arrangements; participatory & community design exercises **Organisational Sense-making**	Systems for Institutional Checks and Balances **Institutional Democracy Design**

Source: Stillman and Linger (2009, p. 259)

(b) Adapted CI object system classes

Figure 1.4 Object system classes for ISD and CI

They draw on the federated framework for ISD of Hirschheim et al. (1996, p. 17) and state that it can be adopted to CI. The fact that Hirschheim et al. (1996) use social action theory and change as a lens to analyse the framework makes it more relevant to CI (Stillman and Linger, 2009). Figure 1.4 shows the original and the adapted object system classes framework. It is presented as a two axis matrix, where one axis shows the *domains of change*, the objects that are changed by ISD, while the other presents the orientations that show the purpose of change instrumented by ISD. Importantly, the framework also distinguishes between two control objects: humans and artefacts. However, there are two issues with this framework from the perspective of CI. First, it is a one-dimensional perspective, as noted by Stillman and Linger (2009), that lends itself to repetition and overlap in concepts across domains; second, the role of technology is only being utilised and prevalent at the instrumentation of control. To address this literal gap in the framework and to contribute to the field of CI, this thesis presents research that explores the role of data and technology in all the orientations of the framework and towards active citizen engagement, participation and instrumentation of change.

Although work has been carried out exploring technologies and mechanisms for citizens as active data users, it has been limited to focusing on one aspect of interactions with data (e.g. production, access, or use) and has not taken the holistic approach needed (Clement et al., 2012, ch. 2) to address the dual focus of this research (Stillman and Linger, 2009). Additionally, research has not looked at such issues thought the lens of digital civics (Olivier and Wright, 2015), civic advocacy, activism and action. Furthermore, a lot of the work around the design and engagement of data with citizens in a smart city has been carried out by researchers from social science or humanities backgrounds rather than from the perspective of an information engineer working with citizens and trying to address community problem-solving activities. According to Bradley (2006), this type of research requires a *'new type of engineer'* to be trained, one who has a much broader grasp of social issues and their importance towards the design and uses of technologies,

while at the same time having the technological know-how and background not to resist ICT and think more about the aspects of HCI (or HDI) and interpretation in CI (Stillman and Linger, 2009).

1.6. Summary

This literature review has provided an overview of the concept of the smart city and the role of citizens within it, as well as offering a thorough review of the different ways that citizens can participate in the smart city through data, either passively or actively. The literature review has helped put forward the concept of a more actively engaged citizen (i.e. a smart citizen) of the smart city who participates in producing, accessing, understanding and making use of data to participate in decision-making processes. Furthermore, the chapter has elucidated the current state of open data; the way it is made available, accessible, comprehensible and usable to citizens; and the stakeholders involved therein, thus ensuring an understanding of what it means to make data usable to citizens in a civic context and for advocating. In addition, the chapter has highlighted the current role that expert professionals would have to play in all this (since the data the smart city is concerned with is large at scale and complex at nature) and whether such expertise can be handed down to citizens or pushed into intelligent IS. Although there is an overlap in critiques of participatory design (or social design) and ISD for active civic engagement and participation, the tools implemented and the research conducted regarding the use of data for purposes of civic advocacy have received little focus, with scant empirical research being conducted. Additionally, the research in this space tends to occur either in humanitarian fields or in systems engineering, with little crossover between the two. Therefore, there is a need for empirical studies with citizens to investigate the active use of data in complex civic processes in order to find a middle ground between technological determinism and social constructionism, thus providing a holistic model developing pragmatic practice that responds to critiques and helps guide such research in the future.

Chapter 2. SenseMyStreet: Sensor Commissioning Toolkit for Communities

With data collection, 'The sooner the better' is always the best answer.

— Marissa Mayer

This chapter describes the iterative user-centred design process, development and deployment of SenseMyStreet (SMS), a sensor commissioning toolkit that has been set up together with Urban Observatory (UO), the UK's largest urban data sensing network. The toolkit enables citizens and communities to use and commission data from scientific-grade environmental monitors, determining where they will be located on their streets and collecting evidence to inform local issues. A strong case can be made for evidence-based decision-making by using data as a resource to make educated and informed decisions. However, there seems to be a disjoin between the people who are affected by these decisions and their access to tools and resources to collect data in order to provide the needed evidence for change. To truly democratise this process and for citizens to become active producers and makers of data, there is a need for new mechanisms of citizen data production. This chapter begins by explaining the motivations for establishing a sensor commissioning toolkit for the community arising from critiques in the literature and the current state of the art of citizen sensing. In addition, it also outlines the context of the study, documents the process of designing and setting up the toolkit, provides an overview of the technical specifications, and describes the development and deployments carried out by the author in collaboration with technicians from the UO. Furthermore, an analysis of the use of the toolkit and its associated platforms and technologies is provided. The toolkit continues to operate as part of UO at Newcastle University, enabling communities and citizens to gather data relevant to issues important to them at a hyper-local scale.

Related Publications and Acknowledgements

- Section 2.2 has some similarities with the study context in Peacock, S., Puussaar, A., and Crivellaro, C. *Sensing our Streets: Involving Children in Making People-centred Smart Cities*, a publication under submission that is co-authored by the PhD candidate. It is due to the fact the two studies shared some common participants and took place in the same context.
- Acknowledgements go to The West Oakland Air Quality team for doing the initial research comparison of environmental data platforms, Thomas Maskell for co-designing the initial workshop, Sean Peacock for helpful discussions and supporting the engagement work, Rob Anderson, Thomas Nappey and Edward Jenkins for their technical expertise on various technologies, and Prof. Philip James, Dr. Jennine Jonczyk, Neil Harris, Dr. Luke Smith, David Pearson, Richard Turland from the UO for helping to support the SenseMyStreet project.

2.1. Motivations

This section expands on the literature and related work presented in Chapter 1 surrounding active citizen participation and production of data in the smart city, with a specific focus on work related to environmental data in citizen sensing and how such data is currently generated; who has access to it and in what way; how it can be used for community problem-solving activities; and how the generation, access and use of this data may be democratised in the future.

2.1.1. Sensing the Urban Environment

Monitoring the urban environment is an important part of figuring out how the city operates and how people are influencing the city as a whole through their everyday activities. The collected data from the city can be useful for understanding these complex processes, but it can also be used for monitoring and evidencing the changes implemented in the city. Motivations for monitoring could be, for example, creating a digital twin of a city (Bolton et al., 2018), understanding city soundscapes (Adams et al., 2006) or air quality (Department for Environment, 2018) to inform policies around planning and infrastructure changes in the city, or building a smart city analytics tools for smart service delivery (Puiu et al., 2016).

The Tools

The affordances of ICT and IoT have enabled a new wave of smart city sensing carried out by citizens. These initiatives can be connected to a research effort or a citizen science project (Haklay, 2013), but they can also be initiated by individuals or groups of people using some type of citizen sensing toolkit (Table 2.1) or commercially available monitoring device, examples of which could be devices put in one's own house to monitor the indoor environment[1,2] or kits that people could use outside and publish data to a bigger cloud platform[3]. These smart home devices are helpful in increasing self-knowledge or understanding, thus optimising and improving one's living space. Using these devices for making a home 'smart' can often be quite expensive, and usually such devices do not provide any opportunities for customisation. However, with the increasing popularity of the open-source hardware movement and the availability of micro-controllers such as Arduino, RasberryPi and MicroBit, people (those with the capacity or a particular skill set) can now build their own sensors to use in citizen sensing. Although people can build their own kits using off-the-shelf components[4], plug and play citizen sensing kits are already available. Table 2.1 provides an overview of popular devices for citizens carrying out environmental monitoring and reviews their openness and ability in regard to getting the larger community involved.

These sensor kits have found usage in science education, but because of their low barriers to entry, such devices are particularly good for involving people in gathering data about the

[1] https://foobot.io
[2] https://www.airthings.com
[3] https://www.netatmo.com/en-gb/weather/weatherstatio
[4] https://wonderfulengineering.com/10-best-raspberry-pi-sensor-kits/

Name	Hardware Customisability	Community	Website
AirBeam	None	Open map and data, link sharing	http://aircasting.org
SmartCitizen Kit	None	Open map, data and API	https://smartcitizen.me
AirQuality Egg	Choice of Sensors	Open map, limited data access	https://airqualityegg.com
AirPi	Full Customisation	Manual data uploads	http://airpi.es

Source: Author

Table 2.1 List of popular citizen sensing kits

environment to spark discussions around environmental issues with the community. There are multiple crowdsourced monitoring projects for citizen sensing around the world, where volunteers use low-cost sensing equipment to collect data about environmental issues and then use this data to advocate for positive change, examples of which are non-profit groups like GASP[5] and INHALE[6], which recruit people to participate in local air quality monitoring; AirCasting[7], which provides a full platform for air quality monitoring, mapping and exploring; and university initiatives such as London Air[8], which provide up-do-date information about regional air quality. Industry organisations have also come together in projects such as HAMNAIR[9] and Aclima partnering up with Google to instrument Street View cars with monitors and publish data on the Google Maps platform to provide a new source of accessible environmental data (Aclima, 2018).

The Approaches

A lot of these projects, however, come from already-framed agendas, similar to citizen science projects, which are sometimes presented to people simply as 'have a go'. Hakken (2003, p. 384) is critical of these kind of approaches, stating that real constructive learning is replaced by participation in the development of scientific knowledge. Although contributing to scientific knowledge and open resources is important for advancing human knowledge, it does not provide individual citizens with the power and choice to create knowledge around issues important to them. As a response, new types of citizen sensing and community investigation tools are surfacing that are more focused on citizen-led projects and enabling citizens to come up with their own problems, conduct investigations and create knowledge to act upon. Examples of such tools are Open Data Kit (ODK) (Hartung et al., 2010), EpiCollect (Aanensen et al., 2009), Sensr (Kim et al., 2013), wq.io (Sheppard, 2012), CitSci.org (Newman et al., 2011) and DisCoPar (Zaman and Meuter, 2015), which enable people to propose their own investigations through a reusable mobile or web application that can be reconfigured and deployed to a specific project. Garbett et al. (2016) have taken this one step further and set up a commissioning framework called App Movement, where citizens can co-create dedicated applications for their community investigations and data collection. Furthermore, research around smart cities has started to look into issues of spacial inequality and inclusiveness, with projects such as Spatial Inequality and The Smart City (Franklin, 2019) leading the way. It is important to understand and identify these

[5] https://gasp-pgh.org
[6] http://www.inhaleproject.ca
[7] http://aircasting.org
[8] http://www.londonair.org.uk
[9] http://www.hamnair.ca

risks and their impacts on citizens when implementing smart city technology in order not to create sensor deserts and instead nurture a smart city of all citizens. Clearly, the city needs to serve the needs of all its citizens and not just a select few. Collecting data about the city is just a small part of this – there is also a need for mechanisms that help incorporate the wishes and desires of citizens into the picture of how a city operates.

Indeed, it seems that the majority of drivers for carrying out city-scale monitoring come from institutional actors, and such projects are mainly run by experts who make assumptions about the population and the issues important to them (Tenney and Sieber, 2016; Townsend, 2013). Furthermore, citizens sensing projects are often disregarded by decision-makers because of a lack of trust in the data quality or accountability of the process – having the perspective that people are simply going around using 'toys' (Gabrys and Pritchard, 2018). It seems that while this is partly to do with citizens not having the resources to acquire the equipment for monitoring equivalent to that used by the city, it also relates to a lack of mechanisms that would enable citizens to access already existing resources to conduct their own investigations.

2.1.2. Data and the City

Cities around the world are working hard to implement their sensor networks not only to collect data but also to pull other sources of data (e.g. administrative data, official statistics, operational data, social media data, and location-based data) together into one place to answer questions about the city. In addition to using that data in decision-making and governance, cities are also publishing the data for others through data dashboards (Kitchin, 2014), for example, CityDashboard London[10], City of Los Angeles Hub[11], City of Sydney Dashboard[12] and Smart CEI Moncloa in Madrid[13]. It seems that a prerequisite for starting to call a city 'smart' is at least having one smart city interface (Al-Hader et al., 2009a,b). However, people have criticised this by pointing out that having access to ubiquitous technologies and fast broadband is not the only thing that will make the city smart – human factors also play a big role in making the city smart (Nam and Pardo, 2011). The city might have many infrastructures in place for collecting data about its operations and population, but it needs an engaged public to make use of that data (McMillan et al., 2016).

Whether data in the city is generated through citizen science or crowdsourcing efforts, research, or governmental initiatives, the way it is presented to people largely depends on the purpose of generation – whether it is for public awareness, policy making, research or civic engagement. In order to figure out what purpose city data serves, it needs to be examined from the perspective of its potential users. There are two distinguishable types of potential users of city data: experts and non-experts. Table 2.2 highlights different types of data users, in addition to showing the overlaps of experts and non-experts within groups.

[10] http://citydashboard.org/london
[11] http://geohub.lacity.org
[12] http://citydashboard.be.unsw.edu.au
[13] https://ceiboard.dit.upm.es/dashboard

Non-Expert	Expert
Citizen-scientists	
Policymakers	
General Public	Researchers
Community Groups	Private Companies

Source: Author

Table 2.2 Categorisation of potential city data users

To provide an insight into the ways environmental data is published, a set of popular data platforms (n=21) were reviewed[14] from the perspective of different user groups and in terms of the *feature selection* on these platforms. Table 2.2 provides an overview of how data is published based on different user groups: non-expert and expert users.

Features	
Non-Expert	**Expert**
Emphasis on health and quality of life	Emphasis on research
Colour themes for background or site	Sliders, levels, dials, larger colour ranges for detail
Mapping: local, overall, basic	Mapping: granular
Storytelling	Agnostic
Emojis	Agnostic
Simpler vocabulary	Technical, scientific vocabulary
Popular icons	Scientific icons
Fewer numbers	More numbers
Summaries or aggregates	Data layers
Pollutants shown as aggregates or not specifically named	Types of pollutants
Charts	
Scale: local and national	
Search	

Source: Author (adapted from The West Oakland Air Quality team's research)

Table 2.3 List of features on environmental data platforms

Looking at these features helps to understand some of the motivations for people to engage with city data. With non-expert systems, there is a clear emphasis on features that put data in a form that gives people immediate benefits (e.g. providing daily levels of pollutants), linking it to health and quality of life, in addition to importance also being given to personal place awareness through making links with places of importance (e.g. schools, nurseries, and shopping areas) and activities that people engage in there (e.g. storytelling). Non-expert systems also try to give people already processed data (e.g. air quality indexes and summaries and readings represented through emojis and popular icons) rather than fine-grain readings from the sensors.

There are vast quantities of data being generated from sensing the city, and this data is increasingly becoming available to people through different platforms. However, the question of what citizens can do with those datasets remains. Looking back to the ladder of participation of Arnstein (1969) (Chapter 1 in Figure 1.1a), it seems that most of the current ways of projecting city data to non-expert citizens are still at the *tokenism* phase. It is becoming more evident that issue-led approaches are having a greater impact on the lives of citizens than those implemented through organisational strategies. Examples of such projects are the Bristol Approach, initiated by Balestrini et al. (2017), which takes a people- and issue-led approach to tackle community

[14] Initial research was conducted by The West Oakland Air Quality team and was expanded on by the author of this thesis.

issues through capacity building and usage of sensor technology, and Public Lab[15], which brings together community groups, researchers and technologist to develop low-cost solutions and tools for exploring local issues. This is also evident in Citizen Science, where platforms like SciStarter[16] enable citizen scientist to form community networks around issues important to them; in this sense, instead of just volunteering their time or acting as 'citizen sensors' (Goodchild, 2007), they can engage with other citizen scientists and scientists working on different projects. To help citizens move up on the participation ladder, research needs to look into ways that data can be used by citizens to improve their personal and community lives, a good starting point for which may be understanding and helping communities voice the issues impacting on their neighbourhoods and examining ways that data and technology could act as resources for exploring and tackling those issues at the hyper-local scale.

2.1.3. Data for Community Problem-Solving Activities

It is becoming apparent that an important prerequisite for using data by communities is to identify the needs of the community through bringing common concerns to the surface (Marres, 2007). The literature review in Chapter 1 Section 1.5.1 looked at different ways that groups, CoPs (Lave and Wenger, 1991) and *publics* (Dewey, 1954) form and come together around issues to achieve common goals. Furthermore, it was pointed out that a mix of different groups are often referred to as *community networks*, which support interactions between people in different groups, help articulate shared concerns and values, facilitate information gathering and build shared knowledge (Carroll and Rosson, 2003).

Although community networks move away from organisational ties and issues take precedence, organisations or community-based institutions (Carroll and Rosson, 2003) often organically form around issues of importance. This is particularly prevalent with regard to environmental justice, where groups of individuals (from different professional and non-professional backgrounds) come together to fight issues linked to climate change and sustainability, e.g. an international NPO such as Friends of the Earth[17], who campaign for climate change, nature conservation and environmental protection, or more localised citizens' groups like Group Against Smog (GAS) in Pennsylvania, US and SPACE for Gosforth in Newcastle, UK, who are campaigning for healthier, safer and sustainable environments. These groups are examples of potential active users of city data and are often already actively leveraging existing data to raise awareness for community problem-solving activities.

Nowadays, the wider availability of the internet and ICT means that community networks are able to form quicker and mobilise better. Digital tools and network capabilities are enabling people to participate in community life even when they cannot attend face-to-face meetings. The internet has provided people with mediums for communications and for new networks to surface, e.g. posting information to public websites and blogs, sharing content on social media platforms like Twitter and engaging people in discussions around local issues on FB groups.

[15] https://publiclab.org
[16] https://scistarter.com
[17] https://friendsoftheearth.uk/

In addition to the popularity of social media platforms and online communities (Afzalan and Evans-Cowley, 2015; Crivellaro et al., 2014), the formation of community networks is also supported by increasing the push to create digital cities, digital twins (Bolton et al., 2018) and digital communities (Carroll et al., 2018). Carroll and Rosson (2003) make a distinction between community networks and network communities, in that the latter is limited only to having interactions on the internet. However, they acknowledge that the availability of different technologies (i.e. community network resources) is changing the way community networks and *proximate communities*, which associate with common activities and places in the community, operate. Furthermore, with the increase of civic-minded (Olivier and Wright, 2015; Vlachokyriakos et al., 2016) approaches to service provisioning, more resources are becoming available for communities to conduct and help engage people in inquiry-driven investigations, example of which are online-resources such as Beautiful Rising[18], which provides resources for activist groups to voice their concerns, and Digital Civics Toolkit[19], which aims to help educators engage youth in exploring civic issues and the capabilities of digital tools for political participation. Although ICT provides more ways to increase communication and exploration of common issues, it can also inhibit participation by creating network communities limited to people who have access to these online digital services, resulting in a bigger *'digital divide'* (Gurstein, 2003). Additionally, there is often more tangible value generated from in situ and informal social encounters where information exchange in happening. Carroll and Rosson (2003) point out that the social context helps to concentrate focus and make the activities more goal-oriented. People are more likely to commit to responsibilities and make something of vital importance happen because success and failure in community-led projects largely depend on individual initiative-taking (Carroll and Rosson, 2007). However, how things are actually done with regard to the responsibilities to make something happen is face-to-face and in small core groups of communities that have combined people with different skills.

Multiple concerns have surfaced around the way city data is currently generated, made available and used by communities to raise attention to and address local issues. In this sense, community groups are often limited in terms of the tools and resources they can access and use to participate in decision-making processes, and they are often dependent on professional services and solutions provided by governments and experts. This means that somebody else might be controlling the knowledge-making processes around these issues. On the other hand, while there are already examples of groups actively using the data, for example environmental justice, there may be more groups with different agendas and skills who are yet to benefit from the city data. There is a need for mechanisms that support different community groups and individuals coming together to form networks, consisting of multiple stakeholders at different levels of expertise and also power structures. Furthermore, there is a need to create a synergy of online-offline infrastructures and participation methods in order to address inequality issues around participation and local decision-making using data.

[18] https://beautifulrising.org/
[19] https://www.digitalcivicstoolkit.org

2.2. Study Context

The following sections will provide an overview of the context within which this research was conducted, i.e. the particularities of the context in terms of the challenges of and opportunities for the research.

2.2.1. Effects of Urbanisation

Every citizen needs their space to live, work, play and get around for their everyday activities. However, with more than half the world's population now living in cities, this is becoming increasingly more difficult. It is projected that the rate of urbanisation will increase to the point that about 68% (Ritchie and Roser, 2018) of the world's population will live in densely populated urban environments by the year 2050. As space in the centre becomes increasingly scarce, cities become dispersed and more people inhabit the outskirts of urban areas. Although people live on the edges of towns and cities, they often work or have to travel because of work to the centres. According to the UK government's Department of Transport, the total number of vehicle registrations in the UK has increased every year since the Second World War, with an average growth of 630 thousand per year[20]. This increased volume of traffic and people commuting to city centres consequently creates problems of congestion and air quality, which means that cities are faced with a multifaceted problem (i.e. 'wicked problem') requiring collaboration from policymakers, city architects, transportation and city planners, public health officials, and also citizens.

In July 2017, the Department for Environment, Food and Rural Affairs (DEFRA) and the Department for Transport announced a plan to tackle roadside nitrogen dioxide concentrations in the UK, a direct response to the illegal levels of air pollutants in 29 of the local authorities in the UK[21], among which were Gateshead Metropolitan Borough, North Tyneside and Newcastle in the north east of England, the focus areas of this case study. The local authorities have had to come up with a combined task force to develop proposals and solutions to tackle these illegal levels of air pollutants. Newcastle and Gateshead, separated only by the River Tyne and connected via multiple bridges, are both old coal mining areas, and with North Tyneside and its connection to the North Sea, the area also has strong ties to shipbuilding. Although both of these industries are not active anymore, they are still part of people's historical identities, places and spaces. Newcastle is the biggest city in the north east of England, with an area of 44 square miles and a population of 300,196, averaging an age of 37 years, with 58.3 per cent of households owning or having access to a car (according to the 2011 Census). As new developments are taking place and more people are moving into the area, the roads need to cope with increased amounts of traffic.

There are already two 'Air Quality Management Zones' in Newcastle, which have been under constant monitoring since 1995 and the release of the Environmental Act[22]. Similar to other

[20] https://www.gov.uk/government/statistics/vehicle-licensing-statistics-2018
[21] https://www.gov.uk/government/publications/air-quality-plan-for-nitrogen-dioxide-no2-in-uk-2017
[22] http://www.legislation.gov.uk/ukpga/1995/25/contents

cities and towns in the UK, the area is undergoing an urban shift to create an environment for modal shift and multi-modal transportation. Alongside attempts to improve public transportation networks, there is a particular focus on improving cycling infrastructure and encouraging cycling as a means of transport in the area. To date, the local council (i.e. Newcastle Council) has taken steps to promote alternative means of travel and has invested in infrastructure for active travel through the UK Government's Cycling and Walking Investment Strategy (CWIS)[23]. Despite these measures, the problems still persist, and there have been discussions around implementing more drastic measures, such as closing roads and implementing Clean Air Zones. Even if these kinds of measures are put in place, with the growing urbanisation trend, the estimation for zero emissions journeys by 2050 seem unrealistic. It became clear quite early on that the only way to tackle issues surrounding air quality was to tackle the sources, which means a wider engagement with citizens to increase public awareness around the issue and working together to come up with viable solutions for getting around the city for everyone.

2.2.2. The Air We Breathe

According to the 2016 report 'Every breath we take - The lifelong impact of air pollution' published by the Royal College of Physicians and the Royal College of Paediatrics and Child Health[24], outdoor pollution is linked to around 40,000 deaths each year in the United Kingdom. In addition to explaining the repercussions of bad air quality on people's health, the report also provides recommendations for action and research for national and local government, business and industry, schools and the National Health Service (NHS), as well as communities and individual citizens. Among the recommendations for tackling air pollution as an individual were the following: *'be aware of the air quality where you live and harness technology to stay informed and monitor air pollution effectively'*. Having access to this kind of information locally could not only help raise awareness around the issue, but could also help people make better decisions about their travel choices. However, not everyone has the means to monitor air pollution 'effectively', even less so in their local neighbourhoods.

Newcastle hosts one of the six national urban observatories in the United Kingdom[25]. Each city takes a slightly different viewpoint, but all are focused on environmental monitoring and data collection in some way or another. Newcastle's UO[26] aims to provide a resource for understanding the impact of the city and the complex interactions within it through sensing and data collection. The UO operates as openly and transparently as possible to provide the best public access to environmental data from the city and to do so in such a way that promotes information being used by a wide range of stakeholders – environmental scientists, city officials, universities and businesses – but also regular citizens to promote initiatives that involve community-led investigations. The UO's data portal[27] has had over five million downloads

[23] https://www.gov.uk/government/publications/cycling-and-walking-investment-strategy
[24] https://www.rcplondon.ac.uk/projects/outputs/every-breath-we-take-lifelong-impact-air-pollution
[25] https://gtr.ukri.org/projects?ref=EP/R013411/1
[26] https://urbanobservatory.ac.uk
[27] https://newcastle.urbanobservatory.ac.uk

from when it was initially set up. Despite the best efforts of the project to make data accessible and usable for everyone, the data were mainly used by individuals or groups of expert users, thus not living up to the observatory's slogan – 'Data for Everyone'. Having the UO already established served as an excellent opportunity to investigate ways we could open such resources up for citizens and community use, helping to understand whether the public actually cares about environmental data or if there is a need for a different approach towards citizen engagement.

In order to obtain a better understanding of the challenges in this context and use the available opportunities, a more participatory approach was taken to explore ideas of active citizen participation in the smart city with the data produced by the UO. The aim of the research was to come up with mechanisms to democratise data production and use in the smart city by enabling citizens to collect and explore data relevant to the issues important to them at the hyper-local scale.

2.2.3. Study Design

The motivations outlined in Section 2.1 within the context of local issues resulting from increasing urbanisation that affect peoples' health and how they get around in the city (Section 2.2.1 and 2.2.2), and the potential availability of resources that could be used to shed light on such issues, opened up an opportunity for research to understand how data production could be democratised and used to better engage communities affected by these issues.

The case study followed the overall methodology of the research conducted in this thesis (see Section 2.2), which takes a PADRE approach to understand the communities' needs through engagement activities and user-centred design processes so as to develop digital tools and processes for supporting people-led investigations and advocacy efforts. The initial interest of the community was explored in a two-and-a-half hour workshop (Section 2.3.1), where two activities were run in order to explore advocacy groups' current practices and communities' data needs and technologies that would support the collection and effective use of that data for civic advocacy and action. The workshop was recorded and transcribed for further analysis. Transcriptions and written material from the workshop were analysed and coded, drawing on Braun and Clarke (2006), in order to identify themes around concerns and flesh out initial design ideas for the usages of data by communities.

Configuring Participants

One of the areas of Newcastle already had an established community action group who were advocating for changes to city infrastructure so as to promote walking and cycling. It consisted of people of different demographics, skills and backgrounds. The activities of the group included lobbying and raising awareness around the effects of air quality and the infrastructure changes being implemented in the city. The main communication of ideas from the group was carried out through public meetings and by publishing articles on a blog to members of the general public. They were also actively using the data produced by some of the sensors deployed by

the UO at Newcastle University. From initial meetings with the group initiators, a two-year ethnographic engagement with walking and cycling advocates in the area followed, which helped to understand the exciting practices of the groups and how they were leveraging the data and technology in their advocacy work. The core members of the action group became the key people for making connections with the wider community. In addition to having meetings with the group organisers, the author also attended the group's annual meetings, where they planned their focus and activities for the year. Reflecting on the ethnographic work and conversations with the members of the group, there seemed to be a great deal of interest in using data as a resource for advocacy and action. Furthermore, there were public funds for infrastructure changes allocated for public proposals, and people were keen to gather evidence about air quality and traffic movements in their local areas to use as a basis for backing up their proposed changes.

From these reflections, an initial workshop (Section 2.3.1) was designed to explore further ways to promote the usage of data for raising awareness about the issues in local neighbourhoods and for civic advocacy and action. The initial workshop participants were recruited by means of snowball sampling through existing advocacy group members. Overall, 22 participants took part in the workshop activities. The workshop not only provided a deeper insight into people's perceptions of data and its uses for civic advocacy but also helped broaden participation and make connections with a group of interested residents in another part of the city (two miles from Newcastle city centre) that had a scarcity of hyper-local environmental data. The area itself has historically been heavily populated by students, but is now shifting towards being more residential, with a bigger proportion of young professionals and families settling down in the neighbourhood.

Participant	Roles
P1	local charity trustee, retired council worker (city planner) from a different council
P2	professional working in ICT
P3	father, manager working in academia
P4	father, general practitioner working at the NHS
P5	father, lawyer
P6	mother, professional working at the NHS
P7	father, public relations officer
P8	consultant

Source: Author

Table 2.4 People involved in the co-design process

The main group of people who helped guide the design for setting up the sensor commissioning toolkit consisted of eight local residents (Table 2.4). Each participant had a degree of knowledge around modern ICT technologies and social computing. All the participants were professionals or ex-professionals working across different disciplines: academia, planning, law, healthcare and ICT. All participants had previously taken part, or were involved in, public consultations (i.e. new cycle schemes and infrastructure changes) in the area and took part in advocacy activities in some respect. These efforts manifested in taking part of group bike rides, organising community meetings, raising awareness in the community (e.g. by engaging local schools, churches and talking at community events) and having an active social media presence. Four participants were involved in forums like the Healthy Streets board, which discusses transport,

health and air quality, and the Cycling Stakeholder Forum, which brings together people who have an interest in cycling in the city. One participant was also involved with a local charity that promotes cycling and up-cycling old bikes for donation. All participants were active advocates in their respective fields, for example, publishing videos from the perspective of a public health professional raising awareness about increasing health issues caused by a lack of activity and air pollution, or analysing and publishing data about car traffic to support the promotion of active travel in planning polices.

Clearly, all the participants in the group were part of a set of diverse community networks and were already involved in official and unofficial or *insider* and *outsider* activities (Asad and Le Dantec, 2017) around the issues – insider activities being those connected to institutional strategies and outsider activities being more indirect, often involving community organising and using ad-hock digital technologies to participate in civic advocacy. All the participants were regular cyclist themselves and used their bikes for communing, leisure and, in the case of one participant, also for sport. Additionally, three participants, who were parents with young families, were encouraging active lifestyles for their children by doing school runs on bikes, scooters or on foot.

Design Process

The case study consisted of two parts: (1) the design, development, deployment and analysis of the community commissioning toolkit usage; and (2) the evaluation of the use and value of the commissioned sensor data by the community (Chapter 3). The first part of the case study, described in this chapter, focused on setting up the sensor commissioning toolkit through working together with a group of residents who were advocating for active travel and infrastructure changes in their local area and the city. This was conducted through an eight-month iterative design process, which consisted of: building relationships and understanding the community; gathering design requirements and responding with technologies that could help people better generate data; and assessing the use of the tools and processes by the community. The process was recorded through ethnographic engagement with an advocacy group, involving regular meetings, emails and discussions through an FB group; observations at group-organised public events and meetings around the issues of concern; and collected statistics about the use of different digital tools of the toolkit.

This part of the case study analysis mainly focused on the small sample of participants in a particular advocacy group, which limits the findings being generalised to the whole population. Additionally, the particular group involved in the design of the toolkit consisted of people who were professionals in their respective areas and had a very good knowledge of the issues they were engaged with. However, outside this group, more people have signed up through the infrastructures that have been set up as a result of this research. The success of the project and the interest from the community is also illustrated by the fact that additional organised advocacy groups have been created based on the model of this one, particularly by using data and digital tools that have been designed through the project to raise awareness and push for positive change

in the community. Although the issue of air quality took precedent in the case study and was definitely one of the key drivers for the local authority and the citizens involved, it is not the sole focus of the commissioning toolkit or the research around active citizen participation and data production in the smart city.

2.3. Designing Community Resources

The following section will describe the process of the gathering requirements for setting up the community sensor commissioning toolkit. Five month's of engagement, involving a workshop, community meetings, focus groups and ethnographic work, resulted in the first prototype of the SMS. The work carried out will be outlined in chronological order in the following subsections to better document the design process.

2.3.1. Design Workshop

A community workshop called 'Beyond Pins on a Map' was run by the research team[28] in spring 2017, focusing on how people could collect meaningful data about their everyday journeys and take ownership and use that data in advocating for positive change in local communities. The workshop engaged with 22 local residents from different parts of Newcastle who took part in two activities to explore the opportunities and challenges of using data for positive change in the community. The workshop already had a previously identified focus on air quality and pedestrian and cycling safety (Sections 2.2.1, 2.2.2, 2.2.3)

The first activity looked at some challenging scenarios related to collecting data about air pollution and safety, which included speculating about the use of low-cost citizen sensing (Section 2.1.1) equipment and bespoke solutions (Maskell et al., 2018) to gather data from people's everyday commutes, in addition to thinking about the city datasets already published by the UO. In groups, people explored issues such as privacy, reliability of the data and how other people not involved in the study might trust the collected data. There were lively discussions around the tables, which were captured and transcribed for further analysis. The second task prompted participants to think about how that data could be used to advocate for better spaces for people to spend time in and move through. To inspire them, people looked at the issue in hand in groups, using examples ranging from art installations to activist campaigns. The groups then brainstormed ideas about what might work in their neighbourhood with the resources they could have and fed back a shortlist of their chosen tactics. Building on Beautiful Rising[29], material was made for the workshop to guide the design process. All the materials produced by the participants were archived, accompanied by audio recordings and photographs of the workshop.

Workshop findings revealed that people were interested in environmental data, but often from places that it was not available and from where there were no known plans from city officials or decision-makers to collect it. Further analysis also showed that although people felt a

[28] The research team consisted of Thomas Maskell and the author. Through collaborative effort, workshop activities were created, conducted and analysed to help understand the community data needs.

[29] https://beautifulrising.org

little uneasy about the accuracy of the data (collected from the low-cost citizen sensing devices described in Section 2.1.1), they thought that it was more about how things are framed, and that having something out there for people to discuss was better than not having anything. The figures from the published report about the effects of air pollution and its consequences on people's health (Section 2.2.2) can feel remote to citizens just going about their everyday activities in the city. However, everyone in the community is affected by it in some way or another. In order to tackle the issue, it had to be made visible within the community by the people living there. There was consensus among the participants that the issue needed local people to come together not only to collect data but also to make sense of it and to decide how it might be used in order to take action and make local streets more liveable places.

After the initial design workshop, the author was approached by residents from an area a couple of miles from the city centre (Section 2.2.3). There had been multiple articles in the press about the air quality in cities around the world, which had sparked discussions among residents in different community meetings. One of the more proactive resident had also acquired a diffusion tube[30] for monitoring air quality and had fitted it near their house. This had provided some concerning results, raising questions and a desire to find out more. Some of the residents from that group also attended the initial design workshop. In contrast to the active lobbying group mentioned previously, who were using data from the UO, this group did not have sensors in their local area, and there was a lack of information to share with the community. Furthermore, according to the group, the DEFRA air quality model[31] showed that the pollutant levels were in accordance with the legal limits, and the area was not part of the two Air Quality Management Zones in the city, which meant that there was no known plans from the government to start monitoring the area. However, people who used the spaces daily pointed out that they were seeing issues related to the infrastructure changes and increased volumes of traffic. People were also looking at the increasingly high readings of pollutants in other areas of the city that had monitors, which led to a small group of residents getting together to examine the topic in more detail. Residents expressed particular concerns about the effects of harmful pollutants on children and their developing lungs. This, together with increased levels of through traffic trying to get to the city centre, made residents worried about the lack of space for active travel and other infrastructure funding not going towards making their local streets healthier and safer for everyone to get around.

2.3.2. Surfacing and Mapping Community Concerns

The findings from the initial design workshop provided indicative evidence of community interest and also of a community keen on getting involved with the co-design process. Furthermore, there were available funds from the UO to acquire environmental monitoring equipment; however, up to this point, there was no commissioning toolkit in place to manage this shared resource, nor was there a clear view of the places and the sort of environmental indicators people were

[30] https://laqm.defra.gov.uk/diffusion-tubes/diffusion-tubes.html
[31] https://uk-air.defra.gov.uk/research/air-quality-modelling

interested in measuring. Together with the research team[32] from the UO, we came up with a set of potential monitors that could be acquired or repurposed, provided there was interest from the people. These included scientific-grade air quality noise and traffic monitors that could be located on lampposts by UO engineers and hand-held environmental monitors that people could use themselves on commutes.

In order to explore these avenues and get an idea of what people were particularly interested in measuring, a meeting was set up with the group of residents (Table 2.4). At the meeting, the group provided some background to their area and talked about the areas of concern and their desire for data to help residents make better decisions and advocate for change. To make the conversation more related to the community, a map of the area in question was printed out and used. After the group learned about the environmental indicators that could potentially be monitored using a selection of scientific-grade sensors, they collaborated around the map (see Figure 2.1) to point out the areas of concern and marked the potential locations for different environmental monitors.

Having no data available from that particular area, the residents had to rely on their perceptions and local knowledge when deciding what the most important places were that needed investigating through environmental sensing. Through discussion, people had to agree on where to focus their efforts and use the limited resources to gather environmental data about the area. As five of the participants were parents whose children went to the local schools, the focus was on areas where young people may be exposed to the harmful effects of roadside traffic pollution. This focus was a recurring theme with almost all community groups engaged throughout the project. In this sense, advocates were trying to voice the concern of those who may not have been able to do so themselves (e.g. young children, the elderly, and people with reduced mobility) but were often the ones affected by the issue the most. Although, this type of engagement activity worked well in a focus group setting, it needed everyone to come together in a physical space and needed support from the researcher to facilitate the mapping. Reflecting on the engagements, it seemed that in order to scale this process and enable broader participation from the community, it needed to take the form of an online participation system.

2.3.3. Planning Community Investigation

The first set of meetings and engagements with the community group helped identify places in the neighbourhood and the issues that people wanted to explore in those places. The next step for the community group was to come up with an action plan. After the initial meetings and conversations with the community group, they came up with a proposal called *'Air Pollution in Heaton - The desire for data to help residents make decisions'*, which consisted of: *a*) Background; *b*) Areas of concern; *c*) A desire for data; *d*) Action plan; *e*) Areas for fixed monitors; and *f*) Areas for hand-held monitoring.

[32]The research team of the UO consisted of Prof. Philip James, Dr Jennine Jonczyk, Neil Harris, Dr Luke Smith, David Pearson and later Richard Turland.

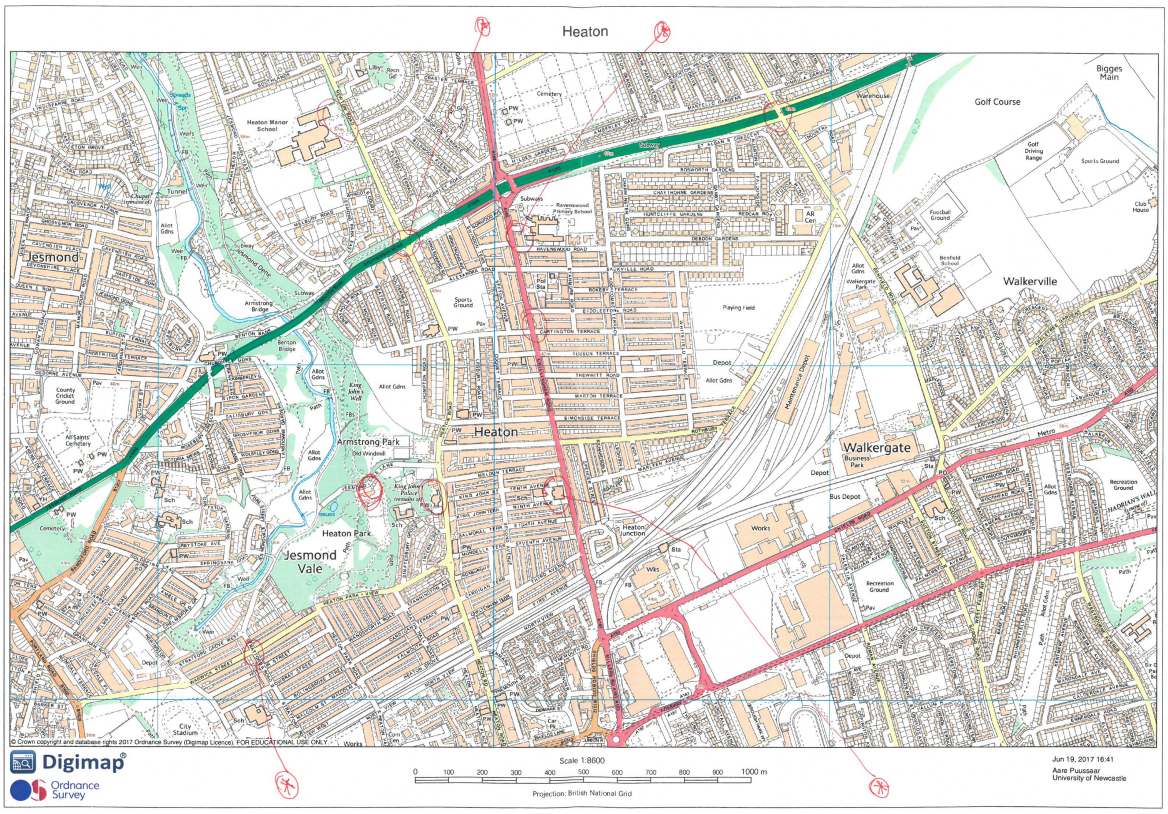


Figure 2.1 Initial mapping of areas of concern

People were driven by the fact that there was no data available from the area to make an informed statement about the air quality. The first aim for people was to establish a baseline on the level of pollution in the area. This document also set out a clear action plan for doing so through collecting the data, but also for using the data for advocacy and activities for local benefit. The following is a statement from the air pollution bid submitted by the residents.

> *The hope is that the data collected from these monitors will fill in the current gaps in air quality data for the area and provide a good baseline for decisions of further research around measures designed to improve air quality.* (SPACE for Heaton, 2017)

At this point, there was no fixed process in place that would set the criteria of how resources would be allocated to the people, nor any mechanism of transparency. The back and forth conversations between the the author and the community group that guided the creation of the proposal were not documented and stayed within the close group. In addition to the issue of transparency, there was also no opportunity for others to learn from these processes. Although the exercise of creating an action plan (i.e. proposal) was good for setting the agenda for the community-led investigation, the knowledge shared and obtained in the process stayed within the specific community group. This knowledge, however, could potentially be leveraged by another community group, and lessons learned from it would help improve the process of commissioning. Reflecting on the process of the community developing a data collection plan, there seemed to

be a need to design a mechanism that encouraged wider participation of the community through active discussions and feedback loops.

2.3.4. Initial Community Investigation

In addition to enabling people to commission environmental monitors from the UO and get them deployed in their community, the people also wanted to be more actively involved in the investigation. This meant that there was a need for equipment that people themselves could use on their commutes to work or when taking their children to school. By doing so, people could not only get a personal perspective and idea of individual exposure, which links to health impacts, but also cover more ground in the neighbourhood and provide a good basis for decisions regarding further investigation around air quality in the area. Although these sensors would provide a snapshot of data, they would still provide a basis for discussions and decision-making for the residents. In a way, this part of the process was less linked with advocacy and raising awareness and more about self-knowledge and taking ownership of the investigation surrounding community issues. The main concerns pointed out by the members of the groups were linked to the increased volumes of traffic having an impact on people's safety and air quality. Through the UO project at Newcastle University, a high precision, hand-held particulate monitor[33] was then purchased and loaned out to the community group to start investigating the air quality in the area. At this point, all the communication with the advocacy group was carried out through face-to-face meetings or email correspondence, which also included the scheduling of the sensor handover and return to the UO. This meant that there was always a reliance on the researcher to make connections with community groups and coordinate the sensor loans, which brought to attention another design challenge and requirement for the technology. This stage of the investigation was highly dependent on the contributions of people. After they got the sensor and a brief introduction about how to use it, they needed to figure out the monitoring schedule (Section 2.3.3), time allocation and logistics around exchanging the monitor within the group to meet the monitoring schedule. People were given an overview of data collection methods, sampling, variability of different indicators and robustness of the data, but the actual monitoring plan was intentionally left for them to set out. An additional reason for promoting community-led investigation was the emphasis on peopleÕs experiences of living and moving around the area, which provided local tacit knowledge not obtained by the research team. These experiences were also important for understanding the issues and different perspectives around them. In order to share these perspectives, there was a need for a mechanism that would be able to capture them, in addition to the sensor readings from the hand-held sensor.

This case study was conducted through a user-centred design process, where the designed technologies were direct responses to the requirements that surfaced from engaging with the people. A number of requirements for the design of the technology surfaced from the initial engagement with the community group around local issues. However, because of the participatory

[33] https://www.palas.de/en/product/fidasfrog

nature of the research, the main focus was not only on acquiring and providing people artefacts (i.e. the environmental monitors) that would give them capabilities to carry out community-led investigations, but on the whole design process and the *coupled technologies*, which were considered as a way of creating knowledge, expanding community networks and helping to build the capacity of the community. This approach and the strategies linked to it echoes those of scholars working on civic engagement (Asad and Le Dantec, 2017) and participatory design (Binder et al., 2011; Vines et al., 2013).

2.4. The SenseMyStreet Toolkit

From the beginning of the research, the focus and emphasis was on supporting initiatives where citizens are involved in problem definition, data collection and analysis, thus taking a *extreme citizen science* (Haklay, 2013) approach and helping people to achieve more control (Arnstein, 1969). The findings from the initial engagements – conversations with the community members, observations and the design workshop – and the process of designing the community resources (Section 2.3) all fed into the first prototype of the SMS toolkit. When designing it, the aim was not to develop stacks more digital technologies because there was already an abundance of smart city technologies available, but also to look at how existing technologies could be reconfigured to be put to use by citizens for community problem-solving activities. Through the development and use of participatory GIS and mapping, leveraging environmental sensing equipment and models of community commissioning Garbett (2017) and HCI for geospatial technologies, the first prototype of the sensor commissioning toolkit was built. The following sections describe the process (Section 2.4.1) and the implementation of the digital technologies (Section 2.4.2) linked to the first prototype. While the toolkit continues to be used by communities and is constantly involving, for this case study analysis, the focus will be on the first prototype.

2.4.1. The Process

This section will provide an overview of the process for the sensor commissioning toolkit that was derived from the initial engagement work (Section 2.3), in addition to explaining how certain design decisions were made at each step of the process. Figure 2.2 shows the overall makeup of the first prototype, which consisted of four steps, each of which had its associated processes and technologies: (1) Identify the issues; (2) Plan and propose; (3) Get the facts; and (4) Data for everyone. Each step is broken down into smaller activities to provide better details of the whole process of issue identification, citizen data collection and sensor commissioning.

Identify The Issues: Negotiating with the community

Similar to the initial engagements (Section 2.3), the first step of the toolkit was to aid people in identifying the issues that they were concerned with and wanted to explore in their neighbourhood. This may have been something an individual was concerned in particularly or something that a group of residents might advocate for; nevertheless, these issues were almost always connected

Figure 2.2 Illustration of the four steps of the toolkit process

to places in the community and something that could potentially nourish the creation of a community network. The back and forth communication between the group and researcher, and the planning and organising of the sensor handover, was replaced with automated systems that people could access online. Figure 2.3 illustrates the steps relating to identifying and exploring issues.

Enabling the community to use an online map to identify issues expanded the participation to people that may not have been present at the community meetings and also to people who may not even have been aware of these issues in the neighbourhood. They could then engage with the other residents and start a discussion around the issues of concern and plan further investigations with the community. To extend the exploration of the flagged issues, the toolkit enabled people to investigate them further through conducting environmental monitoring with hand-held monitors. People could identify areas of concern and then apply online to borrow hand-held monitors from the UO. A set of hand-held monitors was acquired based on the types of issues people flagged

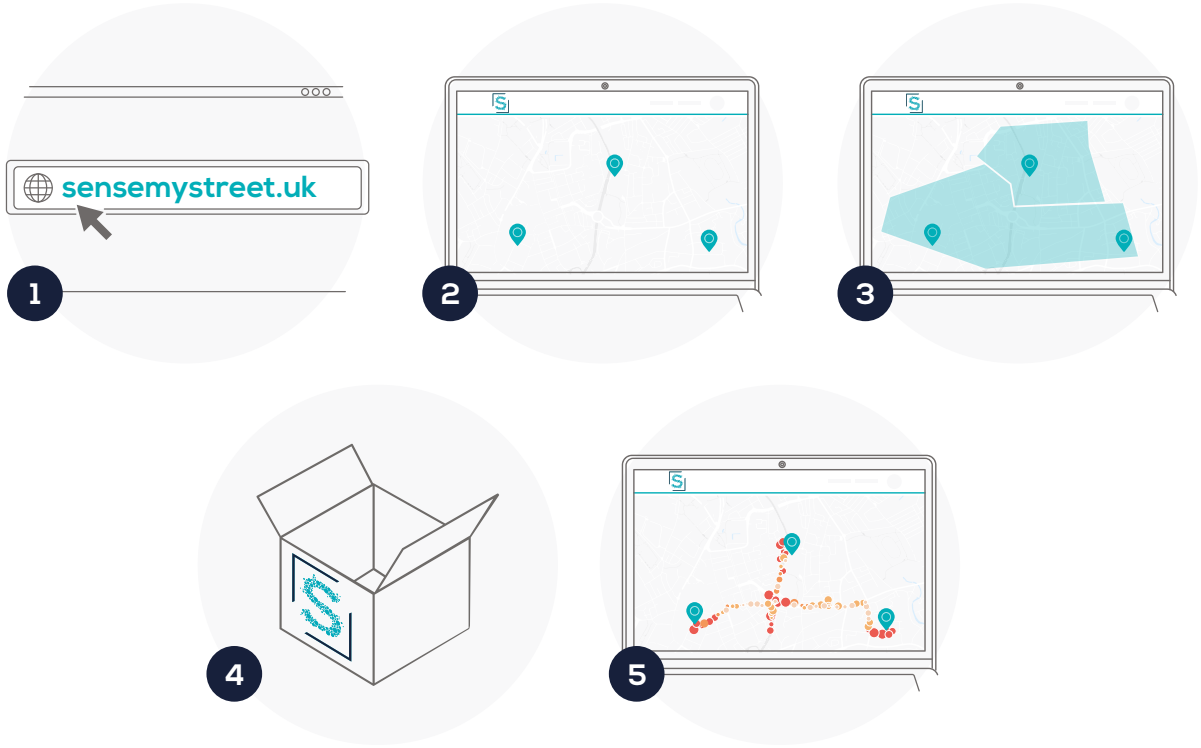


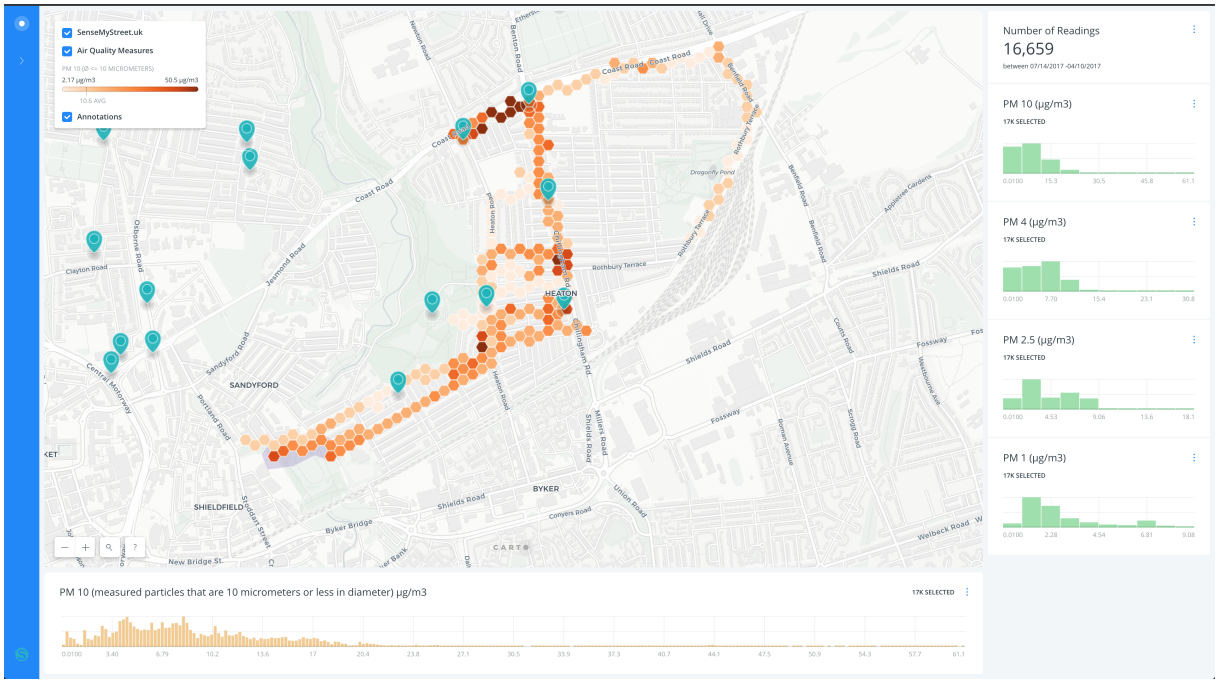
Figure 2.3 Process of identifying the issues: (1) visit the project website and sign up; (2) flag the issues; (3) identify neighbourhood areas for monitoring and apply for a monitor; (4) loan a hand-held monitor; and (5) conduct preliminary monitoring.

through the website. The first prototype of the toolkit had three different types of hand-held monitors: particulate monitors, noise monitors and traffic counters[34]. What distinguishes SMS from most other citizen-sensing initiatives is that the types of monitors uses in the toolkit were scientific-grade monitors, which are considered to be high precision and often cost ten times more than the usual low-cost equipment. Once resources became available, people received or picked up a hand-held monitor from the UO, which they could use for a maximum period of one month to monitor their commutes. Sensor data accompanied by individual perspectives was then uploaded and shared with the community to help them further understand the issues (Figure 2.4).

Plan And Propose: Negotiating with the UO and other stakeholders

Using hand-held monitors provided people with an opportunity to collect data from places and about issues that were not possible before. People could use that data to look at whether the issues they flagged were also showing up in the data to see if their worries around the severity of the issue in those locations were confirmed or not. The snapshot of data not only provided an overview of personal exposure on everyday commutes, but also laid a good base for planning and proposing areas that needed more focus and additional monitoring using fixed monitors. People could use this data as a discussion point for deciding where they would need to continue monitoring in order to get a better picture of the issue. The Plan and Propose step was similar to the *Planning Community Investigation* activity people engaged with in the design of the toolkit described in Section 2.3, but now it was also undertaken online (Figure 2.5).

[34] https://sensemystreet.uk/sensors



Source: https://sensemystreet.carto.com

Figure 2.4 Example map of readings from community monitoring with a hand-held sensor

Figure 2.5 Process of plan and propose: (6) visit the proposal website; (7) create and collaborate on a proposal; and (8) plan the deployments.

Get The Facts

After the community had consolidated their discussions into a proposal and it was approved by everyone who participated, they moved to the next stage of scheduling the deployment with the UO. This step of the toolkit did not introduce any additional technologies to the process but instead relied on the information and scheduling obtained from the proposals created by the community group. Once deployment was agreed, engineers at UO scheduled it in with the overall deployment plan. From this point forward, the responsibility for data collection was handed over from the community to the researchers and engineers at UO, which meant that they had to guarantee successful deployment and deliver the best possible quality of data. Initially, the maximum deployment time was one month, but because the monitors needed some time to settle, the deployment time was extended up to two months. However, the community could reapply for another deployment in the future once the resources were available again.

Data For Everyone

Once the monitors were successfully deployed, people were notified and could access the data from their commissioned monitors using UO's city data portal (Figure 2.7). To avoid creating another data silo, the data collected from the commissioned sensors was made accessible in real time on the UO city data portal – similar to other monitors deployed in the city. The monitors were recording up to nine different environmental indicators, saving a reading every minute. Hence, in a month, one monitor would have recorded *9*43,800=394,200* readings from a commissioned place in the neighbourhood. People could then use the portal to view the latest readings, graph 24 hours or 7 days of readings, and use the data download functionality or API to get data on any period of the deployment. Figure 2.6 illustrates the last steps of the process – *Get the facts* and *Data for everyone*.

Figure 2.6 Process of getting the data: (9) getting the commissioned monitor deployed by UO; (10) getting the data uploaded to the UO's city data portal; and (11) accessing the commissioned data from the portal.

Source: https://newcastle.urbanobservatory.ac.uk

Figure 2.7 View of a commissioned monitor page on Urban Observatory city data portal

2.4.2. *The Implementation*

The first prototype of the SMS toolkit consisted of four main development stacks – the SenseMyStreet platform, an automated sensor loan scheduling system, hand-held monitoring tools, and the SenseMyStreet Proposals platform Ð as well as some external services (Figure 2.8). The SenseMyStreet platform consisted of a web-based application written in PHP leveraging make.place, a reconfigurable GIS platform developed in Open Lab at Newcastle University (based on Silverstripe framework), and a RESTful JSON API for communicating with the MYSQL database. The main website had multiple reconfigurable elements and was deployed using Docker, which made the application scalable and easily deployable on virtual or physical infrastructure. Software included reconfigurable features such as: configurable geographical surveys; API to access the survey data programmatically; commenting and voting on survey responses; a full Content Management System (CMS) to input content and design the look of the website; a configurable dynamic filtering interface; and theme customisation using Docker variables. These features provided additional ways of integrating data with other systems and platforms in the SMS ecosystem (Section 2.5.5). The SenseMyStreet website provided all the information about the project, its resources (i.e. available environmental monitors), description of the sensor commissioning process (Section 2.4.1) and contact details for the research team. The website also acted as an onboarding component for the SMS toolkit, where people could create an account on the website, learn about the project and its resources, and start flagging areas of concern using the participatory mapping tool on the platform. Furthermore, platform users could explore previously identified areas and have a discussion around the issues with the community.

The automated sensor loan scheduling system consisted of monitor loan forms[35] and an automatic monitor loan scheduler[36] and deployment calendar[37] built using Google Apps Scripting. The sensor scheduling system automatically allocated resources and scheduled hand-held sensor handovers to the community, while also automatically publishing the list of available resources and the deployment calendar. This enabled people to receive notifications about when the resource would become available and when they could expect to receive a monitor to start doing their investigations. Furthermore, this provided one place for the coordination of resources that was always up to date.

The hand-held monitoring tools consisted of hand-held monitors; the SenseMyStreet mobile application[38] (based on SpokesPeople App originally from Maskell et al. (2018)) with a bluetooth Flick[39] button integration to control the functionality on the application; an automatic sensor data uploader (using Google Drive); data parser[40] written in Python (using Pandas library) for parsing hand-held sensor data and merging with GPS readings; and Carto[41] maps for visualising

[35] https://goo.gl/forms/7A9eVtB4KwyuM0AA2
[36] https://goo.gl/14wrgM
[37] https://calendar.google.com/calendar/b/1?cid=c2Vuc2VteXN0cmVldEBnbWFpbC5jb20
[38] https://play.google.com/store/apps/details?id=uk.sensemystreet
[39] https://flic.io
[40] https://github.com/aarepuu/fidasfrogparser
[41] https://sensemystreet.carto.com

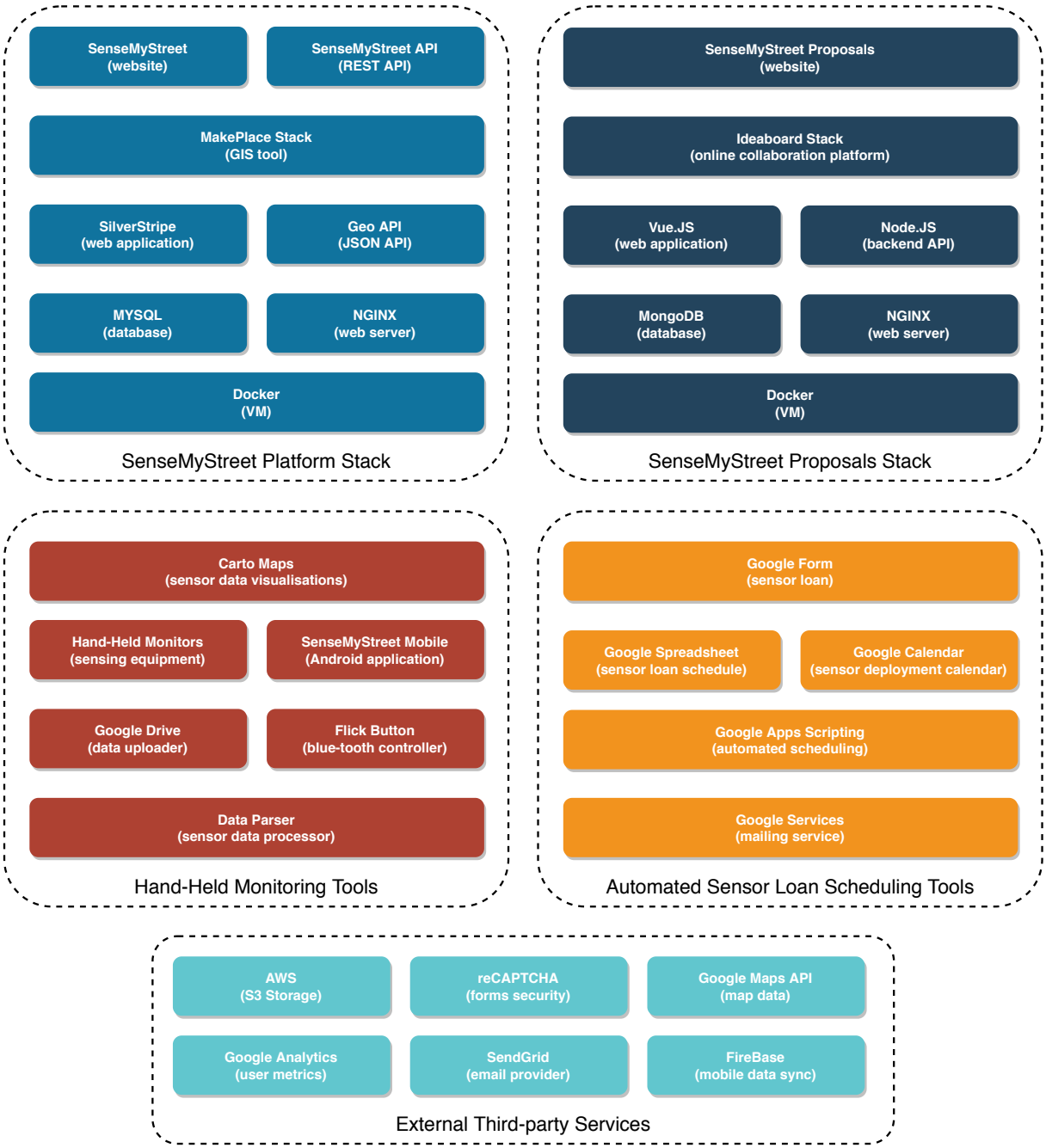


Figure 2.8 SenseMyStreet toolkit architecture

data from hand-held monitoring devices. Hand-held monitors were loaned out with easy-to-use instructions (Appendix A), which were developed through the process of community investigation (Section 2.3.4). The additional mobile application enabled people to record their monitoring journeys, insert textual comments and recordings of audio snippets while they were monitoring and reflect upon their journey once finished. The system had a mechanism to automatically send hand-held sensor readings, GPS and personal reflections to the cloud from the sensor commutes once the devices were connected to the internet. Furthermore, having data available online would enable the building of a pipeline, where the readings would be automatically parsed, processed and visualised.

The SenseMyStreet Proposals platform was a web-based Node.JS application leveraging Ideaboard[42], a reconfigurable online collaboration and feedback platform developed in Open Lab at Newcastle University (based on Vue.JS framework), and RESTful JSON API for communicating with the MongoDB database. People could use the platform to create an account and start collaborating around proposals for fixed monitors. The system enabled people to have discussions around issues and ideas using modern social media functionalities such as online chatting and discussion boards, media sharing, polling, video recording and calling, collaborative whiteboards and online forums. Furthermore, the platform provided sharing functionalities for other social media platform feeds (e.g. Twitter, FB, and LinkedIn) to get the wider community involved with the proposals.

Additionally, there were several external services used to add functionality or support the SMS toolkit: Amazon Web Services (AWS) for deploying two web-based platforms and guaranteeing the data backups (using S3 storage) on the systems; the SendGrid email delivery service for sending out email notification for subscribed users on different SMS systems; Google Maps API for mapping functionalities on the SenseMyStreet web platform's GIS system; Google Analytics to provide metrics on user visits; Google reCAPTCHA for protecting against malicious automated software activity; and Firebase for syncing data between the mobile application and the SenseMyStreet web platform's API.

2.5. Interactions with the SenseMyStreet Toolkit

This section will describe the different interactions that people could have at different stages of the issue exploration and sensor commissioning process. Each stage is described by means of a user story (i.e. a system walkthrough) and data flow diagrams.

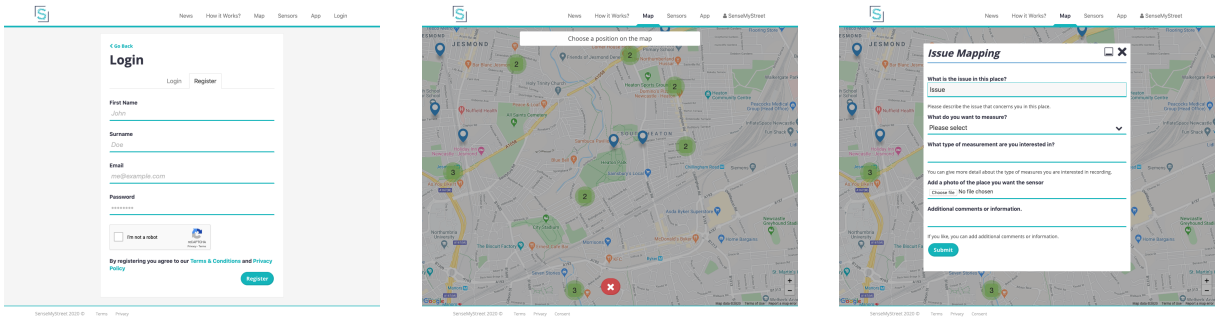
2.5.1. Onboarding and Issue Exploration

Initial engagement with the SMS began with a community member going on the SenseMyStreet website[43] to identify and flag an issue on the map. After creating an account on the website (Figure 2.9a), people could use the map to flag areas of concern and submit issues (Figure 2.9b and 2.9c) explore already identified areas and issues (Figures 2.10a and 2.10b) and have discussions around the issues with the community (Figure 2.10c). The issue submission form also enabled people to add a photograph of the place they were interested in having environmental monitoring done.

2.5.2. Hand-held Monitor Loans

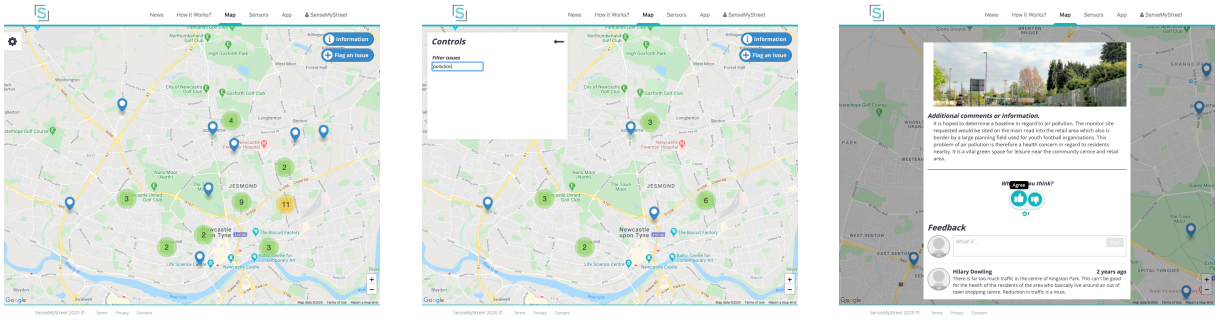
To investigate the issues, people could loan hand-held environmental monitors from the toolkit. The SenseMyStreet website provided people with information about the sensors and a link to the sign-up form (Figure 2.11a), in addition to allowing them to explore the loan schedule (Figure

[42] https://ideaboard.co.uk
[43] https://sensemystreet.uk



(a) Register or login **(b)** Choose a location on the map **(c)** Fill form to submit an issue

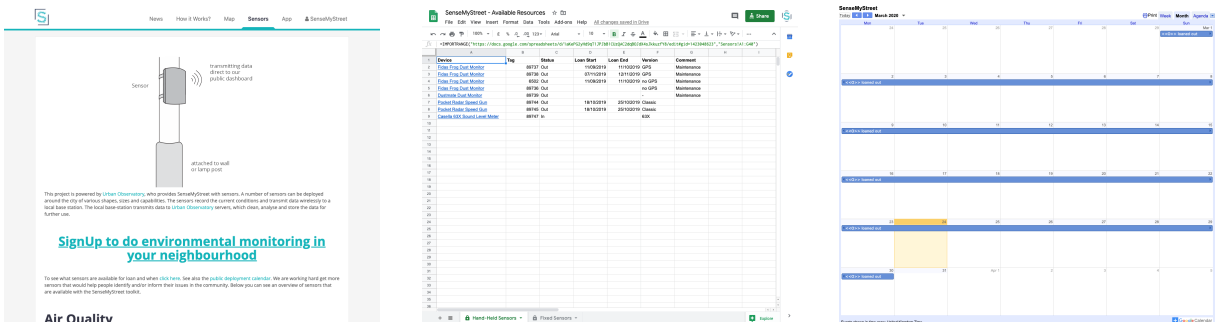
Figure 2.9 Issue flagging on the SenseMyStreet web platform



(a) Explore flagged issues **(b)** Filter by issue **(c)** Discuss and show support

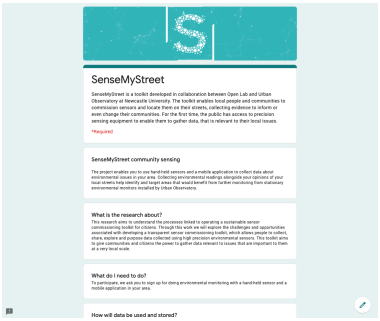
Figure 2.10 Exploring issues on the SenseMyStreet web platform

2.11b) and the deployment calendar (Figure 2.11c). Generally, all the monitors – hand-held and their fixed counterparts – could be categorised into three themes: *air quality*, *noise* and *traffic*. To sign up for a sensor loan, people needed to use a dedicated online form (using Google Forms), which consisted of three parts: a brief description of the project and research consent agreement (Figure 2.12a); the sensor loan agreement (Figure 2.12b); and the sensor equipment and handover details (Figure 2.12c). The details that needed to be provided were the type of sensor to be loaned; the reason for the loan; the time period for the loan; and contact details for sensor pickup and drop off.



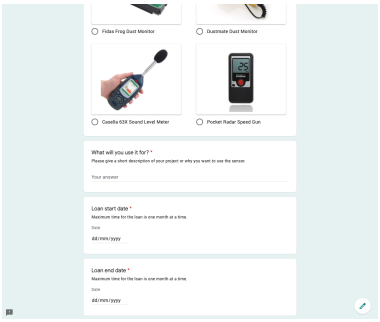
(a) Learn about available sensors **(b)** Look at the loan schedule **(c)** Explore deployment calendar

Figure 2.11 Exploring hand-held sensors and deployment information



(a) Learn about the research project

(b) Read the loan agreement



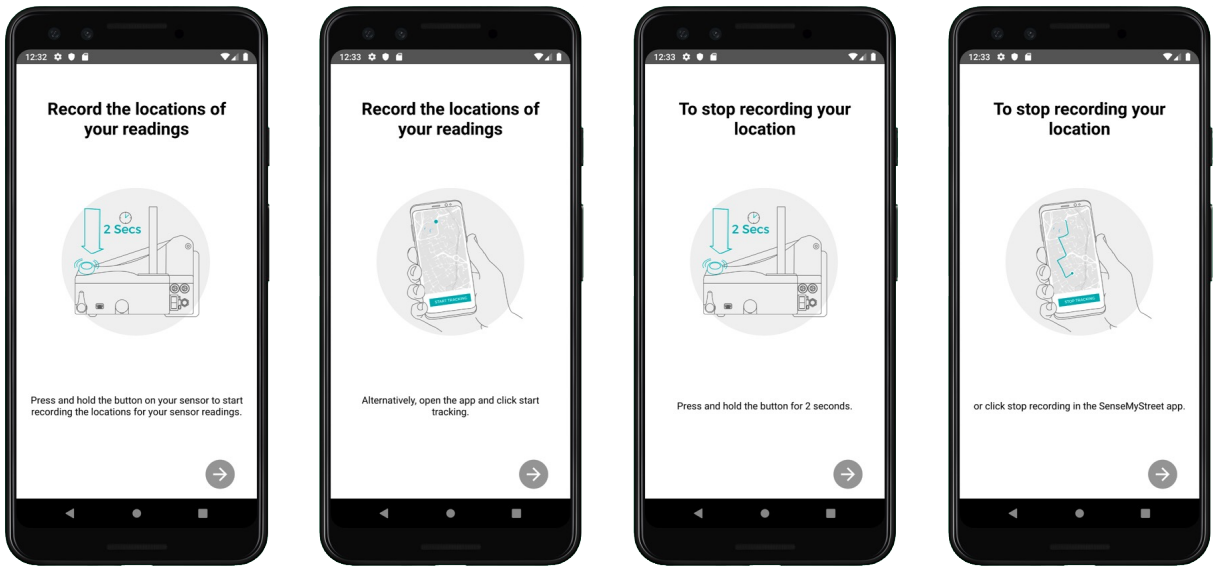
(c) Choose a sensor to loan

Figure 2.12 Hand-held sensor loan process

Figure 2.13 SenseMyStreet hand-held sensor kits

2.5.3. Community Sensing

Once an individual or a community group received a hand-held sensor kit (custom designed kits shown in Figure 2.13) with instructions (Appendix A), they could start monitoring in their local area. Since there was no fixed monitoring schedule, people had to come up with their own, which often meant negotiating with others in the community to cover more ground. Hand-held sensing kits came pared with the SenseMyStreet mobile application, which enabled people to record additional qualitative data about their experiences. The mobile application enabled people to record a GPS track of their monitoring journeys (Figure 2.14); record events (Figure 2.15a)



(a) Start recording GPS via blue-tooth button

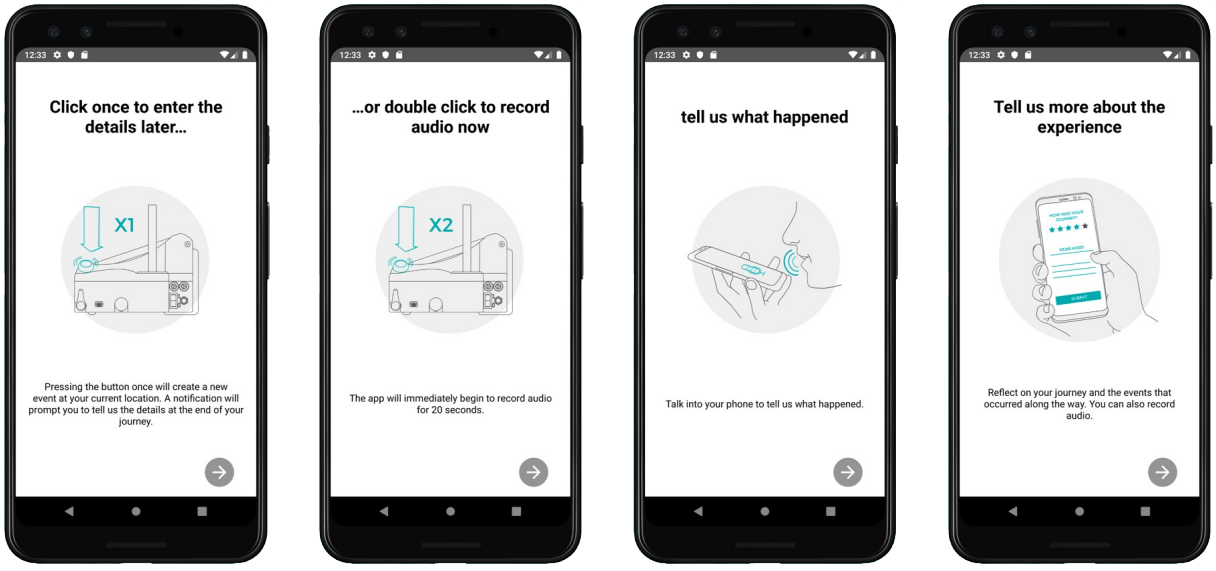
(b) Start recording GPS via app

(c) Stop recording GPS via blue-tooth button

(d) Stop recording GPS via app

Figure 2.14 Recording GPS locations while doing hand-held monitoring

and audio snippets while they were monitoring (Figures 2.15b and 2.15c); and also reflect upon their journey once finished (Figure 2.15d). All that data was then automatically uploaded and made available for everybody through an online map (Figure 2.4) that allowed people to look at the data and reflect on the hand-held sensor readings.



(a) Click to record an event to reflect on later

(b) Double click to record audio reflection

(c) Use app to record audio reflection

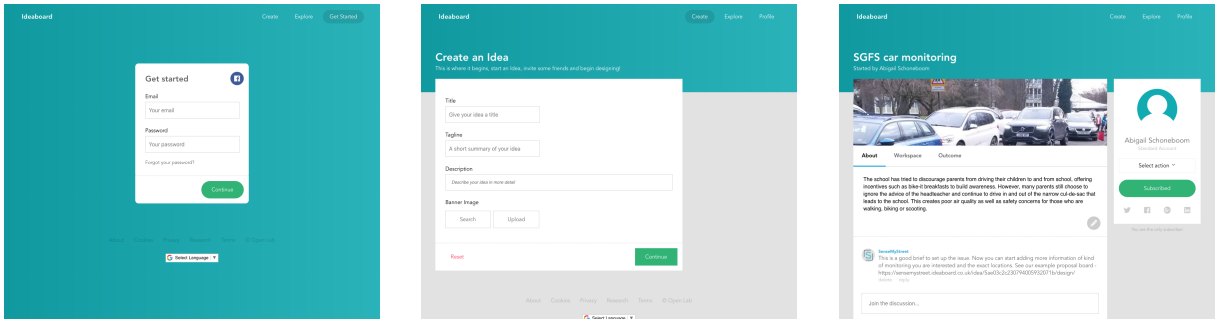
(d) Reflect on the journey and events

Figure 2.15 Adding reflections to journeys made while using hand-held monitors

2.5.4. Sensor Commissioning

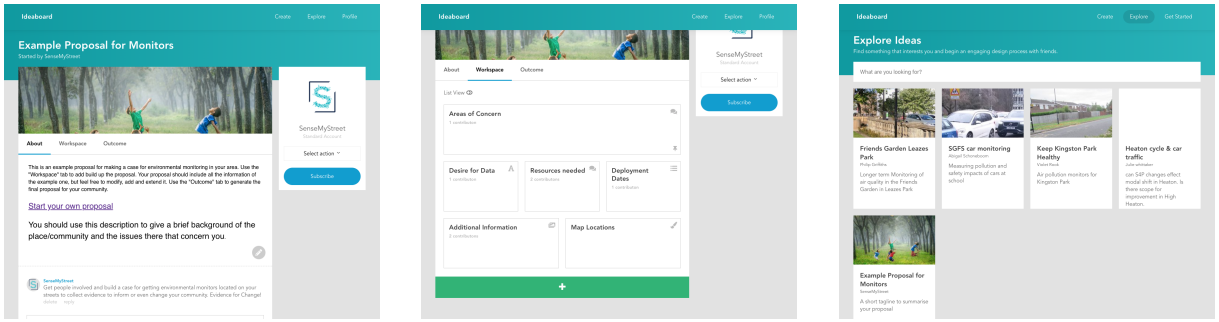
After the community sensing step and reflecting on the results of the monitoring, people could start collaborating around a proposal for sensor commissioning from the UO. The planning for

this could have taken place offline, at community meetings, but in order to propose extended monitoring, people had to use the SenseMyStreet Proposals website. To start collaborating around a proposal, a person needed to register on the platform (Figure 2.16a); create an idea with a title, tagline, short description and banner or image (Figure 2.16b); and start off the discussion with the community (Figure 2.16c). There was no fixed structure for proposals or the discussion; however, an example proposal (Figure 2.17a) was provided on the web platform, which was modelled on the first proposal submitted by the first community group (Section 2.3). The elements that were essential for the proposal were related to questions of *what, why, where, and when*: (1) what data is needed?; (2) why would such data be useful?; (3) where should the monitors be deployed?; and (4) when should the deployment take place? Additionally, people could explore other proposals, learn from them and join the discussion (Figure 2.17c). Once the proposal was ready, the platform enabled people to consolidate it into a proposal document (Figure 2.18).



(a) Register or login **(b)** Start a proposal **(c)** Start the discussion

Figure 2.16 Process of starting proposal for sensor commissioning



(a) Look at the example proposal **(b)** Learn about the proposal **(c)** Explore other proposals

Figure 2.17 Example proposal for sensor commissioning

2.5.5. Data Flow

Figure 2.19 shows the data flow through the whole of the SMS toolkit, illustrating how data was exchanged and synced using the Geo API functionalities of the SenseMyStreet web platform. Using Firebase, the mobile application synced GPS tracks to the SenseMyStreet web platform using the Geo API, which was then used by the data parser to merge with sensor readings.

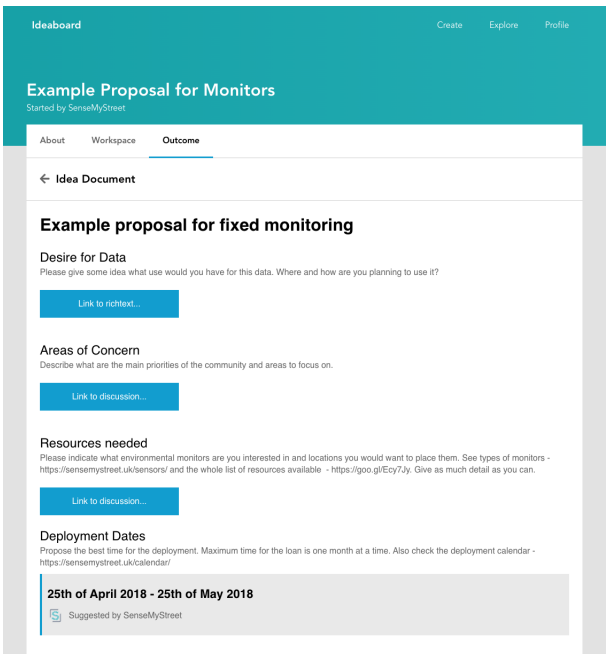


Figure 2.18 Example sensor commissioning proposal document

Additionally, the locations of issues flagged by people on the platform were synced together with the processed hand-held monitor readings (from the data parser stored in Google Drive) to the Carto platform to be visualised on the map (Figure 2.4).

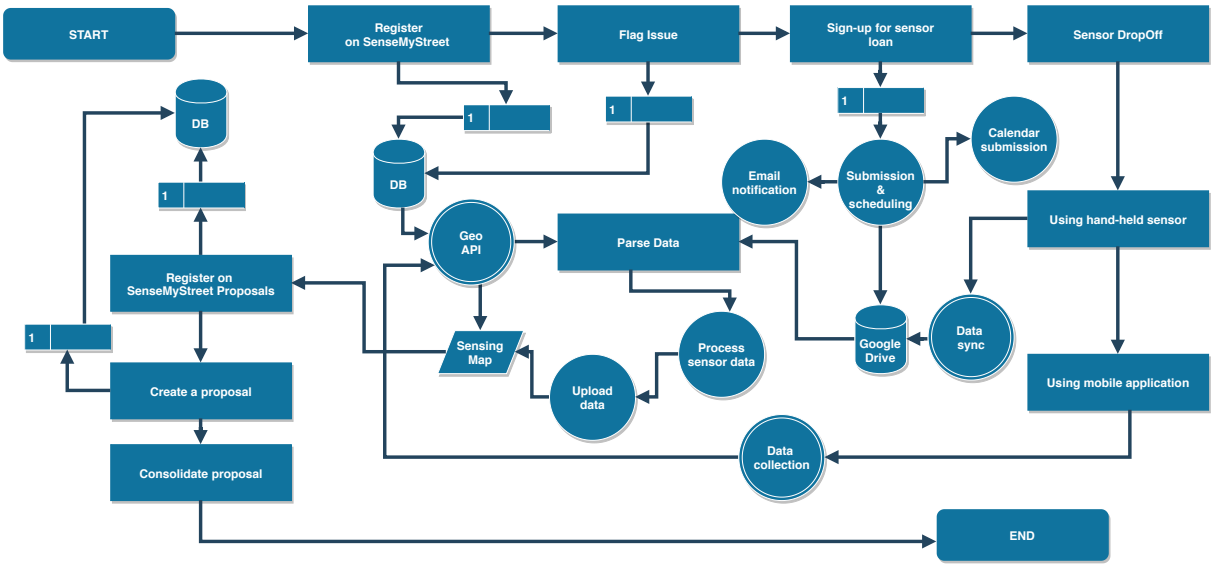


Figure 2.19 SenseMyStreet toolkit data flow

2.6. Toolkit Usage Analysis

2.6.1. Collected Data

This section reports on the usage data of SMS and the observed interactions with it across different platforms and tools linked to the toolkit. Participants included in this reporting are everyone who took part in community engagement activities and the design of the toolkit or

who engaged with any of the toolkit's digital platforms. An overview of the submitted issues are given with their geographical division in relation to the Index of Multiple Deprivation (IMD)[44]. Additionally, this analysis provides an overview of activities for different groups and individuals who took part in the case study and engaged with the digital systems.

The SMS toolkit enables the collection of metrics from three different bespoke user-facing platforms: the SenseMyStreet website, the automated sensor loan scheduling system, and the SenseMyStreet Proposals website. Each system uses its own technology (Section 2.4.2) to collect statistics about the engagement:

- The SenseMyStreet website allows reporting on the number of people signed up and reported issues and provides locations and information about inserted issues;
- The automated sensor loan scheduling system allows reporting on the number of people that applied for conducting hand-held sensing and their advocacy group association; and
- The SenseMyStreet Proposals website allows reporting on the number of proposals created and the groups involved;

In addition, Google Analytics provided metrics collected from both websites (SenseMyStreet and SenseMyStreet Proposals), while additional external third-party platforms in the SMS toolkit, such as Carto Maps[45], provided the number of views on each map; however, they were excluded from the analysis on the basis of their lack of use in this particular evaluation.

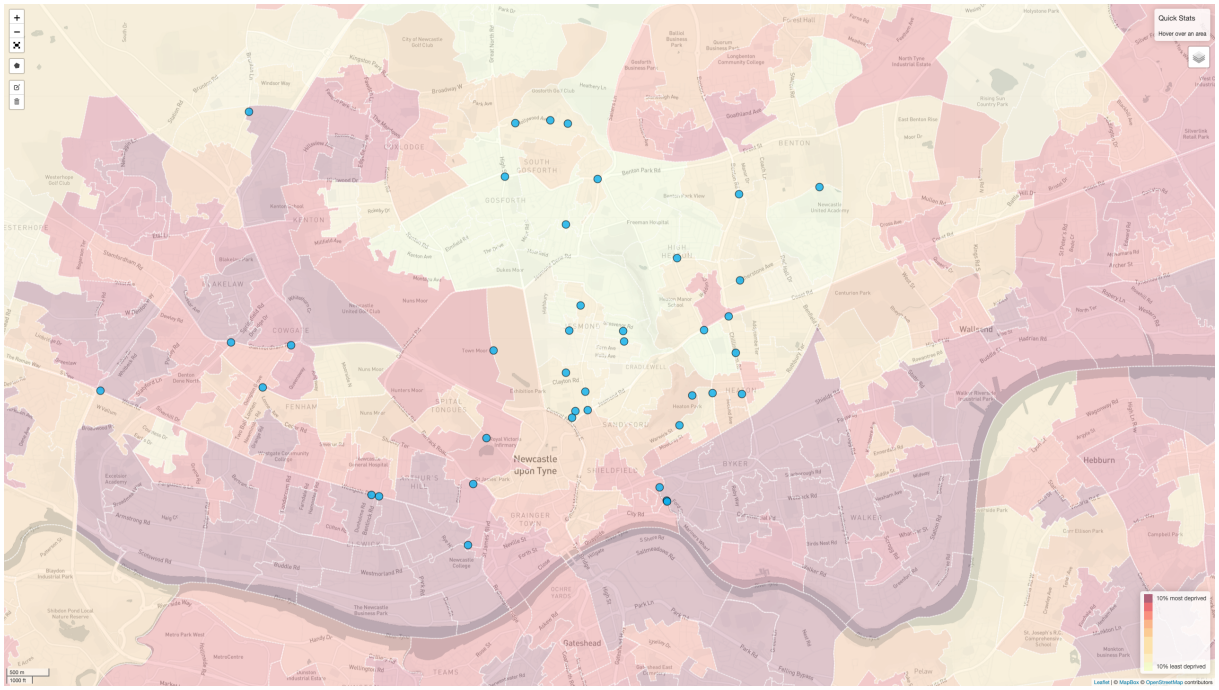


Figure 2.20 Flagged Issues mapped against Indexes of Multiple Deprivation (IMD) statistics: from 10% of most deprived to 10% of least deprived

The SMS toolkit was officially launched in the summer of 2017. Over two years leading up to this evaluation, 65 people had signed up through the SenseMyStreet website and 47 issues

[44] https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015
[45] https://sensemystreet.carto.com

had been submitted using the mapping tool on the site. From there, six groups (G1-G6) and seven individuals (I1-I7) signed-up (Table 2.6) to borrow hand-held monitors, using the sign-up form linked to the automated sensor loan scheduling system. Following that, four proposals were submitted through the SenseMyStreet Proposals platform for commissioning fixed monitors to be installed by the UO, and four deployments have been carried out. Additional data used for this analysis originated from ethnographic work, including field notes from private meetings, community events organised by the group and public meetings around particular issues of concern. Although each platform was independent and often required users to sign up separately, ethnographic work conducted by the author enabled links to be made between each participant's activities across different platforms and an overview of each participant's level of engagement to be compiled. However, the data collected to report on the case study is not definitive as the toolkit is still operational, is constantly evolving, and is being used by communities to gather data for exploring and evidencing issues in their neighbourhoods.

2.6.2. Analysis

Table 2.5 shows that the majority of issues were linked to air quality in the neighbourhood, which was also one of the main drivers of this case study (Section 2.2.2). Issues with traffic were also mainly related to the air quality issue; however, one of the traffic monitoring submissions was to do with counting the number of people using the new cycle lane. The issue form also enabled people to flag issues about matters the toolkit could not measure at that point. The other issues submitted where about measuring water level and quality in the river and vibrations from the building works in a residential area. This could help expand the toolkit in the future and acquire new sensing equipment to respond to citizen's concerns.

Issue Type	**Count**
Air Pollution	38
Noise Pollution	3
Traffic	3
Other	3

Source: Author

Table 2.5 Distribution of submitted issues on the SenseMyStreet web platform

Furthermore, having the mapping component on the SenseMyStreet website and the availability of Geo API enabled analysis of the issues through geography. Figure 2.20 illustrates all the submitted issues mapped against the IMD, which is the official measure of relative deprivation for small areas in England and Wales, as outlined by the Office of National Statistics (ONS). These statistics are published at the level of Lower Super Output Areas (LSOAs) that contain an average of 1,500 residents within any given boundary. The IMD uses seven domains to produce this overall measure: income, employment, education, skills and training, health and disability, crime, barriers to housing and services, and living environment. Using this metric, it is possible to look at the socioeconomic characteristics of the areas and draw a comparison between different neighbourhoods within England and Wales. When looking at the flagged issues against the IMD,

it can be seen that the majority of the issues flagged are in the 10% of the least deprived areas according to the metric. This may be an indication that the people who engaged with the toolkit were living in those areas. However, the map shows that some issues were flagged on the border areas of different LSOAs, which could indicate people living in either area.

Based on participants' self-reporting and the ethnographic work, an analysis of the overall engagement was conducted, which reports on the activities across the digital platforms developed to support the process of community sensor commissioning. In order to understand the factors influencing active citizen engagement with the commissioning toolkit, a comparison of activities for each community group or individual has been analysed. Table 2.6 shows the level on engagement and usage of the different technologies across the toolkit. The key aspects considered are: existing community *group*, formed *issues* mapped on the SenseMyStreet platform, engagement in hand-held *monitoring*, formulation of a *public* (Dewey, 1954) around the issue, submission of a *proposal* for sensor commissioning, an action *plan* for data use, and usage of *data* by the community. Additionally, each individual's or group's communication *channel(s)* with the community are provided as a comparison.

The analysis indicated that the existence of a community group is not essential for starting engagement with the toolkit. In many cases, groups and publics formed alongside participating in the activities related to the SMS toolkit. However, it was paramount to identify the issues within the community to focus on at specific locations in the neighbourhood. Participants who did not identify the issues or did not have a specific agenda in mind did not manage to get further from hand-held monitoring. There was one special case where an individual's advocacy efforts might have had negative consequences on their quality of life, which meant that the issue was dropped by the individual. However, it could be picked up again if enough people are interested in the issue and want to do something about it. This indicates that there is a need for a public to emerge at some point in order to ensure the successful use of the toolkit and the data it generates. The findings also indicate that participant(s) who set a prerequisite action plan following data generation managed to make use of the data. This was prominent on both cases – when people used hand-held monitors to collect measurements or using data generated through commissioned monitors. There were instances where participants engaged with the toolkit up to the point of getting fixed monitors commissioned from the UO, seeing that as the end result of the engagement. Although the data generated may be useful in the future, e.g. when infrastructure changes are implemented in the area, without any promotion or intended use by citizens, it does not serve the purpose with regard to the effective use of data by the community for civic participation, advocacy and action.

2.7. Discussion

The data that is being produced about the way the city operates and how people interact with it allows us to see the city through a completely different lens. This is also changing the way citizens interact with the city and how they express their concerns and participate in decision-making processes. An integral part of this is the integration of civic technologies that aim to

Participant(s)	Existing Community Group	Mapped Issues	Hand-Held Monitoring	Formulated Publics	Submitted Proposal	Planned Action	Used Data	Channel(s)
G1	✗	✓	✓	✓	✓	✓	✓	Blog, FB, Newsletter, Twitter
I1	✗	✓	✓	✓	✓	✓	✓	Blog, Twitter
G2	✓	✓	✓	✓	✓	✗	✗	FB, Website
I2	✗	✗	✓	✗	✗	✗	✗	None
G3	✓	✓	✓	✓	✗	✗	✗	FB
G4	✓	✗	✓	✗	✗	✗	✗	Website
I6	✗	✗	✓	✗	✗	✗	✗	None
I4	✗	✗	✗	✗	✗	✗	✗	None
G5	✓	✓	✓	✓	✓	✓		Website, FB
I3	✗	✓	✓	✓	✓	✗		None
G6	✓	✓	✓	✓				FB
I5	✗	✓	✓					None
I7	✗	✓	✓					None

Source: Author

Table 2.6 SenseMyStreet toolkit usage: Blank fields indicate that the engagement is ongoing (no data yet) and 'None' indicates no data available

engage people to participate in these processes. This section discusses how commissioning infrastructures, such as the SMS toolkit, could open up new opportunities for more active citizen data production and civic participation in the smart city.

2.7.1. New Models of Active Citizen Participation

The literature review (see Chapter 1) and motivation of this study (Section 2.1) covered a variety of different technologies and modes of participation (i.e. passive and active) that aim to engage people in all matters of civic life. Furthermore, scholars in design and HCI have also started to distinguish between designing for immediate use or a response to a particular inquiry and designing for future unseen uses (Björgvinsson et al., 2010; Dantec and DiSalvo, 2013; Ehn, 2008). This is what is referred to as *infrastructuring*, which essentially means using design practices to support capacity building in the community.

However, designing a commissioning platforms situates itself somewhere in between, where the aim is to develop a generic infrastructure that could also be appropriated by people in their pursuit of issues, e.g. systems such as App Movement, where people come together to leverage technological infrastructure to commission bespoke location-based mobile applications (Garbett et al., 2016). Although it is packaged as a finished platform or toolkit (i.e. product or a thing), it is driven by a participatory design process, facilitates the discovery of issues and also promotes the creation of publics (Dewey, 1954). This contradicts what Ehn (2008) has suggested; however, similar findings were discovered from a community project around the design and use of a Community Resource Messenger (CRM) system for homeless communities living in urban areas (Dantec and DiSalvo, 2013). Although the CRM system was designed as a response to a practical need of a population, it still facilitated the creation of publics and their attachments around issues. Community commissioning platforms such as App Movement and SenseMyStreet, while not shying away from being designed as purposeful systems, are built for uses that are defined through engaging with it and issues of concern.

The SMS toolkit and the process of sensor commissioning successfully provided mechanisms to carry out a citizen-led deployment of environmental monitors and helped generate data about issues important to communities. The toolkit also created a situation were data was demand-driven rather than the usual open data portal's ideology – 'If we put data out there, people will use it'. Furthermore, taking things online enabled the commissioning process to scale and extend the participation through different forms of communications for people unable, or unwilling, to engage in face-to-face dialogue; for example, taking the proposal creation online provided a way for people who could not make or were not aware of community meetings to participate in the discussion and get their voices represented. Additionally, it also helped to document the process for transparency and reproduction, sharing the knowledge with other groups who may want to carry out their own investigations and submit a proposal of their own. Proposals were also looked over by expert professionals at the UO who could give communities guidance on how to get the most out of the deployment in terms of data quality and placements of the fixed monitors.

Indeed, technology is changing the way we understand infrastructuring. Commissioning of resources and provisioning of infrastructuring through platforms like SMS certainly raise a new set of questions regarding the way infrastructuring ought to be configured. Commissioning platforms have the ability to scale up or streamline the processes of infrastructuring, at the same time as preserving the components of participatory design for constituting publics to develop attachments to act upon (Dantec and DiSalvo, 2013; Marres, 2007). Within this somewhat generic system for engagement, it is important to recognise the diverse nature of these formed publics and provide different levels of support for taking action.

2.7.2. Physical Infrastructure and Infrastructuring

Moreover, there needs to be a distinction between providing physical infrastructure and infrastructuring engagement around it. With the case of SMS, the aim was to help people access the physical resources of the smart city in order to engage in the exploration of local issues. Hence, the question is: Can the act of using or deploying the sensors be seen as a successful effort of infrastructuring? The availability of physical tools (i.e. environmental monitors) can provide support for community action through enabling people to participate in the creation of new sources of data that did not exist before. So, in some regards, certainly, commissioned senors helped generate data about the local community and inform issues of concern in the neighbourhood. Furthermore, in some cases, it enabled people to voice their concerns by making them more visible to others through appropriating the toolkit resources. However, the availability of physical infrastructure and the act of participating itself did not facilitate the creation of, or was not sufficient enough by itself to form, the publics (Dewey, 1954) needed to move towards citizens taking action. In this sense, it needed to be integrated with activities supporting the usage of these physical things, i.e. it needed infrastructuring around the physical resources. The SMS toolkit approached this through a four-step exploration process that facilitated the framing of issues, gathering support and developing attachments. Issue discovery was built into the system

through participatory GIS and citizen sensing to align data generation with particular concerns and to help form publics around those issues. This aligns with the ideas of Marres (2007) and Dantec and DiSalvo (2013) regarding infrastructuring, where socio-technical processes and resources are put in place to support imagined futures.

However, without going through the process, the physical infrastructure did not necessarily provide the effective use of the resources. This was evident in a couple of cases where people signed up to borrow hand-held monitors without defining the issue (i.e. mapping it through the platform) they were interested in (by skipping the first step of the toolkit). Some of them did not collect the monitors and some just tested the monitors for a brief while and then stopped doing so. That is not to say, however, that self-discovery and learning about technology are not useful for framing issues. When reflecting on the process of commissioning, it seems that getting the commissioned sensors deployed was actually marking a start of additional activities that related to the effective use of data by citizens. In a similar way, as pointed out by Dantec and DiSalvo (2013):

> *[...] infrastructuring comes as a result of the reconfigurations that occur around and with a technological intervention; that is, the deployment of the technology is a beginning, not an end.* (Dantec and DiSalvo, 2013, p .249)

It has also been documented that ownership plays an important role in infrastructuring because of the way it steers people towards future action (Dantec and DiSalvo, 2013), whether it is taking ownership and building attachments to issues (Asad and Le Dantec, 2017; DiSalvo et al., 2009; Marres, 2007) or taking ownership of the designed technology (DiSalvo and Lukens, 2009). The SMS toolkit worked well in helping people take ownership of the issues by publicly posting them and enabling people to attach personal opinions on the matter. Taking ownership of the designed technology (i.e. the toolkit and all the resources), however, appeared to be much more multifaceted. Commissioning systems are usually designed as generic platforms or systems that connect multiple platforms and bespoke sub-systems, which is referred to as design 'appropriation' (Dix, 2007) or an 'unplatformed' design Lambton-Howard et al. (2019) approach, implying the use of online platforms for purposes that they where not initially designed for. The SMS toolkit could be considered as an unplatformed system because of its use of multiple configurable platforms like MakePlace, Ideaboard, Carto, and Google Drive, Forms, Calendar, mobile app and Scripting (with the addition of physical infrastructure in the form of scientific-grade environmental monitors). However, a hidden part of the toolkit was also the communication channels (e.g. FB, Twitter, and blogs) that groups used to exchange ideas and share information. These channels often already had an existing community that could be integrated into the system. While it would have been impossible for the community to take ownership of the tools and technologies, people could take ownership of the resources generated as a result of the process, or, putting it in Dantec and DiSalvo (2013)'s terms, people taking ownership of the data generated and knowledge created through the commissioned monitors and their role in shaping future actions.

As designers, and particularly as engineers of these technologies for participation and civic engagement, there is often a desire to respond to people's requirements with a particular solution that addresses the issue. In the case of a commissioning toolkit, it could still be a product or a useful system, but it has to facilitate infrastructuring. Designing provisioning of resources through infrastructuring is more likely to support capacity building (in terms of skills and knowledge) that would benefit the community in the long run and provide recurring and sustained participation around issues of concern.

2.7.3. Active Citizen Data Production

There is ongoing debate around data ownership in regard to its generation and rights to access[46]. This also extends to the idea that, by law, data has no owner, whereas the collection of data does. However, if the data is commissioned by the community, does this mean that the community owns this data? The SenseMySteet project and the UO, who deployed the fixed monitors, both work in the spirit of transparency and openness, which means that the data is made publicly available using open licences. If data is in the public domain, who does it belong to and can the community take ownership of it?

The act of citizen data generation was actually a two stage process: (1) citizen sensing using hand-held monitors (2) and commissioning environmental sensor data from the UO. In the first stage, people engaged in the data collection themselves by using borrowed, hand-held monitors to learn more about the issue and what goes into environmental monitoring – the issues with the technologies, the uncertainties with the readings and the value of the data. This could have helped the community get a personal perspective and take ownership of the issue in hand. Enabling citizens access to the smart city tools that are already used by city officials and scientist responded to critiques of citizen sensing projects, which are often derided because the types of sensors used by the public are considered to be 'toys' (Gabrys and Pritchard, 2018). In the second stage, people commissioned stationary monitors to be placed in their neighbourhood by the UO.

Having two types of data generation presented itself to be a kind of double-edged sword. On the one hand, having people do their own investigations helped them better understand and frame the issues. The data collected by people using hand-held monitors on their commutes was not used in advocacy efforts in the community but was predominantly used by the groups themselves to explore the issues and target particular areas for prioritising. Furthermore, because there were no official regulations regarding how to conduct the monitoring, the research team could not guarantee the rigorousness and transparency of the processes. For people, it was easier to take full responsibility and ownership of the data because they had invested their time in it, and it was important to them to get the best data out of the process for future actions.

However, the second stage of data generation was carried out by professionals at the UO. This meant that the community-commissioned monitors received the same attention and went

[46] https://royalsociety.org/-/media/policy/projects/data-governance/data-ownership-rights-and-controls-October-2018.pdf

through the same transparent deployment process as they would if the data had been collected for scientific research or policy purposes. Additionally, having the data for the whole city accessible in one place helped avoid creating another 'data silo' and also enabled people to compare the sensor readings from their area to other areas, thus giving them an idea of their neighbourhood matched up to others. When communities handed over the task of data collection to the research team at the UO, they put a lot of trust in the process; however, that often meant they also handed over responsibility for appropriating the collected data. This was evident with some groups that participated in the study, where they commissioned the sensors to provide data for the community without having a particular aim to use it themselves in the future:

> *Air Pollution Monitors along Brunton Lane near the roundabouts would be beneficial to everyone.* (posted on SenseMyStreet Proposals platform)

Unfortunately, this meant that once the stationary monitor was installed, the group stopped interacting with the toolkit because there were no next steps planned with the data. Although the commissioned sensor data was published on the UO city data portal, which was available to everyone, it was mainly up to the citizens to make use of the data and turn it into activities for local benefit. The toolkit had created an opportunity for communities to access resources to investigate local issues; however, generating and making data available is only the first step in its effective use by the community. The initial aim of the case study to come up with mechanisms to democratise data production and use in the smart city was met through the design, development and deployment of the community sensor commissioning toolkit. The indicative findings from the toolkit usage suggest that there is an importance to planned action and establishing a strong community network. It is not only important for keeping up engagement with the toolkit, but also for sharing skills, increasing social capital and building community knowledge. There is a need to identify the underlying relations that help constitute networks as supporting infrastructure for proximate communities. Furthermore, these should also be integrated into the design of the digital tools and processes to promote creation of these networks, particularly for communities that are placed in a disadvantaged position. However, the utility of this infrastructure to act as a catalyst for civic advocacy and action needed to be assessed through an evaluation of its use by the communities. Hence, an extended analysis of the case study will be presented in the following chapter, which provides an in-depth chronological breakdown of the activities of the initially engaged group of residents (Sections 2.2.3 and 2.3) using the commissioned sensor data for advocating for change in their community.

2.8. Summary

This chapter described the design, development, deployments and analysis of SMS, a sensor commissioning toolkit for communities that enables people to use scientific environmental sensing equipment in order to investigate local issues and commission environmental sensors from the UO, placing them in the neighbourhood to gather data relevant to community issues at a hyper-local scale. The chapter outlined, in detail, the iterative design process of setting up SMS and provided an overview of the processes and digital technologies linked to operating a sustainable sensor commissioning toolkit. Furthermore, a detailed description of user interactions on the platforms was provided, in addition to how data is exchanged within the SMS ecosystem. Finally, a analysis of the toolkit usage was offered to report on the ways people engaged with the technologies. An extended analysis of how different groups engaged with the digital tools of the toolkit was also provided so as to understand the key factors influencing the toolkit's effective use. The purpose of this step was to achieve an overview of the usage of the system, while also developing a conceptual understanding for the analysis of the case study in the following chapter. In this regard, Chapter 3 will provide an analysis of the use of the data by a particular advocacy group to further develop an understanding of the effective use of data by communities.

Chapter 3. Evaluating SenseMyStreet: The Effective Use of Community Commissioning Resources

Never doubt that a small group of thoughtful, committed citizens can change the world: indeed, it's the only thing that ever has.
— Margaret Mead

This chapter is part of Case Study I and focuses on an evaluation of the effective use of the SMS toolkit 'in the wild' by communities. The SMS sensor commissioning toolkit's process, which was described in detail in Chapter 2, consisted of four main parts: (1) Identify the issues; (2) Plan and propose; (3) Get the facts; and (4) Data for everyone. Citizen engagement and civic participation have become an important focus across engineering, science and technology and participatory design studies. As more technologies are being integrated into the city, design researchers and systems developers are exploring new ways of enabling citizens to participate in the formulation of issues and take part in decision-making processes. Setting up the SMS toolkit (Chapter 2) illustrated how user-centred participatory design processes can be used to engage people in the discovery of local issues using scientific environmental sensor technology. The toolkit not only aimed to help empower citizens investigate issues relevant to them, but also tried to engage the wider public with the work of the UO and the bigger concept of a smart city by helping to get people's voices into the city's datasets. Expanding upon the findings related to the toolkit's usage by communities presented in the previous chapter, this chapter focuses on understanding engagements with the toolkit by a particular community group that was involved in its initial design (Chapter 2 Section 2.3). Evaluation was conducted over a period of two years and reported in two phases: (1) analysis of the discussions of a specific advocacy group on FB (i.e. community discussions) and (2) analysis of audio-recorded reflections of the core members from the same group on their advocacy efforts and the role of data generated using the toolkit (i.e. community reflection). Through these phases, it was possible to analyse the engagement of the community in different stages of the toolkit process and present findings on how the community responded in each step of the process, how they made use of the data and how the focus of the group changed over time when new challenges surfaced or opportunities became available.

3.1. The Evaluation Design

The evaluation of the SMS toolkit was conducted over a period of two years (Figure 3.1). Drawing on the findings of the toolkit usage presented in Chapter 2, the evaluation was carried out in two phases using different data sources and methods:

PH I. (*community discussions*). This phase of the evaluation examined people's discussions around using the toolkit for civic advocacy and action and how people perceived the tools, appropriated them and used the data generated by the tools in their advocacy efforts. It focused on the discussions between people in a community group involved in the design of the toolkit (Chapter 2 Section 2.3) and analysed posts, comments and replies made on a private FB group that the researcher was given access to. The coding guide used for analysing the empirical data was derived inductively from the initial findings of the toolkit usage presented in Chapter 2 and from the fieldwork conducted by the author.

PH II. (*community reflections*): This phase focused on the participants' reflections on the SenseMyStreet toolkit, collected through a focus group with the key stakeholders from the same community group studied in PH I. It carried out a qualitative analysis on the transcribed focus group data using a thematic analysis approach (Braun and Clarke, 2006) with confirmatory content included from the FB data analysed in PH I.

Finally, the evaluation includes an overview of the *community outcomes* from the project, which are presented through example uses of the data, generated through the use of the system and the outcomes of people's advocacy efforts using that data.

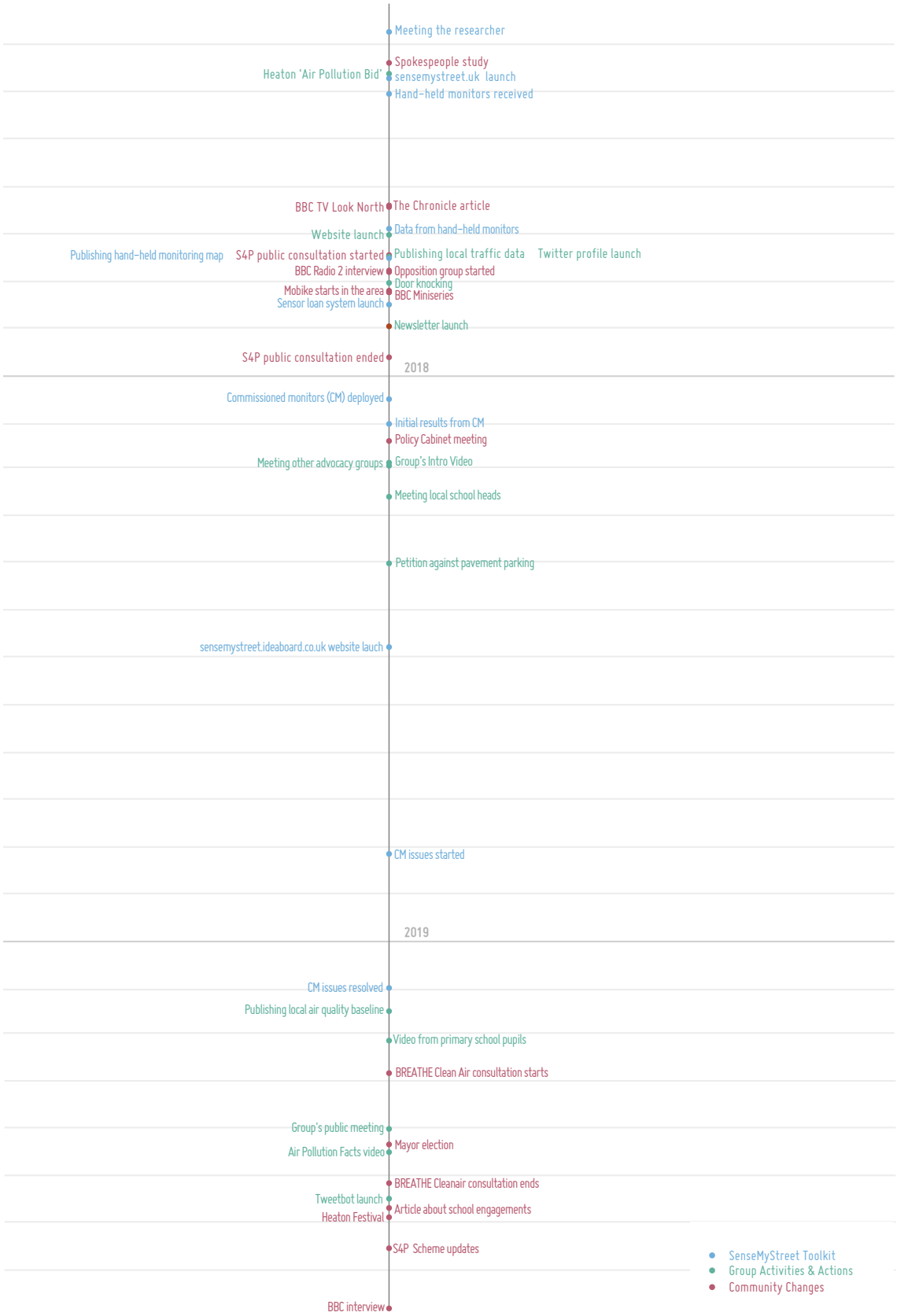
3.2. Data Collection and Analysis

This section describes the different datasets included in each phase of the evaluation and how they were collected and analysed. Each subsequent phase of the analysis builds on the insights learned from the previous phase so as to build a detailed evaluation of the case study.

3.2.1. Community Discussions: Phase I

About the same time that the first design activities for the SMS toolkit kicked off, a group of residents (n=8) in a local community (Chapter 2 Section 2.2.3) started a private FB group to exchange information, discuss local developments and organise their advocacy activities. The same people were the core group involved in the design and development of the SenseMyStreet toolkit described in Chapter 2. The author was added to the FB group, initially to exchange information and coordinate the deployments; however, he was retrospectively given permission to use the discussions in the analysis of the case study.

The data used in the first phase of the evaluation was retrieved using the FB graph API consisting of 265 parent-posts and 1,092 comments and replies (a total of 1,357 digital objects) that were posted on the group's page between April 2017 and July 2019. Taking a contextual



Source: Author

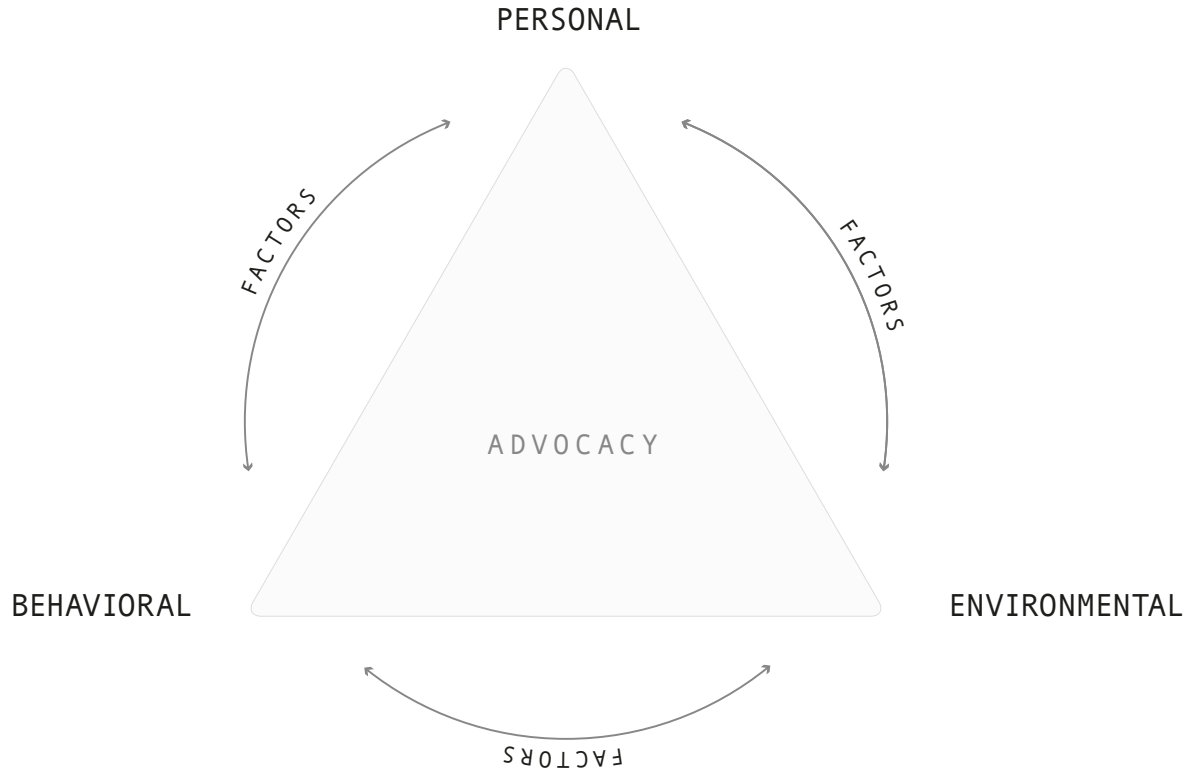
Figure 3.1 SenseMyStreet case study timeline with events

design inquiry approach (Beyer and Holtzblatt, 1997) through the analysis of the toolkit usage presented in Chapter 2 and the fieldwork conducted by the author – which studied different groups using the toolkit and appropriating generated data using the environmental monitors – three key factors were inductively derived that influenced the use of the toolkit and the data it provided: *personal*, *environmental* and *behavioural* factors (Figure 3.2). Similar factors have been used to describe social learning according to social cognitive theory, originally developed by Bandura (1977). These three factors where used to analyse and cluster the empirical data in terms of each factor's importance in a particular advocacy effort (Figure 3.1). Each post discussing an advocacy effort was assigned a code according to the main factor of influence. The importance was derived from an absolute number of advocacy efforts discussed linked to each factor, which was analysed from the perspective of how data and sensor technologies were appropriated as tools for civic action. However, an advocacy effort was often linked to multiple factors, at times to all three. This is not surprising because the factors are tied to each other in a reciprocal manner, meaning that they are in constant interaction with one another, which in turn enables learning, creation of knowledge and action to happen. However, the factors have a power relationship between them at any given moment, which enabled codes to be assigned with weights to each data snippet individually linked to an advocacy effort. The weights of the codes were the following: the strongest factor got 1, the one after that 0.3, and the last one 0.1 (if connected to all three). Furthermore, to filter the dataset and better analyse the interactions between people related to the different factors of influence, the following four rules were applied:

1. posts must have at least one reply;
2. posts must be replied to by at least one person other than the original poster;
3. posts must not be created for archival purposes (e.g. sharing a document or an image); and
4. posts must be more than simple information request related to the group administration activities (e.g. adding a member or sharing contacts).

Initially, these rules included a minimal length for the post. However, on closer inspection, documents and files were shared without a long textual description, which prompted discussion and feedback within the group. Some examples were an email thread between a council worker and a group member, or sharing an image of a local newspaper, posted through the letterbox of one of the members. Both of these started a discussion, and subsequent actions were taken by the group. However, the dataset excluded 'like' type messages (e.g. 'thank you', 'great', 'good work' and other confirmatory expressions), where people were agreeing with the original poster or expressing their gratitude. Although this would be an interesting dataset to study on its own, e.g. to look at the dynamics of the group, it was out of the scope of this particular analysis. After applying these rules, the dataset left for coding consisted of 193 posts and 647 comments and replies. In addition to the reporting on the key factors of influence for advocacy efforts and how they changed over time, chronological events and anonymised quotes from the dataset are provided to exemplify the activities linked to advocacy efforts. This reveals the main

focus of advocacy efforts for the community group, how it has developed over time and what role the sensor commissioning toolkit played in implementing strategies for civic advocacy and action by the community. Although the factors of influence could be derived from the previous findings and, in relation to social cognitive theory, code the data, there was little understanding of the relationships between the factors in the context of citizens using technologies and data in advocating for change. The aim of this analysis was to get a better understanding of what the relationships between the factors were and what role they played in the process of civic advocacy and action.



Source: Author

Figure 3.2 Key factors of influence for advocacy efforts

3.2.2. Community Reflections: Phase II

Deductively applying the factors of influence on data in the previous phase of the analysis helped gain insights and enabled the author to compile an informed view of the process of sensor commissioning by local communities. The second phase of the evaluation built on this and sought feedback from participants on the author's informed view. The data used in this analysis originated from a audio-recorded focus group with participants (n=3) who were the key members of the community group. The focus group was organised as fairly unstructured conversations, where participants were reflecting on the overall process of the community commissioning through the SMS toolkit and the group's main strategies for advocating for positive change in the community. It started off with an introduction, the processes they engaged in within the scope of the toolkit, initial strategies of the group and the current status of the group in relation to

utilising the data generated using the toolkit. This was followed by an open group discussion facilitated by the author, who also provided occasional prompts to keep people on track: *'What is our agenda?'*, *'What could we potentially do with the data?'*, *'What are we actually going to do with the data?'*, *'How are we going to do it and who is going to do it – the action plan?'*. The audio-recorded focus group was fully transcribed by the author, and a qualitative analysis was conducted on the data using a thematic analysis approach (Braun and Clarke, 2006). The key factors of influence (Figure 3.2 were applied to code the focus group data in order to gain understanding of the ways that people purposed the sensor commissioning toolkit as a resource for community problem-solving activities and for exploring local issues. Additional confirmatory content was also pulled from the previous analysis (i.e. posts and comments from PH I) to saturate the data. Under each theme, the sensor commissioning toolkit and its functionalities are reflected upon and reported using anonymised quotes from the transcripts to illustrate how the capabilities of the toolkit can be appropriated for citizen-led advocacy and action.

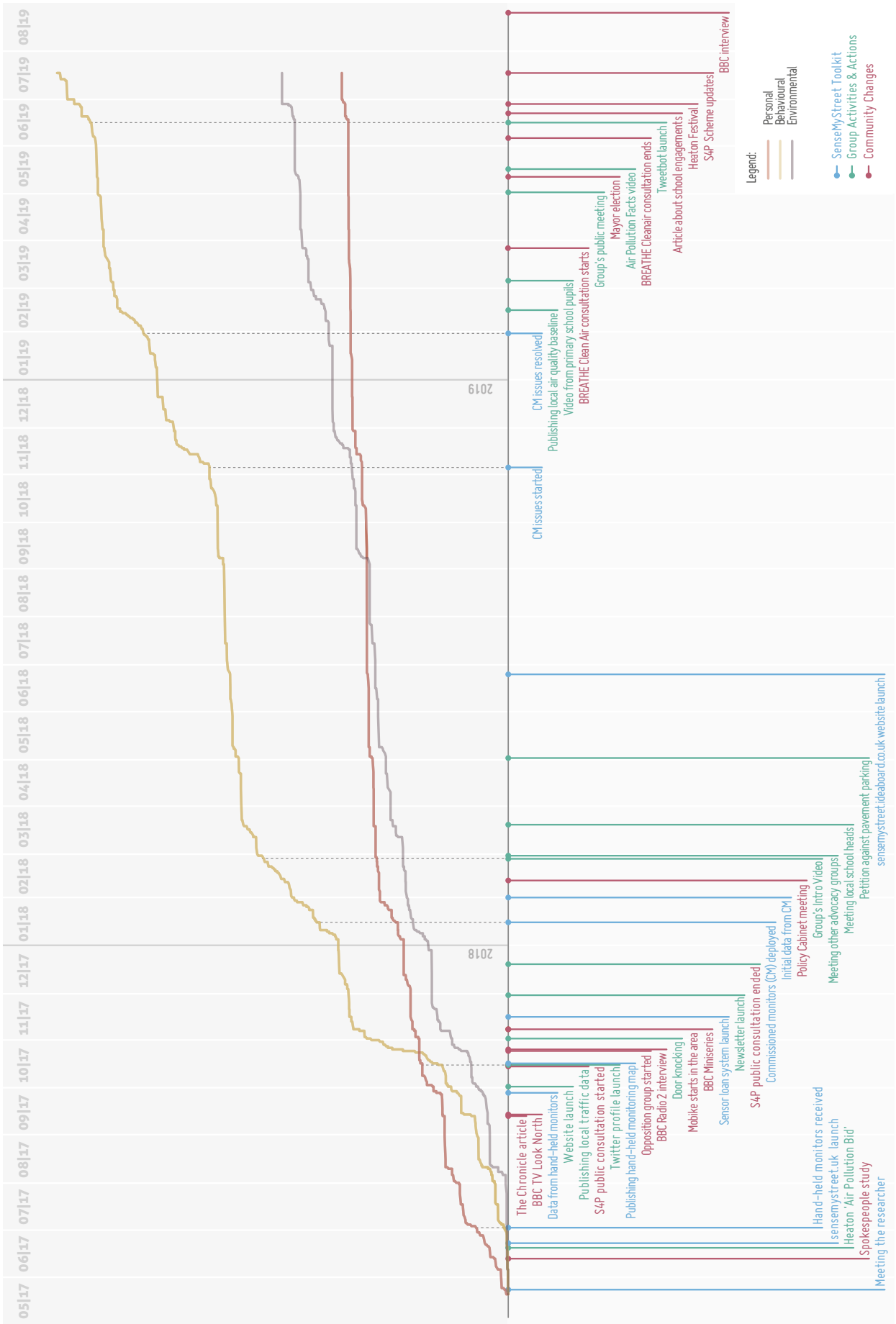
3.3. Findings

Taking a multifaceted approach and using the key factors of influence for advocacy efforts (Figure 3.2), this section reports the findings from both phases of the analysis to understand if and how communities could use the SMS toolkit and the data generated through the commissioning tools for community problem-solving activities and for civic advocacy and action.

3.3.1. Analysis of FB Data

To better analyse and represent the findings of the coded FB data, they were visualised by the author using the project timeline (Figure 3.1). Figure 3.3 presents the timeline graph of coded FB data in terms of the key factors of influence – *personal*, *behavioural*, and *environmental*. The graph shows the evolution of different strategies linked to the key factors of influence (Figure 3.2) over a period of two years. Additionally, three types of events are marked on the timeline: activities and events related to the SenseMyStreet toolkit, activities and actions of the advocacy group, and the events and changes in the community. Furthermore, sections of the the graph are highlighted (using a dashed line) where there is a surge of focused activities or where the power relationships between factors of influence changed.

As highlighted on the graph (05/27 – 07/17), the main factor of influence was initially related to *personal* factors – experimenting, experiencing and forming a good understanding of the situation personally, meaning that people where intrinsically motivated to start taking action. This also came out from the initial meeting with the group, where people stated that the immediate concern was to establish a knowledge base for the group to understand the issues and their severity and to start collecting the data to help make decisions. Similarly, this was also highlighted as a major motivator in the action plan (i.e. 'Air Pollution Bid') and in the FB discussion:



Source: Author

Figure 3.3 Coded FB data plotted on a project timeline with events

> *Firstly, there doesn't seem to be any information about the air quality in [the area] that is readily available to members of the public. I think it would be good to establish a baseline on the level of pollution within [the area], particularly along the main roads.* (posted on 06/06/2017)

Furthermore, once the group received the hand-held monitors for initial monitoring, they had to learn how the monitors worked and how they would collect the data and make sense of it. However, once people got the hang of it and could understand the data, they started to think about how to present it to the wider community:

> *We need to start on some explanation to go alongside the plots explaining what's being measured and what the results mean before putting it on our site. We'll need to explain particle pollution and safe limits too, but we can talk about that on Thursday.* (posted on 17/10/2017)

The second highlight (10/17) on the graph shows how the main factor of influence changed from *personal* to *behavioural* about the time that data, such as hand-held monitoring and local traffic data, become available to publish to the wider community. Furthermore, the group started their website and Twitter account to publish the data and engage with the local community online, in addition to organising door- knocking to do the same offline. At the same time, there were multiple events happening in the community: an opposition group or counterpublic (Warner, 2005) emerged, local consultations for infrastructure changes started, and media exposure for the group and the project increased. While the importance of the *behavioural* factor was steadily growing in the advocacy efforts, the group also started to focus more on the *environmental* factor. This was related to actively participating local consultations for built environment change, but also speaking at public meetings organised by the local council about issues in the neighbourhood and the data the group had collected.

The next major leap in advocacy efforts, as highlighted by the third dashed line (01/28), occurred about the time the community-commissioned monitors were deployed by the UO in the neighbourhood and the data was made available for everyone on the UO's city data portal. This presented an opportunity to publish the data to the community using the group's online channels; however, people again had to develop a better understanding of the issue by working together with the researcher and the team at UO and making sense of the sensor data. Hence, the group had discussions around posting an update to the community without going into great detail too early:

> *Good dispassionate set out of the position, [group members name]. It seems odd not to put in any 'editorial', even if just saying what we (SPACE) want and saying that the [Anon] Rd level looks like it might be in excess of [the] legal limit assuming other months were similar. Also mention its the only part of [the area] in AQMA [Air Quality Management Zone]. We could say we are concerned and will continue to monitor and share with Ward Councillors, schools and parents.* (posted on 09/01/2019)

The graph also shows a period of no activity in the middle of the evaluation, starting from the fourth highlight, which refers to the period when longer baseline data was obtained to help inform a better view of the air pollution issue. As illustrated by the events on the graph (02/18 – 05/18), the group used that time to develop relationships with local residents, schools, businesses and other advocacy groups from the city in order to expand their community links and social capital. Furthermore, the availability of the commissioned sensor data on the official city data portal made it visible to other stakeholders in the city, which meant that that data could potentially be used to advocate for policy and built environment changes. The graph shows how the influence of the *environmental* factor had become more predominant compared to the *personal* over that period.

The fifth to sixth highlight (02/18 – 11/18) on the graph shows the issues with the sensor technology that were flagged by the community and the researcher to the UO. Working together with the people and the research team at the UO, the issues were resolved by changing out the faulty sensors. Following that, there was a series of activities and actions linked to both (but not equally) *environmental* and *behavioural* factors of influence, showing a strong link between these two. The final highlight (06/19) on the graph shows the launch of the automated Twitter bot (Tweetbot), which was designed by the group and built by one of the members. This enabled the group to publish daily summaries from the commissioned sensors to the community online and print out graphs to take them to face-to-face community events or public meetings.

3.3.2. Thematic Analysis

This section presents the thematic analysis of transcribed audio data from the focus group with three key stakeholders from the advocacy group (Chapter 2 Table 2.4). The aim of the focus group was to seek feedback from the advocacy group in regard to the informed view of the author on advocacy efforts and the commissioning process established (and the uses of data) to the observations and FB data analysis in the previous section.

Personal Motivations

The initial motivator for the local people was a lack of information about the issues concerning them. When the author was first approached by the residents after the initial design workshop, this was the main reason why people wanted to start conducting environmental monitoring in their area 2. This was also reflected by the participants in the focus group:

> *Well, [we've] got to remember why we started it in the first place, which was to just find out where we were, really, wasn't it? See if things were good, bad, or getting worse or getting better.* (P2)

There was some data available on higher levels from the governmental agencies (e.g. DEFRA-modelled data) and already established baselines from other parts of the city, but not at the specific locations and the granularity that would confirm or refute the concerns of residents. Residents

were finding it difficult to contextualise the national and city-wide discussions without having any local information about those issues:

> *I think we were aware that there were a couple of city council run monitoring sites, you know? Sort of on the Cradlewell Bypass and the Civic Centre, but the scale of those is fairly large, and having the ability of zoom in to street by street level, you know, sort of neighbourhood level, was really interesting.* (P3)

Although there were already people who were intrinsically motivated and had come together because they wanted to do something about these issues, there was not enough hyper-local data available to build up a strong informed view of the issue. The community was relying on people's personal experiences of living in the area, and their view was often informed by stories and anecdotal evidence. This also meant that it was difficult for them to advise their fellow residents and the wider community on matters of concern:

> *I mean, it reflected the concern that various residents [that] were included had about air pollution in the city or part of the city [in] particular, um, and not knowing whether it was good, bad or indifferent. Not knowing where the problems lay and knowing how to advise people who talk to us or we relate to, you know, about whether there is a problem and where there's a problem and what, if anything, could be done about it.* (P1)

Furthermore, there were available statistics about about traffic in the area from one of the city-wide consultations, but there was no data about what it was doing to people's health and wellbeing:

> *I think what it also reflected [is] the fact that congestion in our part of the city has, over the last five or ten years, increased substantially. Lots of standing vehicles, the rush hour, 'so-called rush hour' is getting longer, both in the morning and the afternoon. And there was concern [about] what that increase in traffic is actually doing in terms of the health and welfare of people, particularly younger and older people.* (P1)

Additionally, the were national- and city-scale changes happening (e.g. clean air zones and transportation infrastructure changes) that could impact the local area and how people get around in the neighbourhood. The group was interested in establishing a baseline or a benchmark before the changes:

> *And I think, obviously, they [the city council] model it from DEFRA about [Anon] Road being one of the air pollution monitoring zones. I think one of the modelling from DEFRA is pointing towards that being a problem area. And what we didn't want to happen is that the council propose to route more traffic through [the area], where people live, to solve the pollution on [Anon] Road. You kinda really wanna benchmark what is going on before they start doing something, you know? Because these are essentially residential areas, you know?* (P3)

Most of the starting motivators and the focus for efforts were driven by personal agendas of motivated individuals wanting to establish an understanding of the current situation of the issues in the area. These findings where also confirmed by the FB analysis, which showed how the activities (e.g. monitoring with hand-held monitors) and discussions on FB were connected to *personal* factors of influence.

Environmental and Behavioural Factors

Although a lot of motivations for gathering local data for the group were connected to getting an accurate picture of the situation (e.g. 'establishing a local knowledge base'), data was also seen by the group as a tool for raising awareness. The FB analysis showed how the *behavioural* factors started gaining more importance as time went on and new data became available, which was also confirmed by participants in the focus group reflecting on the process:

> *And actually, to me, [to be] absolutely candid, I think there was sort of another agenda as well, and that all us actually believe in using the city to travel around by means other than private car. And the worsening air pollution problem was seen by us – we're not pleased by it – but it actually gave us an opportunity to talk to people and encourage people to adopt more sustainable means of travel. Because doing anything else was chocking us all! So it was useful – unfortunately – a useful tool to use to get over to people the need to travel more sustainability for their own good and everybody else's.* (P1)

The local data the commissioning toolkit enabled people to collect was seen as a useful tool for not only helping to change social norms but also perhaps nudging the community to change their behaviour:

> *I mean, the thing with, the difficulty with air pollution is that you cannot see it. That is why we need the data and need to think of a way [of] making that accessible to people. And to link it to how they are getting about, and to a possibility of changes in the area and beyond, improving the air quality situation whilst also improving the sort of health and lifestyles in general.* (P2)

Moving from community investigations with the hand-held monitors to getting stationary monitors commissioned and deployed by the UO opened up new possibilities for the group. Instead of having a snapshot of data, the commissioned monitors provided a constant flow of data from a particular place in the community chosen by the people, which made them think about new possibilities to evidence issues with the built infrastructure and try to show cause and effect:

> *[...] sometimes, you see, I think that says it takes, the travel journey is 15 minutes to get through. So, is the pollution worse during the 15 minutes journey time or, you know? Can you draw a link between, you know? And that is all you are trying to do [...] to demonstrate that air pollution is worse when there are more vehicles*

> *standing and/or travelling on that route as the traffic count goes up. And you are trying [to] kind of draw connections to it, you know?* (P3)

In addition, using that data helped to correlate events that were happening in the built environment and to understand the changes in the community and their impacts on people who felt that this additional data could be used to draw attention to how environmental changes influence the data:

> *[A] lot of those events are quite easy to find out about, aren't they? But, I mean, finding the data about that window is quite difficult, isn't it? And it would be good at some point to be able to perhaps, umm, build up a list of things linked to the data. Say that, 'Well, I looked at the data for yesterday and also noticed this was happening,' so that it's recorded for the future, so you can sort of say, 'Well, so was that related to this.' And you know in a few weeks' time, the [Anon] Road might be shut again, and you can say it's the same pattern there.* (P2)

People in the focus group agreed that it is important for the data to be reliable and representing the situation as it actually is: *'But we need to keep getting reliable data – Reliable and trustworthy.'* (P2). Although it was important to have a consistent stream of highly reliable data, the participants also noted that they personally would not be using the high frequency data because they would not be the people who would start engineering solutions and infrastructure changes in the city. Most importantly, the participants wanted reassurance from the engineers at the UO and the researcher that the data is the highest possible quality and the processes of collection and calibration would be transparent to them. Furthermore, people's attachments were also shown from their investments in helping to solve the problems with sensors when they occurred on the first deployment. Although people saw data as a useful tool for evidencing and advocating infrastructure changes, because of the complexities of these lengthy processes, people started to see more benefit in helping to make data accessible for the community:

> *But that seems to me to be giving us clues if we were trying [to] actually engineer a solution. And I don't think we are trying to do that, yeah? I am completely with you in saying we want to see the data, but can we find a way of expressing the data in a way in which anybody would understand and would be meaningful.* (P1)

Furthermore, participants where thinking about how they could put out data in simple digestible forms or as pervasive information for the community to make use of and adjust their behaviour accordingly:

> *I was just thinking. Sorry, from my point of view though asking about the datasets and how easy it is to manipulate them to produce [a] certain sort of information set. It [is] almost like it is trying to, there is a sweet shop of information, and if you don't know what it is, it's very hard to ask for exactly what you want, you know? You are looking at things like the DEFRA daily air quality index. So that's...is a mix isn't it? So daily air quality index, and that takes nitrogen dioxide, sulphur dioxide, ozone, pm2,5 and pm10 and then maps them into a chart. It comes up with a colour,*

> *like purple – don't breathe, green – everything is fine, and everywhere in between, you know? And you wonder if there was a script that pulls – and that is based on a forecast for the next five days – so you know Newcastle is green for the next five days because we have strong winds coming, you know? But you're wondering [...], if there was a computer program that would just get those five variables and then mapping them and then just producing a cell, you know, that was a colour, and say: 'Today's daily air index quality in [the area] is red, green.'* (P3)

They also felt that in order for the community as a whole to start changing their behaviour, there was a need to change the social environment and have the data available and in front of people all the time:

> *And ultimately, it will be that we have an app on our phones that actually gives us, everyday, the colour and those [that are] interested press the button, and you get an explanation [about] what the reading is. I probably said this to you before what cooked me up from all of this was about eight years ago going to China and being appalled about the air pollution. And finding that there were some apps that, still that long ago, in China, that gave you real-time data about the pollution. And now there are hundreds [of apps] in China, and we don't have any really.* (P1)

For participants, the data ideally would have been something that people could check like they check their social media feed. It needed to be embedded into people's everyday busy lives to start having the desired effect:

> *The trouble is that it needs someone to do something that is reasonably polished. You can spend a next year saying to people: 'What do you want, do you want this? ' What you need is something that works fairly reliably and is easy to use. People will start using it and go, 'Uh, look at the air pollution today.' People have busy lives, most people aren't [the] slightest interested in this on a day to day basis because it doesn't visibly, physically affect your day. Um, but neither [does] reading lots of tweets about nothing, but I still do that ((laughing)). If I got an app that distracts me for something, then I will look at it, but I'm not gonna put any thought into how it gets there. It need someone who has access to it and just do it, you know?* (P2)

Additionally, participants saw ways of getting through to different demographics, for example by engaging the local schools to point out these resources to them:

> *Can you point [local] primary school to this data, with some confidence, that you can say to the school that you can now find out what is the pollution outside your school. What you do with that data as a school is up to you. You can't, you know, you can educate the kids to a certain degree. You can try to inform the parents, you know, but they are primary school kids. There is a conversation to be started around air quality, but that's probably, you know.* (P3)

This was also confirmed by the activities of the group that were analysed in the FB data analysis, where the group had meetings with local schools, organisations and businesses to draw their attention to issues and provide them with updates about the data and community investigations. All of this was part of the group's larger plan to start changing the way people looked at the built environment and got around the neighbourhood doing their everyday activities. Here, the participant was offering ideas about how they could present that data to the community, which was later written up by them as a blog post on the group's website:

> *So, if you are able to say what, during an average week, the levels of pollution at this are. And if that period of collecting the data was long enough, then it would actually average out all the little bits like the congestion, the wind, etc. And you would be able to give an average reading (at 9?) of a typical day at any one of our points in [the area]. And I think that would be an incredibly powerful bit of data because if the average level's like say 9 in the morning when kids are walking to school – quarter to 9 or half 8 or whatever – or 4 o'clock, we're actually massively higher than the legal limit or getting close to the legal limit, then that is a really powerful story. I think [we've] got to try and show very simple stories with very complicated data.* (P1)

The findings from the community reflections further illustrate the close relationship of environmental and behavioural factors in advocacy effort of the group. Furthermore, with the point about getting sensors commissioned in the neighbourhood and working out what type of data they provide, and what can be done with it, the group had exhausted the personal factors of influence for the advocacy efforts. Although the personal factors did not completely disappear from the picture, they where still connected (perhaps more loosely) to the other two factors of influence.

3.3.3. Multi-Faceted Approach

By engaging in investigation of local issues and the commissioning process, the group became more aware of their different roles as community advocates and how they had to operate on different levels by having multiple hats on. First, they saw themselves as a group that helped provide local information to the community linked to city-wide issues while not providing solutions or policies, since the council needed to take the lead on that:

> *I think [that] because the city council are coming out with this public relations campaign on dirty air and people are going to be asking questions about 'What's it like for me,' we are then going to be in the position where we can say, 'This is what it actually is for you.' And then the city council – 'and watch this space' – the city council is coming up with proposals [about] how to deal with cleaning the air in the city. And at least then you are on to some information as to how bad it is and why it is important. I don't think we should get too heavily – personally – too heavily into*

> *the policy arena at this stage because I think it's up to the council to actually drive – they got more modelling than we got – we are just people who are concerned.* (P1)

Second, they saw themselves as the 'watchdogs' for changes, keeping an eye on the data to see if the built environment changes from the local government had had a positive or negative effect on the community, in addition to also updating the community if there were any changes that could effect their health or the way they get around:

> *But we need to able to, you know, that gives you some evidence to say the air quality is this. And anything new that is proposed in the future, you can say, 'Is it likely to impact this?' positively or negatively. If something does change, you can say, 'as it done [so].' It is not that [this] takes a day or an hour and it's done. This is a just a start, you know?* (P2)

> *Is this story even simpler than that? Space For Heaton have established a baseline pollution level for some of our streets. And [we'll] be watching to see what will happen with this and informing you as residents, you know, whether it could be that. This is how this comes up, compared with the legal limit.* (P1)

Finally, the participants came to the conclusion that all factors were important and they would keep working on them all; however, the main focus for their advocacy efforts and also the main utility of the data was its usefulness in helping to change people's opinions:

> *Well, I can tell you what my thinking about this is. We've campaigned unsuccessfully [with regard to] changing the built environment, building cycle ways, etc., etc., and okay, things get done, but very slowly, very incrementally. The same time all this is happening, actually things get worse, so the net gain is actually minus. You know we are making, we are not getting anywhere with it. So, the conclusion I come to, and I think a lot of other people [have] come to, is actually you got to start changing people's opinions. So it's a 'hearts and minds' campaign. So I think [about] what we are doing about action here, actually not within a margin. We're gonna continue to take some action to support [the] city council in developing cycle route like up [Anon] Road etc, but actually we should be concentrating on how to get people to understand that there is an issue here that needs action, and we are local people who are helping to articulate that for all.* (P1)

Here, the participant expressed similar concerns that came out of the FB analysis in the previous chapter, where the main factor (i.e. the focus) of influence for the advocacy efforts of the group was *behavioural*. However, all the other factors still remained prevalent and instrumental in aiding the focus of behavioural change of the community. The findings from the FB analysis and the community reflections indicated that there was a need for closer inspection of the relationships between each factor of influence in relation to citizens advocating for change in the community.

3.4. Community Outcomes

Through the ethnographic work, the data obtained from the FB posts and focus group provided a better understanding of the relationships between the key factors of influence, in addition to enabling the author to compile a list of examples of different advocacy efforts and strategies taken by the group using the toolkit and the data it provided, some of which are also linked to events and activities shown on the timeline (Figures 3.1 and 3.3). Table 3.1 lists examples of different activities and actions of the group categorised by the domain of desired change: *built environment*, *policy*, *behaviour*, and *social environment*.

Although the analysis of the case study was carried out through examining the key factors of influence (Figure 3.2), the findings indicate that those factors where also part of the domains of desired changes for which people were advocating. This type of categorisation is similar to the ideas of the CI object system classes framework (Hirschheim et al., 1996; Stillman and Linger, 2009) discussed in Chapter 1, which adopted social action theory to understand ISD. Moreover, using social action theory as an analytical lens is also highly relevant because the aim of the evaluation was to look at how the SMS toolkit and the data could be appropriated by citizens for social action. Table 3.1 shows that there are overlaps in activities for influencing different factors simultaneously to achieve changes in multiple domains. This was also illustrated by the linked nature of factors of influence on the FB timeline graph 3.3. Furthermore, by taking an advocacy and social action lens, it is possible to start looking at how to better design for the effective use of the toolkit and data by communities.

3.5. Discussion

The evaluation process revealed new understandings of how people used the SMS toolkit and the commissioned data. The findings show that the effective use of the toolkit was highly dependent on the actions of self-motivated people from the group; however, it was also heavily influenced by key personal and behavioural and environmental factors. By mapping out the advocacy efforts of the group in relation to community changes over time, new understandings surfaced about how the key factors link to each other and to the domains of desired change in relation to community problem-solving activities. This section presents these understandings through a CAF (Figure 3.4) and discusses the elements linked to the framework through the findings, providing an understanding of how mechanisms may be built for supporting the effective use of commissioning resources in the future.

3.5.1. CAF: Towards Understanding Civic Advocacy

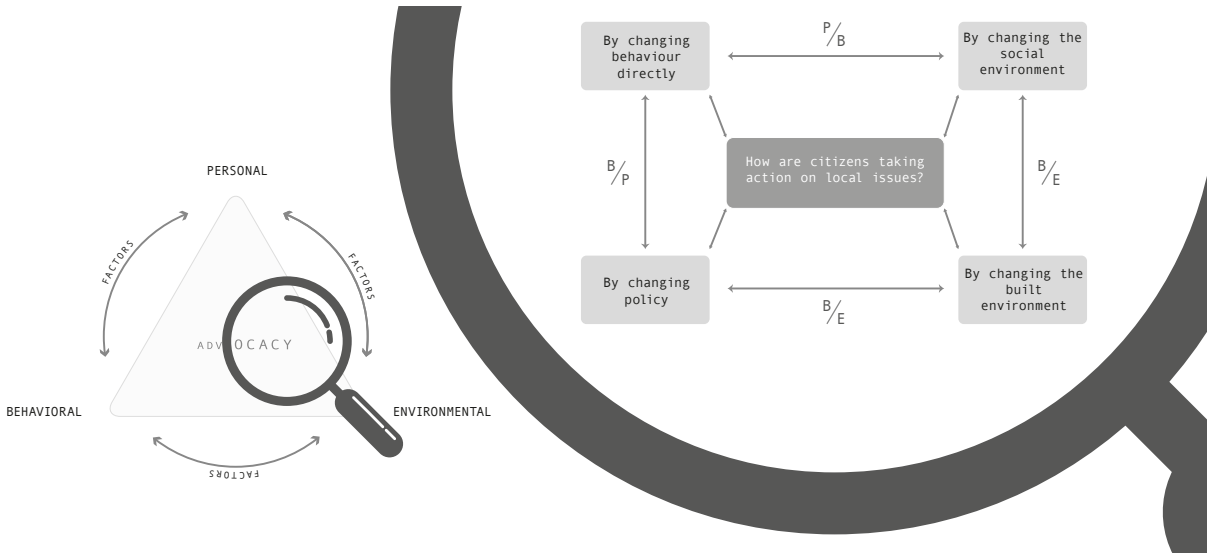
As mentioned previously, the factors of influence (Figure 3.2) were closely linked to one another, wherein focusing on or being affected by one could have an affect on the others. Graphing out the relationships between the factors of influence 3.3, analysing the community reflections 3.3.2 and looking at the community outcomes 3.1 revealed new findings about the ways people used the toolkit for advocacy. A more focused lens was needed to analyse the outcomes of the

Domains of Change	Examples
Built environment	• Combining local knowledge and data for consulting on the development of new schemes for active travel • Using data to submit augmented responses to public consultations around air quality • Using data for activism and drawing attention to built environment issues
Policy	• Using data for activism and drawing attention to issues • Using data for applying pressure on decision-makers • Getting electoral candidates to acknowledge issues and publicly pledge for improvements • Using data to speak up at local governmental meetings
Behaviour	• Helping to raise personal and place-based awareness • Providing easy-to-digest data for people to make decisions that affect their lives – Automatically tweeting daily summaries of commissioned sensor readings – Publishing blog posts and sending newsletters with summaries of analysis of commissioned sensor readings – Providing explanations of proposed infrastructure changes and how they might affect the community – Providing information about alternative ways to get around the area and nudging people to change their behaviour
Social environment	• Raising awareness – Organising public meetings to explain local issues and data – Participating in community events to spark conversation around issues – Automatically tweeting daily summaries of commissioned sensor readings – Publishing blog posts and sending newsletters with summaries of analyses of commissioned readings – Engaging with local businesses and other organisations • Using data for community knowledge creation and for local decision-making • Building community resilience and capacity • Using data to speak up at local public meetings • Engaging schools around the issues and providing support to educate young people

Source: Author

Table 3.1 List of activities using data by the community group categorised with the desired domains of change

advocacy efforts, determine the relationships between the factors of influence, and reflect on and discuss the practices of the advocacy group appropriating the toolkit and the data. Rephrasing the core question of the evaluation from *'How citizens make effective use of the SenseMyStreet toolkit?'* to the more encompassing *'How are citizens taking action on local issues?'* enabled the author to derive more specific categories that are simultaneously liked to multiple factors of influence. Figure 3.4 shows the derived CAF framework, which incorporates the factors of influence but is more catered towards the effective use of data and technologies by citizens for



Source: Author

Figure 3.4 Citizen Advocacy Framework (CAF): *B* – behavioural, *P* – personal, *E* - environmental

advocacy and action and to creating knowledge and positive change in the community. The findings of the study and the outcomes indicate that changing *policy* and the *built environment* is leaning towards influencing physical space and helping to instrument that change, while *social environment* and *behaviour change* have aspects of influencing social norms, attitudes, behavioural trends, self-efficiency and even personal attitudes. Similar to the key factors of influence, the individual parts of the CAF cannot be viewed in isolation. In this sense, they are dynamically interlinked, with each action taken influencing the whole action network. This is exemplified in Figure 3.4 by the different combinations of the factors of influence on the links between the elements, e.g. the correlation between the *behavioural* and *environment* factors shown on the graph. However, when setting a focus for advocating efforts, it is possible to direct the action towards a particular part of the CAF, making that the strongest variable in the action network. Introducing the CAF will not only aid in discussing and reflecting on the processes of SMS and its effective use by citizens but also provide support for setting a focus for the future advocacy efforts of communities. Identifying the key focus points for an advocacy group enables designers to better explore digital technologies that could support citizens taking action on local issues.

3.5.2. Framing Issues, Developing Personal Attachments and Taking Ownership

Researchers working in participatory design, particularly those focused on community-centred projects, have documented the importance of issues and developing attachments to them that help constitute publics (Dewey, 2016) and create spaces for action (Asad and Le Dantec, 2017; Dantec and DiSalvo, 2013; DiSalvo et al., 2008; Jenkins et al., 2016). This work mostly focuses on using participatory design for infrastructuring that helps come up with technical responses to issues of a formed public. In the case of commissioning systems, the technology itself acts as both a technical solution and a socio-technical mechanism for infrastructuring (Chapter 2).

These are systems like App Movement (Garbett et al., 2016) for commissioning community mobile applications, but also SMS, which was designed as community-led urban sensing toolkit that would act as catalysts for grassroots civic action. This is not to say that they have not been designed to work on particular issues and publics. One's first interaction with SMS starts with *identifying an issue* on a map using the GIS mapping functionalities on the project's website. This will set the tone for further activities in terms of framing the issue – what type of issue it is and where it occurs. Furthermore, *place* also has an important role to play in this, as it is an important part of hyper-local advocacy efforts and essential for constituting new publics around issues or expressing and developing the personal attachments of existing publics.

In the case of the advocacy group, a huge driver for them initially was to educate themselves about the issue, which then helped the framing of future actions. Intrinsic motivations for understanding the data were major factors for its use for the advocacy group. Findings from the FB posts reveal that the initial motivations for people were *personal*: figuring out the current policies, how to use the hand-held monitors and how to make sense of the commissioned data. Once the required knowledge was obtained by the group and the personal attitudes formed, they could then start focusing on other strategies linked to advocacy involving the domains of desired change (Table 3.1). When reflecting on the SMS process with the participants at the focus group, and looking at the data produced by the commissioned monitors, participants had a better sense of the figures because they had gone through a process of self-discovery in the first step (i.e. Identify the issues) of the process involving hand-held monitoring. This argument is also supported by the findings of the usage data of the toolkit (Chapter 2 Section 2.6). Having a clear view of what they had (in terms of the current data and its value) and what they would expect to get from the commissioned monitors in the future enabled the group to choose relevant strategies and future actions. Enabling people to imagine the desired futures through the processes linked to the toolkit supported its value to act as a process for infrastructuring (Dantec and DiSalvo, 2013) towards civic advocacy and citizen action.

Furthermore, although there was a handover of responsibilities in the third step (i.e. Get the facts) of the process, the participants were also actively involved in the deployment processes of the commissioned sensors by prompting the researchers for updates. The findings indicate that when the technical difficulties with the monitors on the first deployment emerged, it seemed that the group felt that they were accountable and wanted to see that the issues got resolved. The group worked closely with the research team at the UO, keeping an eye on the readings to detect inconsistencies and get faulty sensors replaced. This also meant that some of the planned advocacy activities were put on hold by the group, and the focus and responsibilities of the group shifted – holding back on presenting the data to the community before it was validated. The ability to adapt to situations and dynamically shift responsibilities, often due to external influences and community changes, and work around problems that might occur have been shown to be strong benefits of infrastructuring projects (Dantec and DiSalvo, 2013). Additionally, as it was their efforts that got the monitors out there, it seemed that they felt that it

was their responsibility to make something out of the data and provide it to the community in a more digestible form.

Research in CI has shown that a community taking ownership of the final outcome of the project is also essential for the sustainability of the initiative (Carroll and Rosson, 2007; Merkel et al., 2004). However, there is a difference between actually taking ownership of the technologies and data (i.e. maintaining and sustaining the actual tools and infrastructure) and having a *sense of ownership* over the investigations and issues (i.e. community initiatives). Analysis of the different groups that engaged with the SMS toolkit shows that when a group takes ownership of community data and develops attachments, it will promote its further use in civic advocacy and action. All communities are different, however, with different issues and attachments, and all need their own approaches; thus, there is a role for infrastructuring and technologies to facilitate the creation of these bespoke methods of action.

3.5.3. Plans of Action

The SMS toolkit successfully helped to carry out a citizen-led deployment of sensors, creating an opportunity for citizens to access resources that were not available to them before. However, that was not the only aim of the toolkit – it also aimed to help citizens put the data into use for activities for local benefit. The analysis of SMS usage suggested that the process of data generation was benefiting the community when it was supported by additional strategies to purpose the data, meaning that data generation had to be connected to a specific goal – either for personal discovery and self-knowledge or for the purposes of advocating for change in the community. Although the commissioning platform was designed as a product that would facilitate infrastructuring, it was still largely up to the community to take it further from data production to informing issues and exploring ways the data could be put into use to instrument change in the community.

The findings revealed that plans for actions often preceded the actual data production by citizens. When participants in the focus group reflected on the process of commissioning and talked about their motivations for getting involved in the project, they pointed out a set of strategies that they had for the data before it was available, strategies that could be tailored and implemented depending on what was possible to get out of the data. However, when it comes to infrastructuring, in addition to attachments being dynamic (Dantec and DiSalvo, 2013), so are the strategies for actions. As shown from engaging participants in commissioning of environmental data, they could shift in focus from *personal* to *behavioural* to *environmental* or the other way around, depending on the affordances of the data. Initial strategies could end up not working at all in the long run; however, having a set of questions, ideas and plans for action gave people a specific goal to work towards. Findings from the FB data analysis also illustrated that planning ahead enabled people to have much more focused advocacy efforts when the data became available. However, based on the analysis of the FB data and community reflections, it seemed that there was a need for an adjacent possible in terms of the data and resources for people to actually be able to take any action. The planning of action was good, but data was

also needed to take those actions. Additionally, there was a clear trend of actions being more focused when there was new data or new information available that the group could share with the community.

Furthermore, there were also events happening in the community that were out of the group's, and the toolikit's, control. These could be positive, such as opportunities to provide opinions and data for public consultations, or negative, such as an opposition group emerging, from the perspective of group advocacy efforts. It often made sense for the group to act accordingly to community changes – directing their efforts or linking to some sort of already ongoing activity or event. Such events are illustrated in Figure 3.3 as *Community Changes* – local and national media interest, local consultation, local elections, city-wide events, etc. Participants reflected that their initial aims were to take advantage of opportunities to influence the polices and built environment (e.g. public meetings and local consultations); however, as time passed, they realised that would take more time because of the official processes in place. Although the group worked closely with the council to help consult on infrastructure changes, they also did not feel that they were in the position to propose solutions for the issues. This was also evident from the findings of the focus group with participants reflecting on the uses of data. They did not feel that the data would help them engineer particular solutions, but they seemed like concerned residents who wanted to expose the issues by expressing them through data in a way that would be understood by all members of the community. In this sense, their focus shifted more to influencing the social environment and coming up with strategies for changing local people's behaviour. This did not mean that other activities stopped completely, which was also illustrated by the connected nature of the environmental and behavioural factors revealed from the FB data analysis and the group's ability to change strategies when there was an opportunity to influence another domain of change, for example when a new consultation was opened up for citizen feedback or when a public event was organised by the local council. The latter event was used by the group as a platform for speaking out about the commissioned data and issues. This illustrated the importance of having capacities and resources (and large social capital) to respond quickly to community changes.

3.5.4. Forming Networks Through Community Links

The importance of publics (Dewey, 1954) in framing the issues should not be underestimated. Across the deployments, it was clear that the existence or creation of publics was essential for the use of commissioned data. Often, people living in one area had similar opinions or concerns on matters, but they were not being vocalised by anyone. Having a group of people that could formulate issues and seek to generate data through the infrastructure provided by the commissioning toolkit to back them up helped to voice issues for the community. In turn, this type of collective representation of the community helped to get more people behind the issues and make them more prominent in society and the eyes of decision-makers.

The author's years of experience in working around data have revealed that 'everyone wants data until you give it to them', meaning that it takes much more effort to make data accessible and understandable. Data science skills and training are required to transform data into a useful

form and into something that the community can use in their advocacy efforts. Rows and rows of sensor readings in a spreadsheet are just not useful for the community. Although there is a huge representation issue with the data, this does not mean it should not be available to the public. Another importance aspect that was discovered in designing and evaluating SMS was the powerful effect of community links. This comes back to building capacity – not only extending the skills and knowledge of the immediate community but also reaching out to others outside the formed public (i.e. social capacity). Establishing links within the community and collaborating with external actors can help alleviate some of these issues with data. Additionally, having the people involved in the processes of data production through infrastructuring enables us to break down barriers in data use by linking it to people's concerns and helping to developing plans for action that could be sustained in the future.

The advocacy groups that were formed though the use of SMS (Chapter 2 and the current chapter) started making links with each other. Although they were all living in separate parts of the city, having their own particular issues and attachments bounded to place, they where exchanging knowledge and resources and learning from each other's experiences of using data as a tool for advocacy, e.g. sharing pledges for electoral candidates to publicly sign or ideas of how to present local data for residents (Table 3.1). The formation of these links was not explicitly facilitated through the commissioning infrastructure; instead, they organically emerged once people started working on strategies for using data. During the focus group, when people were coming up with ideas for future strategies with the data, they often drew examples from more established advocacy groups in the city. As more data about the community was generated though the commissioned monitors, representatives from different advocacy groups started having regular meeting with each other to discuss strategies and develop joint thinking for establishing a combined voice for the people of the whole city, including showing up to community events and council meetings as a combined Safe Pedestrian and Cycling Environment (SPACE) advocacy group. Groups not only shared strategies but also started sharing actual technical solutions developed as a result of the strategies for using the data. As reported in the community reflection, a significant focus for the participants was to bring the issue to public attention and build the knowledge of the community. The proposal of having an automatic system that gives the community daily information about up-to-date readings from commissioned local monitors was later realised by one of the members in the advocacy group. The group set up a Tweetbot that tweets daily about the last seven-day average readings from their local sensor. Later, the group also helped set up a similar Tweetbot for the other advocacy groups.

Links between advocacy groups definitely became a major factor in coming up with strategies for the use of community data. However, links between advocacy groups were not the only connections that prompted the use of data within these groups. In this sense, links with local authorities helped advocacy groups focus their efforts. As outlined in the previous chapter (Section 2.2.3), some of the participants were part of forums (e.g. the Healthy Streets board and Cycling Stakeholder forum) within the local council. The groups often had insider information about the next steps of the local council in terms of upcoming campaigns, consultations, infrastructure

changes and policies, which enabled groups to come up with timely information snippets and explanations for the data and processes for community, including using the data generated by the monitors to compile an informed view of issues. Advocacy groups often posted blogs with explanations of data, linked to things that were happening in the community or in the city at large. This relates back to the insider and *outsider* strategies linked to civic advocacy (Asad and Le Dantec, 2017). In a way, the groups knew their capabilities and limitations regarding what they could achieve with the data, but they were also trying to reach out and link up with external groups and individuals who had knowledge or skills in other domains, whether it was different advocacy groups, professionals in the university or the local authority. A similar situation played out in the study conducted by DiSalvo and Lukens (2009), where the local community, facing regulatory difficulties relating to getting their radio broadcast off the ground, decided to seek the help of professionals who could get it on the air. Although the group increased its reliance on others, it was still attached to the issues and felt responsibility for the community to get the project out there. This echoes the commitments of the advocacy groups that took part in SMS and helped generated data for the community – they felt accountable and wanted to see the data being made visible and accessible to the community. These different ways of creating links and coming up with solutions to accommodate issues could also be looked at through a lens of *collaboration patterns* (Moor, 2009), which emerged from engaging with the commissioning toolkit. According to Moor (2009), there are five different collaboration patterns: *goal*, *information*, *communication*, *task* and *meta*-patterns. SMS has provided practical examples for different patterns of collaboration though the analysis and evaluation of the toolkit, which provides important insight for future research in CI.

For some communities, however, establishing these links was easier than for others, depending on whether they had more social capital or already existing links and connections they could call upon, or sometimes just through luck and timing, e.g. showing up to workshop or neighbourhood forum and meeting people interested in exploring similar issues. There have been successful attempts to use technology to accumulate this social capital. For example, App Movement, a community commissioning platform for mobile technologies, used campaigning as a mechanism to help constitute the public prior to initiating the collaborative design process for the community technology (Garbett et al., 2016). Initially, similar mechanisms were considered for the design of the SMS toolkit. However, there was a concern that the platform would start prioritising someone's issues over others and that the one who had the most 'friends' would get the resources. Instead, the toolkit was made inclusive for everyone who had an issue they wanted to explore, whether that was a group of people or an individual. This proved to be the right move because, in some cases, the investigation was initiated by an individual first, and in time, when there was some data available, more people joined to form a public (Dewey, 1954). This was also supported by the infrastructures of SMS technologies, which enabled voting and commenting on issues on the SenseMyStreet website and collaborating around proposals on the SenseMyStreet Proposals website (Chapter 2). Infrastructuring around SMS still did not provide a guarantee about whether a deployment of monitors from the toolkit was followed by informed

actions from the community. Although the toolkit supported collaborations and the formulation of publics and links though infrastructuring, it was still largely up to the people involved to come up with strategies for acting upon the generated data.

3.6. Conclusions

The first prototype of the toolkit combined multiple tools and digital platforms to enable people to access scientific-grade environmental monitors and use them to explore issues in their neighbourhood. Through the process of engaging people in commissioning the collection of environmental data, a lot of new opportunities for collaboration and community building surfaced. However, multiple challenges also appeared around the readiness of IoT and smart city technology, the representation of data to serve the needs of the community and the community's ability to use data to take local action. Indeed, the current infrastructuring around SMS worked well for groups where additional capacity already existing within their community, whether that was manifested in needed skills or external links to resources or information. However, these links started developing as more groups began to surface and become vocal. Furthermore, the complex configuration of tools and technologies of SMS (Chapter 2) worked as temporary infrastructure because of the constant support and orchestration from the researcher and the UO, which was invisible to the participants. In some regard, SMS stayed true to the vision of Ubicomp (Weiser, 1991), which refers to seamless interactions with computers in the background responding to human interactions and intents. Going forward, there needs to be consideration about the way these tools can be transferred or taken over from the researcher to make them sustainable in the future. However, the value of this case study was not only to show how to configure and build tools for community commissioning but also to illustrate how the smart city could be made more accessible and usable by communities though human-centred design and active citizen engagements. Future reconfigurations of the toolkit may consider building in mechanisms for information exchange with the smart city and making it easier to establish these links. This includes not only using technology to help connect different groups, but also developing better links between various platforms (including the ones already existing in the smart city outside the toolkit) to connect people in one community so as to form publics (Dewey, 1954). With this in mind, future developments of community commissioning and CI systems should consider not only providing resources and access to collect data, but also working more on facilitation and training to enable communities to make effective use of those resources. These ideas will be further explored in Chapter 5.4, which provides a combined reflection and analysis of the two case studies (SMS in the current chapter and Chapter 2, and Data:In Place in Chapter 4) from the perspective of the effective use of data by communities for civic participation, advocacy and action.

3.7. Summary

This chapter provided an evaluation of SMS from the perspective of the effective use of commissioning resources by communities, which involved looking at a particular group's engagement (Chapter 2 Section 2.2.3) with the toolkit and how they appropriated the data produced through the commissioning process. Evaluation was carried out in two parts: (1) chronological analysing of the textual content of a FB group using key factors of influence (i.e. environmental, personal, and behavioural) and (2) analysing transcribed audio from a focus group interview with key group members. Additionally, the activities of the community using the data were presented through *community outcomes*. As a result, a CAF was posited, which enables researchers to discuss activities of civic advocacy and action and helps communities plan and focus their activities for change for the local benefit. From the initial launch of the toolkit, there was a great deal of interest from people to get involved; in this regard, people seemed to care about their streets and wanted to investigate issues in their communities. Having infrastructuring and infrastructure for people to use provided new forms of active citizen participation and civic action driven by communities. Enabling citizens to help provide data as part of the smart city for the community could help inform issues and also challenge uninformed decision-making in the neighbourhood. However, the case study findings also highlighted the importance of community links and social capital for making effective use of commissioning resources. Despite this, having evidence to point to that informs an issue in the community has given people a voice to advocate for change and helped more people become vocal. While this case study mainly focused on the issue of air quality, framed by the motivations and nature of community research, there is no reason to assume that the issues will remain static. New areas of concern may rise from the community in the future that could be explored and evidenced through environmental sensing and the use of SMS.

Chapter 4. Data:In Place: Making Open Data Work for Communities

Information Wants to Be Free. Information also wants to be expensive...That tension will not go away.
— Stewart Brand

This chapter presents Case Study II and describes the iterative co-design and development of the *Data:In Place* platform, which aimed to democratise data access and use by communities. The value of data in supporting citizen participation in processes of place-making and community-building is widely recognised. While the open data movement now permits citizens to acquire governmental data relating to their communities, little to no effort is made to ensure that these datasets are accessible and interpretable by non-professionals. A series of community engagements spanning an 18-month period resulted in the Data:In Place platform, an open-source platform that supports citizens in accessing, interpreting and making sense of open data for the purpose of civic advocacy. Leveraging visual map-based querying, citizens could access official statistics about their community, interrogate the data, and map their own data sources to create data visualisations. Throughout the co-design process with local communities and charity organisations, the author acted as a resident data scientists, collaborating in and supporting their community research with technical expertise. This served as a means to understand what the communities' data needs were and what sort of platform would be of value to them. As new barriers and challenges to the engagement of non-professionals with data science were uncovered, they were responded to through the co-design and development of new features and functionalities for the platform to support the access and use of these techniques (Section 4.6). The platform was designed through in-depth collaboration in two community contexts: (1) a neighbourhood-planning group relying on open data to inform the process of policy-making and (2) two charity organisations reliant upon data to focus their work on the needs of the community and to illustrate their impact so as to extend their funding. Reflecting on the participatory design process and the designed technology, this chapter provides a framework to make open data work for civic advocacy. It is important to note that Case Study II started before Case Study I (Chapter 2) in the actual research timeline and then ran alongside it (see Thesis Map Figure 2 in Introduction). The case studies are presented differently in this thesis for the purpose of a clearer narrative.

Related Publications and Acknowledgements

- The structure of this chapter is based on Puussaar et al. (2018), a publication authored by the PhD candidate. The analysis and writing of this paper was done by the PhD candidate. It has been reworked by the PhD candidate to fit the overall narrative of the thesis.
- The chapter also borrows ideas from the conference paper Johnson et al. (2018). In collaboration with Ian Johnson and the PhD candidate the research design was set up, prepared and conducted with the help of Jennifer Manuel. Dr. Ian Johnson performed the main part of the data analysis which was written up for the conference paper co-authored by the PhD candidate.
- Acknowledgements go to Dr. Ian Johnson and Jennifer Manuel for helping to set up the study design, data collection and helping to conduct the engagement work.

4.1. Motivations

From life-logging to smart-city management, data is at the centre of our research, communities and world. It is estimated that we create 2.5 quintillion bytes of data daily[1], which means that 90 percent of the data in the world today has been generated in the last two years alone, and with emergent ubiquitous technologies, the data growth rate will likely accelerate substantially. While the technology to generate and store such vast quantities of data is ubiquitous, what remains a significant challenge is how to use that data for the common good. For many years now, private sector industries have been mining big data from different sources in order to understand product demand, forecast business trends and identify potential business risks. In the public sector, there is also a growing emphasis on data-driven, evidence-based policy-making. Different government initiatives are emerging that bring together teams of people to address the challenges of society with data science. For example, rather than relying on surveys and the traditional polling of citizens, institutions like Data Science Campus[2] in the UK try to find new data sources and ways of providing early statistical indicators. The development of the open data movement has begun to overcome the first obstacle to using data for the common good, i.e. enabling people other than those who generated it to access and use it. Currently, however, the effective use of big data (Gurstein, 2011) requires the professional skills of the data scientist, which include being able to access, interpret, and make sense of data in ways that make it valuable, actionable information for others (Gitelman, 2013); consequently, data science skills are in high demand[3].

Within HCI and CSCW research, the rush to big data has unearthed a number of unanswered questions: Where do citizens fit in this data saturated world? Is there a role for them beyond that of simply being the source of data feeding 'smart' algorithms? How do and how could citizens use big data for their own civic and community purposes. Is there a role for 'citizen science' and citizen-generated data in smart city management? How could big data be made accessible and usable by citizens? How could big data be used to inform civic decision-making at the level of local communities and local area planning? And how could all of this be done without the need of a professional data analyst to act as intermediary? To start exploring some of these questions, ubiquitous technologies are enabling more HCI and CSCW research to be conducted outside laboratories with a variety of communities in different contexts (Carroll and Rosson, 2003; Golsteijn et al., 2016; Lindley et al., 2017; Rogers, 2011; Vlachokyriakos et al., 2014). Often, researchers design tools and platforms for data collection, exploration and use, which enable institutions and people to engage with datasets around complex socio-technological issues. A variety of data-driven technologies are being applied in different civic contexts, from smart cities (Kitchin et al., 2015), personal data management (Puussaar et al., 2017), community data hubs (Crabtree et al., 2017), deliberative democracy (Johnson et al., 2017), and citizen science (Balestrini et al., 2015; Gallacher et al., 2014). Many of these systems are designed to elicit participation and collaboration and provide a common data resource (Asad et al., 2017;

[1] http://ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN
[2] http://datasciencecampus.ons.gov.uk
[3] http://blog.linkedin.com/2017/december/7/the-fastest-growing-jobs-in-the-u-s-based-on-linkedin-data

Balestrini et al., 2017) for institutions and communities to use and repurpose. However, this can create an unbalanced relationship with regard to professionals using the data to make decisions on behalf of citizens (Van Dijck, 2014), while citizens lack the skills, knowledge and appropriate tools to use the same data in civic advocacy. Furthermore, this gives rise to significant tensions relating to the multiple ways in which these datasets are made legible and interpreted.

The literature review (Chapter 1) in this thesis provided an in-depth overview of how data is being generated in the city and by whom, how it is made available or 'open', who is involved in it, what the potential benefits of it are, and who can access it. What was revealed was that making data available is not the same as making it usable for everybody. Furthermore, there should be a distinction between those who collect the data, those who need to use it (and for what purpose), and those who can access it. When citizens do not have the skills to access or make sense of data and to use it for effective action, it is difficult to claim that it is 'open data' (Gurstein, 2011). Despite aiming to make open data more accessible to communities, many of the existing solutions and tools are often linked to static datasets, are not freely available or lack the functionality necessary to help citizens understand and find value in the data.

In designing platforms to support communities and institutions working together with data, there are significant challenges and questions relating to how data may be relevant and for whom and how to translate and represent it in ways that make it accessible and effective in supporting civic action. There are also challenges and questions as to how different datasets may be combined to better support the communication, collaborative interpretation and use of complex datasets that straddle the boundaries between data that is private, shared or public. A crucial role for the professional data analyst is to understand complex community contexts so that they can identify useful data and help communities make sense of data in ways that are relevant and actionable within their contexts. In HCI and CSCW, researchers have stressed the importance of providing additional contexts to support the process of making sense of data (Boehner and DiSalvo, 2016; DiSalvo and Lukens, 2009), e.g. contextualising data in interpersonal 'worlds' by focusing on the nuances and implications of understanding data as bound to places and communities (Lindley et al., 2017; Taylor et al., 2015). A key challenge in this regard is to develop ways to support individuals and communities in finding data that is relevant to the issues and concerns that are intimately tied to the community's construction of place – the locales, neighbourhoods and regions in which citizens live and their values and futures are invested. The challenge then is understanding what is involved in supporting citizens to make data relevant to place in order to enable effective action in relation to place-based issues and concerns.

4.2. The State of Open Data for Civic Advocacy

This section extends the thesis' literature review (Chapter 1) and *The State of Open Data* presented by Davies et al. (2019) to focus more on the ways that OGD is made available for non-professionals and citizens to use in advocacy and activities for local benefit.

4.2.1. Effective Use of Open Data

In addition to the issues around access and limitations with existing tools, there are a number of other barriers to the *'effective use'* (Gurstein, 2003, 2011) of data within communities: *resources*, *digital literacy*, *competencies*, and *trust and ownership*. Gurstein (2011) points out that we should not confuse access with the use of open data. As described in the literature review, there has been a great deal of effort to make data available for whoever is interested in it, but little has been done towards making data accessible and usable for non-professionals, especially those who are deprived and marginalised. Gurstein (2011) is critical of Tim Berners-Lee's statement that open data empowers 'everyone', because the fact is that it only empowers those who have access to resources and the knowledge to make use of it. He suggests that we should adopt the concept of effective use to distinguish *access* from *accessibility* and *use*, in which case, we should also distinguish *open data* from *accessible data*. In this regard, the former refers to the classical definition of OGD[4], whereas the latter would truly make data open and accessible because simply having access to these different raw datasets does not mean better civic engagement and participation from citizens (Davies and Frank, 2013). In his article, Gurstein (2011) puts forward a seven-layer model to address the *'data divide'* (Gurstein, 2011, p. 3), the layers for which are: (1) internet; (2) computers and software; (3) computer/software skills; (4) content and formatting; (5) interpretation/sensemaking; (6) advocacy; and (7) governance. While the model of Gurstein (2011) adequately captures the challenges of addressing the data divide, Hakken (2003) also argues that these barriers cannot be overcome by technology alone and that human actors play a vital role. In this regard, information and knowledge does not magically appear from data but needs to be constructed by humans and not computers. Furthermore, expert systems of data and information retrieval do not adequately deal with the social and cultural dynamics at the centre of the knowledge networking (Hakken, 2003).

The idea of *'democratising data science'* has emerged from research that calls for new tools and approaches to leverage data science for social good (Choi and Tausczik, 2017; Chou et al., 2014). Choi and Tausczik (2017) investigated collaborations around these kinds of open data projects and found that they are small in scale, mainly involving civic hackers and data journalists (Parasie and Dagiral, 2013). Such projects aim to address problems for the social good; however, it is clear that more could be done to include the community in this knowledge creation. Although there have been efforts to link data experts with NPOs to help them achieve their goals (Erete et al., 2016), this is likely to produce a reliance on experts to provide constant support. In addition, the exclusion of actual problem owners (i.e. the people on the ground) from these processes is worrisome as it can mean that experts and scientists decide which problems to tackle and how to interpret the data. More should be done by creators of technologies that work with data to reach groups with low or non-existent data science capacities (Chou et al., 2014).

Taylor et al. (2015) posit that data should be understood as intrinsically linked to a sense of 'place' – arguing that data should be kept in the context it was collected and questioned to produce better understanding and use. Nevertheless, with open data, it is already collected and taken away

[4] http://opengovdata.org

from those places by outside players, and the people living there do not have any say about how the data has been chosen, collected and interpreted. These datasets often lack a representation of place that is essential to contextualise and situate them within the community and the activities they engage in, the concerns and issues they face and the situations they want to change. Only when such aspects are better understood can we turn our attention to understanding how people reconstruct new meaning and knowledge from these already engineered datasets. Work in this area has reported on *sensemaking* (Furnas and Russell, 2005) with personal informatics (Puussaar et al., 2017) and with domestic data in the workplace (Clear et al., 2017). In both these studies, people created meaning from the data through 'comparison' and linking it to 'real life experiences'. Despite such work showing promise for better engagement with data and an understanding of what data is and what it means, none of the studies have confronted the problem of the effective use of data by communities for civic advocacy and action.

4.2.2. Integrating Open Data with Citizen-Generated Data

Data about citizens and communities is collected and used by institutions for specific purposes. Releasing such data as open data is a useful by-product of those processes, one which can potentially make the processes and purposes more transparent. However, for citizens to be able to challenge the data or processes, they have to be able to talk back to the data either through comments or opinions or with data of their own. In theory, the availability of open-source software and hardware tools (Chapter 2) enables people to take democratic matters into their own hands (Asad et al., 2017; Balestrini et al., 2015, 2017).

However, data does not always have to refer to something that can simply be counted and represented as quantifiable values (e.g. 'big data') (McMillan et al., 2016; Michael and Lupton, 2016); instead, it can be a much broader and diverse set of 'things', often characterised as contemporary media such as 3D printed physical fabrications of data (Nissen and Bowers, 2015); travelling suitcases filled with stories and memories (Crivellaro et al., 2016); and interactive physical graphs (Sweeney et al., 2019). A good way to engage people with data is often through installations and art practices that use technology coupled with engineering practices to place data under a critical lens. Examples of these kind of projects are Take a Bullet for the City[5], which uses data from the New Orleans Police Department to fire a blank shot every time a report of a firearm discharge occurs, and Invisible Airs[6], which uses different 'machines' to visualise the public spending of Bristol City Council. In many cases, this way of 'releasing data' is far more effective than a flat machine-readable file on some data portal and gets people more engaged with the issue at hand. In addition to the benefits of being able to combine digital and non-digital tools to visualise data, situated urban displays also enable people to keep the data in the place it was collected from and with the community it belongs to, for example, using chalk and tape to visualise locally collected data to raise awareness around local issues and promote dialogue between residents (Koeman et al., 2014) or using public spaces to display

[5] http://www.digiart21.org/art/take-a-bullet-for-the-city
[6] http://yoha.co.uk/invisible

private household data to promote behavioural change around energy usage (Bird and Rogers, 2010; Moere et al., 2011). However, it is known that the way in which different data comes to be interpreted is always entangled with and dependent upon both contexts and situations (Suchman, 1985; Taylor et al., 2015). While there is a tendency for civic technology projects to place emphasis on quantitative data – on the assumption that quantitative data may be relevant to people – qualitative data describing things that cannot be easily quantified or measured is not less important. Indeed, it is often this type of 'data' that helps people make sense of what matters and helps generate actionable knowledge.

However, despite available tools showing promise for better engagement, very few opportunities exist for less privileged communities to collect their own data and have an impact on decisions that affect their lives and those of their families and neighbours (Colomb, 2017). Moreover, it is often difficult to identify what data is relevant, worth collecting for the community or even possible to quantify. We can apply the idea that *'not everything that can be counted counts, and not everything that counts can be counted'* (Cameron William Bruce, 1963, p. 1-14) when taking on data-driven decision-making. There are several opportunities for open data to have an impact on the lives of citizens, and even empower them, but this requires citizens with skills, literacy and tools to access, interpret and contribute their own data to existing datasets. Part of the potential is around establishing a more holistic picture of the community for the smart city (Townsend, 2013). Making data accessible and relevant for people also has the potential to transform it into a resource that encourages debate and deliberation and can open up new ways of participating (Balestrini et al., 2015; Hansen et al., 2014).

4.2.3. Existing Open Data Platforms for Non-Professionals

Platform	Datasets	Access	Contextualising	Expandability	Openness
UK Census data	Static Census 2011	Search datasets by postcode or drill down from country level	Displays official boundaries on the map and a fixed set of charts and graphs	Not possible	Free to use, proprietary
Know My Neighbourhood	Fixed set of different socioeconomic datasets	Web and PDF reports based on requested area (e.g. user provided boundary)	Displays official boundaries on a static map. Fixed set of charts and graphs	Not possible	Commercial, proprietary
Community/ Local Insight	896 different socioeconomic datasets (up-to date)	Web application for exploring and PDF reports	Enables drawing areas on a dynamic map. Fixed set of different visualisations	Adding own organisational data via file with postcodes	Commercial, proprietary

Source: Author

Table 4.1 Comparison of open data platforms for local communities and non-professionals

The popularity of open data and the availability of different data sources have stimulated the development of a variety of platforms to make use of these datasets. The literature review in Chapter 1 outlined the ways that data is made available or 'open' for others to make use of. In addition to 'official' specialist data portals, there are also multiple tools that try to make that data more accessible. This section will detail some of the most popular free and commercial tools used by communities and non-professionals (based on our engagements with different communities) to make use of open data. Similar to data portals, this analysis focuses on tools

intended for exploration of multiple datasets. Hence, tools such as Public Health England's Fingertips Tool[7] and open data demonstrators like the Indexes of Multiple Deprivation (IMD) Explorer[8] are excluded because they are meant to help people interact with a specific dataset, illustrating the mean of production of that data. A broad assessment of existing open data platforms was made based on the following criteria (Table 4.1): the ease of *access* to work with the data; the support provided for *contextualising* the data with communities or issues; the *expandability* of the available data for working with; and the *openness* of tools concerning paywalls and collaboration.

The major drawback of the existing platforms is their lack of expandability to support the incorporation of new datasets. The *de facto* standard is to provide a fixed and curated set of open data, meaning that new data captured within communities would not easily be contextualised with the national statistics and datasets. One platform that did allow users to upload their own datasets – *Local Insight* – had specific requirements on the formatting of the data, and that data was only available to the organisation that uploaded it, making collaboration with others in the community challenging. Moreover, as this platform is intended for commercial use, carrying large subscription fees, it excludes marginalised communities. In addition to limiting the availability of datasets to work with, the existing platforms are restrictive in their forms of access to data interrogation and query manipulation, opting instead to provide sets of predefined PDF reports. Though informative and well presented, the reports are merely a curated representation of the statistics as seen by the organisation that published them, with no support to inspect or contest the contents. For a community to take action on matters that affect them, they need tools that support accessing, interpreting and making sense of open data. While the existing platforms can provide support for these tasks, they fall short of supporting non-professionals through the entire process.

4.3. Study Design

Given the agenda outlined, there was a clear need to work with local communities to co-design more inclusive processes and digital tools that would not only enable the community to access open data but also help to democratise the practice of data science itself. According to existing literature and studies, it was evident that there was great potential in using data as a tool for civic advocacy. However, there were still barriers and a lack of available digital tools to access data and help make it meaningful and useful for people. Before jumping into civic advocacy and action, there was a need to take a couple of steps back and address the issues surrounding *access*, *interpretation* and *sensemaking*. Based on the literature review in Chapter 1 and the *Motivations* and *The State of Open Data For Civic Advocacy* in this chapter, Case Study II started by posting that data science practices and digital tools need to be framed by the following design principles:

1. Place should be used as a way to make open data relevant to communities.

[7] http://fingertips.phe.org.uk
[8] http://dclgapps.communities.gov.uk/imd/idmap.html

2. Communities need to be able to make sense of and challenge open data in relation to their own experiences and their own data.

3. Local knowledge should be used to enrich and promote collaborative exploration and evidencing of issues for civic action.

4.3.1. Study Approach

To investigate the challenges for using open data in more depth and validate the principles, the author sought out people or groups who could benefit from having easier access to data and the ability to use it in their endeavours. The case study took a PADRE approach (Haj-Bolouri et al., 2015, 2016) to identify the needs and uses for data within communities and built a system that would respond to these needs, taking account of local practices and cultures. The process of building the system was not a one-off: design → develop → deploy. The platform evolved through an iterative process, with different stakeholders and potential users of the system over a period of one-and-a-half years. The system was co-designed through two-to-three week design cycles, during which features and functionality to the system were added, based on the input and participation of the co-designers from the community organisations who were the collaborators. The data used to report on the case study came from the researchers field notes and observations at community meetings, focus groups, interviews and feedback sessions with different stakeholders – two local government workers (LG), six charity workers (CW) and eight community volunteers (CV) from a range of roles. The interviews, focus groups and feedback sessions were transcribed and analysed to guide the ongoing design of the platform.

4.3.2. Study Context

In December 2016, the researcher started working with a group of residents and members of local charity organisations (Charity B) who were interested in starting a neighbourhood plan for their local area. A neighbourhood plan[9] is one of the processes put in place by the Localism Act (2011) in the UK that devolves power away from central and local government directly to citizens. The local policy process enables communities to write their own statutory planning policy that will be added as a part of the local planning policy (Gallent and Robinson, 2012; Stanton, 2014). As part of the local planning process, the group identified the need for data to provide evidence to inform the creation of their local planning policy. During this time, the researcher also started collaborating with a local charity organisation (Charity A) that focuses on the health and wellbeing of young people by providing opportunities for physical activity, training, coaching and education. Charity A was interested in using insights from data to focus their interventions in areas of the city that could be supported by the external funding sources they acquired, as well as providing other evidence of the impact of their interventions, again related to future funding applications and renewals.

[9] http://www.legislation.gov.uk/ukpga/2017/20/contents/enacted

4.4. Co-designing for Accessible Data

The following sections describe the chronological and in-depth process that informed the final design of Data:In Place, an open-source place-based data exploration system. The online platform brings together data from official governmental sources with data generated by communities and citizens and links it together around *place*. The design aimed to help non-professionals access, interpret and make sense of data for potential uses in local decision-making and civic advocacy.

4.4.1. Accessing Data

Both groups approached the research team[10] with a request to support them in accessing open datasets to support their work. The neighbourhood planning group – made up of various people from community, voluntary and charity sector organisations (Charity B) and local residents – had collated some data on the defined neighbourhood plan area from the local government website in the form of 'ward profiles' (i.e. political boundary). This was a document with collated statistics from various governmental departments (e.g. age, population and employment) represented in tables and graphs. Charity A had its own administrative data, which included some basic demographic information on the clients of their services and some census data for the city. Before people could start to think about what kind of data they would need to evidence their issues, they needed to get an idea of what was already out there. To get an idea of what sort of data they would need, the researcher started attending community meetings where neighbourhood plan priorities were being discussed. During community meetings, residents and people working in the area marked down the priority list of local issues or important areas they wanted to focus on, which were categorised under themes: housing, anti-social behaviour, local environment concerns, open spaces and play areas. The concerns and proposals were local for these people and were things that they thought must be prioritised. This list was taken as a basis to seek out the official data published by the government and other institutions relating to it. For example, anti-social behaviour was evidenced from local police data, whereas data about housing ownership from the Office of National Statistics was sourced for housing. Although this data was already in the 'ward profile', it could not be interrogated at the neighbourhood level. These datasets consisted of different types of data – static documents and reports, machine readable files, links to APIs, etc. All these datasets were then presented at the next community meeting (Figure 4.1).

Although people were excited by the idea of getting their hands on the data that governmental structures use to make decisions in the area, they found it difficult to grasp exactly which areas and what statistics they represented. This meant that people were relying on the research team to access, make sense of and interpret these datasets. Furthermore, people started asking additional questions about the data, which meant going back to the original data sources again with new query parameters. The questions this exercise raised necessitated reinterrogation of the original datasets with new query parameters. This illustrated the issues with static datasets

[10]The research team consisted of Dr Ian Johnson, Jennifer Manuel and the author. All worked together to attend the neighbourhood planning meetings and conducted interviews to understand community data needs. However, the author was solely involved with the work related to Charity A.

Figure 4.1 Compiled set of official statistics about the neighbourhood

and the difficulties of data access for non-professionals and led to the development of the first functionality of Data:In Place – to bring all of this open data from different governmental departments together into one place (Section 4.6.2).

4.4.2. Adding Relevant Data

Aggregating all of the datasets requested by the community into a single system they could access was not enough, however, as much of the data lacked relevance to their specific local community. The forms that data was captured, stored and presented in by institutional structures did not resonate with people. In this sense, people do not see places as numerical codes in a dataset; they associate them with what they see when they are going about their everyday lives. When the data was presented to one of Charity B's workers (CW2) involved in the planning group with the data they had requested, concerns about the lack of context was raised: *'Should be street names or postcodes for people to understand where the data is coming from.'* To make this data relevant to people, there was a need to start with things that were familiar to them. Local government worker LG1 here points out the different ways that neighbourhoods are constructed:

> *Everybody defines their neighbourhood slightly differently. From a council perspective, we're often looking at lower super output areas. We know that's not a...we know people don't have defined neighbourhoods necessarily. Some people would define their neighbourhood literally as a few houses, others as a few streets, [and] others as more of an estate, um, type area. But it's information or stuff that you can*

> *kind of drill down to at a very local level, and things which are important to people at a very local level.*

Based on these ideas, a redesigned of the platform was done to enable the starting point for exploring these datasets to be *place*. Furthermore, geographic reference (i.e. geocoding) was used to automatically link datasets together for people to start exploring in relation to the places they live and/or work (Section 4.6.1). The simple design principle of these features was to improve accessibility of OGD for communities and put it into context with the issues that were important and relevant to them, which meant that data was not only brought together from relevant datasets but also bounded it into place. Further available datasets were added in the system as a direct response to the interest of groups. However, every time a session with the tool was run with a group, new potential datasets of interest were identified. There was a plethora of data available that could be potentially added to the platform, but it might not have been valuable to add all of them and overload people with choice. Initially, data was added by asking people what they did and did not deem useful in terms of what they could relate to in regard to the issues they were trying to evidence or tackle. Again, however, this resulted in a flurry of requests from the collaborators, and therefore a lot of time was spent going back to people with different datasets in order to validate the usefulness of those datasets, which meant that the groups would always be relying on the researcher's availability to access data and the capacity to provide them with data in a timely manner. Instead, a 'request data' feature was added, which enabled people to express a desire or add an idea of what kind of data they required (Section 4.6.4).

4.4.3. Additional Administrative Data

The functionality of adding new open datasets to the platform quickly prompted groups to request their own organisational data to be added to the system. They were keen to compare their own data and interpretations against the open datasets they had access to through the system. For the charity organisations operating in the area, it provided a way to evidence their work. CW3, for example, explained that it would help shape their group's service provision, ensuring they were using their public funding to deliver the services where they were needed most:

> *Our own data is really important, but also the publicly available data we are looking at here is really useful for us for shaping where we want to go and where we want to work. So, for example [...] what we would like to be able to do is map out where those participants of ANON Programme are coming from. And map that onto, like, indexes of multiple deprivation so that we can see that we are actually in there [...], if we say we are working with the most deprived communities, can we prove it?*

They also discussed how this would help ensure the ongoing funding support for the charity by creating evidence of their intervention work:

> *So, it's all sort of very well for the founders. We can say the right things, but for ourselves and our growth and being able to really demonstrate publicly, this sort of*

tool would be really useful for us to be able to say, 'Look, these are the areas in the ANON which we cover' and we are covering all these more deprived areas.

For the neighbourhood planning group, it enabled them to contest some of the official data by adding some of their own data that they collected themselves from the community. For example, they were interested in seeing how people's opinions about where they lived and their feelings of wellbeing mapped against official statistics for crime and employment. There were multiple requests from people wanting to add their data into the system, and often the data was sent in different formats. This prompted implementing the ability for people to upload their data to the system and have it mapped alongside the official data (Section 4.6.3). The ability to easily map and compare their own data with official statistics really brought meaning to the official datasets by putting it into the contexts people were in.

4.4.4. Adding New Data

By this point in the co-design process, the system had added functionality that enabled *access* to open datasets for people to see the data in the *context of their community* and to compare and contrast other open datasets with their *own administrative data*. Issues started arising around what these datasets represented. People were worried that what they were seeing did not reflect the actual issues and concerns of the neighbourhood. They felt they could not find the answers to the community's questions from the data. As such, there was a need to 'create new data' that provided a more accurate picture of their experiences of the community. CW4 pointed this out in terms of the negative and restrictive perception associated with the data: *[...]'they're at the lowest 10% of blah de blah' and this, that and the other and 'they're such a deprived community,' and I think it might be actually quite nice to present some positive data to them, you know? I don't know if there is any, but maybe there is.*

The aim was to incorporate into the platform the ability for residents to capture and upload locally relevant data. CW3 highlighted resistance to a version of the community that was reduced to numbers: *'I mean, I think a lot of people would think of data and think of it purely as numeric, but I think of data as being any kind of information that you've gathered to understand a situation.'*. For this, it was envisioned as utilising photographs, sensor data from citizen science kits, online comments and viewpoints and voiced concerns of citizens. This also posed a challenge of how to get people's voiced concerns into the platform. In order to do this, the research team started recording people's opinions about places and concerns in community workshops, specifically designed to capture opinions about place that could be pinpointed to locations as short audio recordings on the map within the platform (Section 4.6.5).

Throughout the co-design process, the system was iteratively developed in response to research collaborators' requests and questions. As each new feature was added, it was interesting to see how our collaborators' relationship with the data changed. At first, they became more curious about what they could access and then experimental as they asked for an increasing amount of data to compare. Later in the process, they wanted more control and personalisation of

data representation before finally moving towards a level of criticality toward data. The following section oulines how this process informed the design of the system.

4.5. Data:In Place Implementation

The first prototype of the Data:In Place platform consisted of a web-based application written in Angular.js leveraging RawGraphs (Mauri et al., 2017), a visualisation framework built using D3.js[11] to aid people in data visualisations (based on Angular.js), a RESTful API, and the Geo Coder for translating map queries and communicating with the PostgreSQL database (Figure 4.2). In addition to extending RawGraphs' visualisation functions, the web application also consisted of an audio player for exploring opinion data (built using WaveSurfer.js[12]) and a GIS component using open-source mapping framework Leaflet[13]. Data:In Place integrated open standards such as Geographic JavaScript Object Notation (GeoJSON), Web Map Service (WMS), and Web Feature Service (WFS) for further expansion of the mapping capabilities, e.g. adding different base maps or extending geographical manipulation tools.

RAWGraphs (Mauri et al., 2017) was intended to be a front-end-only solution without any server-side logic; for the Data:In Place platform, however, the Accessible Data API and the Geo Coder were implemented as server side logic (using Node.js) to interface with official governmental data APIs and handle geographic queries. Furthermore, Geo Coder was implemented with a *data type parser* that identifies geographical data (e.g. longitude/latitude data, GeoJSON, postcodes, and Office for National Statistics (ONS) codes[14]). Similar functionality can be found in current state-of-the art online mapping tools, such as Carto[15]. However, the Data:In Place platform tries to automatically understand the data by looking at the headers and the content without any human intervention, mapping it on the go. This was carried out through implementing algorithms in regular expression language[16] to identify and match patterns and the GeoJSON[17] parser to enable automatic georeferencing.

The existing functionality of RawGraphs (Mauri et al., 2017) had three ways to load data into the platform: copy and paste plain text from any source, upload a machine-readable file (e.g. CSV, TSV, XLSX, and JSON) or request data from a provided API endpoint URL. In addition, it had a section for loading sample datasets to demonstrate the capabilities of the visualising framework. On the Data:In Place platform, the sample dataset sections were replaced with a list of official datasets that people identified as being important (Section 4.4.1). However, these datasets were not static but were linked to real-time API endpoint URLs that were compiled (using the Accessible Data API) to retrieve data about the specific place defined with the map-based query (using Geo Coder). This meant that the platform was always up to date with the latest official datasets.

[11] https://d3js.org
[12] https://wavesurfer-js.org
[13] http://leafletjs.com
[14] https://www.ons.gov.uk/methodology/geography
[15] http://carto.com
[16] https://regexr.com/
[17] http://geojson.org

Multiple third-party services were also used to support the functionalities of the Data:In Place platform: Google Analytics to gain insight into the platform usage statistics; a map tiling service MapBox[18], using data from OpenStreetMap[19]; and ONS for getting official map boundary data for the UK. In addition, Ambit (Johnson et al., 2020), an augmented reality audio-capture tool built in Open Lab at Newcastle University (based on ARToolKit[20]), was used in Community Conversational (CC) (Johnson et al., 2017) activity to help collect and structure conversational-based data (i.e. 'neighbourhood data') at community workshops (described in Johnson et al. (2018)). Data from the Ambit system was then geocoded using Geo Coder and pulled into Data:In Place as an additional source of qualitative data.

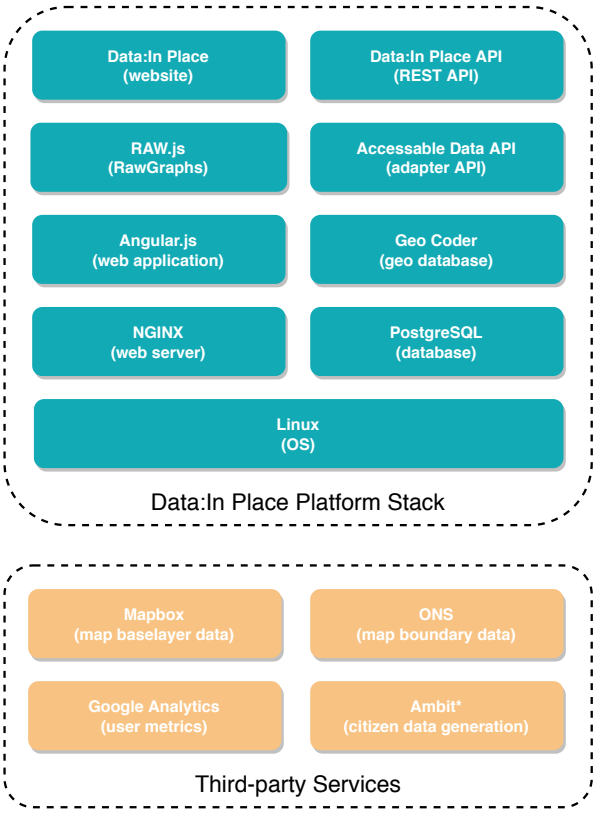


Figure 4.2 Data:In Place platform architecture

4.6. Data:In Place Interactions

RawGraphs (Mauri et al., 2017) interactions comprised of: (1) load your data; (2) choose a chart; (3) choose your dimensions; (4) customise your visualisation; and (5) download. This section will discuss the added features and capabilities built to extend these interactions to make it more usable in the context of open data and civic advocacy, as informed by the design principles (Section 4.3) and co-design process (Section 4.4) with the people. Based on the literature and the experience of working with the communities and the issues encountered around accessing and making sense of open datasets, the platform ended up implementating the following features:

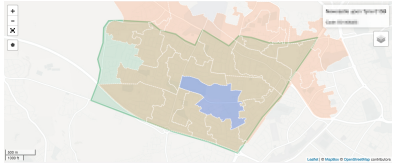
[18] http://mapbox.com
[19] http://openstreetmap.org
[20] https://github.com/artoolkitx/jsartoolkit5

(1) *draw a boundary*; (2) *pick a dataset*; (3) *map your data*; (4) *request a new dataset*; and (5) *explore community opinions*. The features are explained through user interactions (i.e. user stories) with the system.

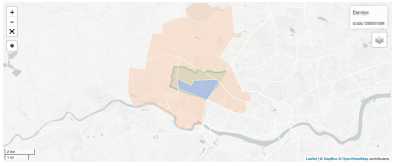
4.6.1. Draw a Boundary: Visual Map Based Query System

A user's first interaction with the system starts with defining a place they are interested in – bounding the *data-in-place* (Taylor et al., 2015). For that purpose, a user could make use of the GIS component of the system to explore and find the place they are interested in. The system enabled people to search for postcodes or just pan around the map to find places of interest (e.g. home, work or a potential engagement area). After finding an area of interest, it could be selected by drawing a boundary around it on a map using the GIS component (Figure 4.3a. When a boundary was finalised, the system automatically retrieved polygons of the intersecting governmental areas[21](Figure 4.3b). By zooming in and out of the map, the system switched between different administrative levels (i.e. output areas, wards, local authorities, counties, regions, and countries). The polygons were updated and redrawn on the fly while zooming in and out on the map (Figure 4.3c). Using the GIS component, people could define their query for open datasets by simply exploring the areas and defining places (i.e. drawing boundaries) of interest.

(a) Draw a boundary to define an area



(b) Get intersecting official area polygons



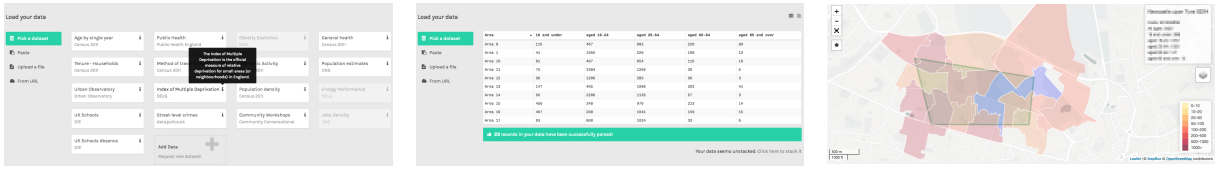
(c) Zoom out to retrieve new level area polygons

Figure 4.3 Visual map-based query system for open datasets

4.6.2. Pick a Dataset: Accessing Official Governmental Datasets

Next, the user could pick from a list of official datasets (Figure 4.4a) to be automatically pulled onto the system defined by the map query specified in the first step (Section 4.6.1). Again, zooming in and out of the map while a dataset was selected automatically redefined the query and retrieved a new set of data (Figure 4.4b) specific to the place on the map. This made the retrieval and comparison of statistics easier, but it also displayed the areas intersecting with the boundary representing contextual information (Figure 4.4c). The initial prototype had a list of official datasets (Figure 4.4a) that people participating in the co-design process identified to be important (Section 4.4.1). However, this list could be further extended in the future (Section 4.6.4).

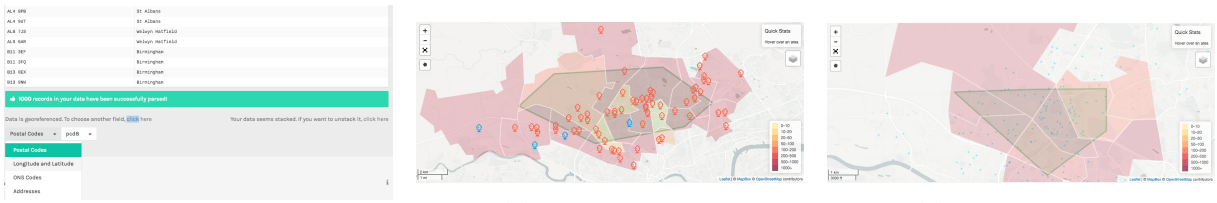
[21] http://geoportal.statistics.gov.uk



(a) Pick an open datasets **(b)** Get data from official APIs **(c)** Explore data with boundary

Figure 4.4 Automatic linking to different governmental APIs to access official data in one place

4.6.3. Map Your Data: Mapping Community Data



(a) Georeference the data **(b)** Map point data **(c)** Map postcodes

Figure 4.5 Mapping community datasets and comparing with the official sources

If the user had their own administrative data to map, they needed to upload that data to the system by either copying and pasting it, uploading a file or specifying a URL destination. Once data was loaded, the system tried to automatically map it by parsing the data fields. If the system did not successfully georeference the data (Figure 4.5a, there were controls for the users to do it manually. The system was able to map most standardised geographic data, such as point-based data (Figure 4.5b) and also UK postcodes, as a heatmap (Figure 4.5c). This feature enabled users to add community datasets for mapping and comparison with official data sources.

4.6.4. Request a New Dataset: Adding Official Datasets

Similar to RAWGraphs' (Mauri et al., 2017) ad-hoc visual models, whereby anyone can add their own visualisation templates to be used by the community, the Data:In Place platform aimed to enable people to add new official data models to the list. The functionality to add a simple dataset from a file and/or API endpoint URL already existed; however, in order to make use of 'map-based queries' (Section 4.6.1), it needed to include the correct parameters to pass in the URL. If the user wanted to get new official data to the system, they needed to use a simple form to request data so that it could be linked to the desired dataset (Figure 4.6). This was then saved to the database, ready to make the link between the new source and map query and add it to the platform. The form also enabled people to add requests for potential datasets they would like to see on the system without knowing the exact sources. This related to the first decision of choosing which should be in the system – they were related to the issues people identified as important and wanted to investigate and evidence through the use of data.

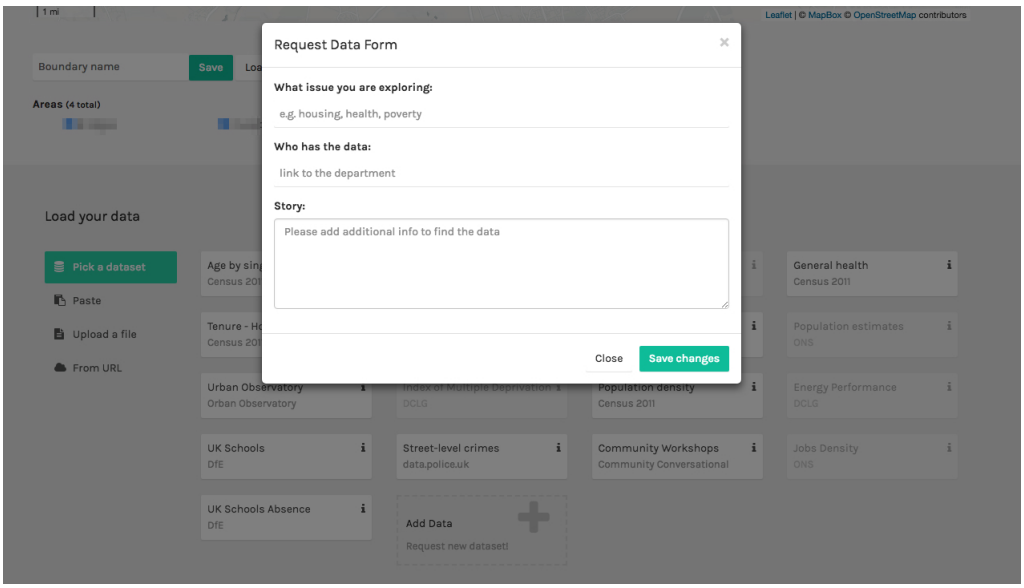
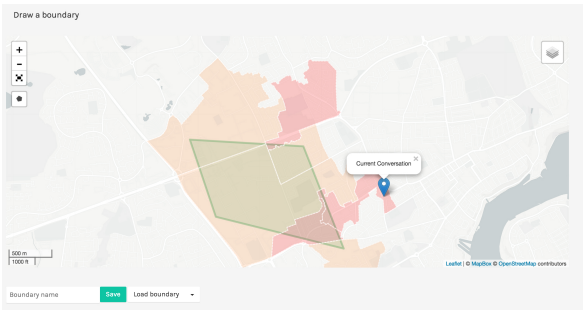


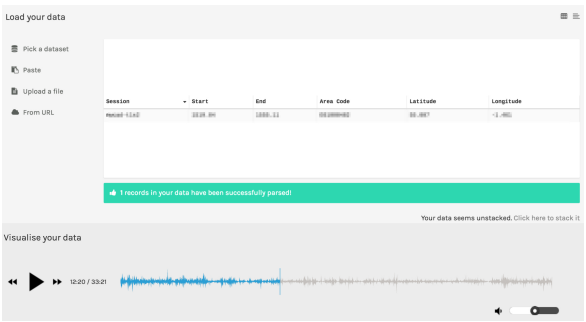
Figure 4.6 Use the data request form to get new official data on the platform

4.6.5. Explore Community Opinions: Adding Citizen-Generated Data

Leveraging the Ambit system, public opinion data (i.e. 'neighbourhood data') in the form of audio recordings from community workshops was added to the platform as another data source. The whole audio was time and geocoded, having a database entry for each coded opinion. This also meant it could be accessed similar to other datasets using the map-based query system and getting all the people's opinions in a specific place. Feeding this data into the system also bounded it within the governmental boundaries, creating a geographic link between public opinion data and socio-demographic characteristics from official statistics. When a user chose an opinion from the list (Figure 4.7b), the system automatically added the marker on a map and started playing the place-based recording (Figure 4.7a). As such, it was possible to integrate the public opinion data into the system side by side with the other datasets. For example, a use could explore crime statistics while listening to people talk about issues around anti-social behaviour on a particular street, or look at the availability of developing land in an area while listening to people talk about the lack of play areas for children.



(a) Define a query using the map



(b) Choose an audio clip to play

Figure 4.7 Listen to citizen-generated data from places in the neighbourhood

4.7. Discussion

Through an exploration of the current state of open data platforms for non-professionals and co-designing the Data:In Place platform with the communities over an 18-month period, a number of issues around using data in the context of civic advocacy and action were brought to attention. It quickly became evident that communities could see the value of data and its importance in civic action, but the barriers to accessing and understanding its relevance to local matters distanced them from using it effectively. This meant that people were often reliant on third-party professionals (or the researcher in this case) to lead in these processes, which in turn raised issues around dependency and the misrepresentation of communities.

The Data:In Place platform and the ideas of linking different datasets together through a place-based interface aimed at promoting the *'effective use'* of data (Gurstein, 2011) and supporting non-professionals and community groups in accessing and making sense of data in ways that were relevant to their matters of concern. Instead of presenting people with a set of information, the system started helping people define a place and then linking data to it. This enabled communities to decide which data was relevant and what granularity was needed to surface the issues they were exploring. The following sections will discuss the values embedded in the platform derived from the design principles and assess their realisation through the co-design process. This includes reflecting on the key parts of the system and its utility and current limitations as a tool for civic advocacy.

4.7.1. Accessing Open Data

Although national statistics and other official datasets (e.g. OGD) were seen as a valuable resource for neighbourhood planning, they were very difficult for people to access. Even at the level of query-based searches, national statistics relating to local neighbourhoods are often categorised using postcodes, area names and ward boundaries, a system quite unlike the ways in which communities reference their own neighbourhoods. To overcome this, a map-style interface was developed that allowed people to identify neighbourhoods in ways they were familiar with and a map-based query system (Section 4.6.1) that enabled people to obtain national statistics for the neighbourhoods and issues they were considering.

Relating back to the first design principle with regard to *making data relevant through place*, it was learned that an important aspect of contextualising data is making it relevant to the particularities of place understood both as a geographical place (i.e. locale) and a social place characterised by a community's interests and issues. The problem with big data in relation to abstract plans and procedures is that it always underspecifies the particularities of place. Work or procedures have to be abstractions relevant to the particular circumstances of their uses (Suchman, 1985, 2007; Taylor et al., 2015). What was discovered was that data, particularly big data, government data, etc., is no different; its abstract materiality means that its relevance to place and local issues is not always apparent, but it is something that has to be constructed, worked at and made visible in order to provide value in these contexts. Only then can people start

to construct knowledge from these abstract datasets, which will then open up spaces for enquires. This was experienced when map-based queries were integrated into the system and people started exploring available datasets situated in their neighbourhoods. Immediately, new questions were raised, as well as requests for additional datasets to be linked to the system for exploration. Aiming to address this, *data request forms* (Section 4.6.4) were integrated, which people could use to request new datasets not currently linked through Data:In Place. While this significantly increased accessibility to new open datasets on the platform, there was still a technical input required from the researcher, thus limiting the extent to which the platform could be totally community sustained. Such issues are challenging but not insurmountable. One way forward here could be to collaborate with official data providers (e.g. national and local governmental structures) to automatically link and publish their data through Data:In Place. Alternatively, the *dataset linking* could be made more modular, promoting the open-source community to contribute to the project and start linking available open datasets to Data:In Place's map-based query system. By following a tutorial, people could commit additional code snippets to extend the project hosted on the open code repository. This is even closer to the ideas of RAWGraphs' (Mauri et al., 2017) *'custom charts'*, which enable people to add their own chart templates for data visualising.

4.7.2. Diversifying Participation

The second design principle was to enable communities to challenge open data through their own experiences, which was realised in the platform by linking people's opinions on a place with official statistics. This give a more complete picture of the neighbourhood – providing context to open data or a way to challenge official statistics, but also bringing the real needs and aspirations of the community to the surface. The importance of this was pointed out by one of Charity B's workers, CW2: *I think there needs to be more of those conversations: 'This is happening, this is how it is in the reality.' And I like that, but I'm just thinking how we can make more of those conversations happen?'*

Furthermore, open data can be used to contextualise community data, and representing both in the same platforms opens up the possibility for communities to speak back to government. There is often a presumption around both big data and open data that its use-value for communities is both self-evident and veridical. This presumption is given credence by virtue of the idea that the data is somehow 'raw'. In contrast, Gitelman (2013) argues that the term *'raw data'* is self-contradictory since data is never raw; instead, she argues that data begins its life as an act of imagination in response to a need or a question. It is created in order to meet a need, and thus it is always framed with respect to a human purpose. The analogy she uses is that of creating a photograph – a human decision is always made as to where the camera is pointed and at what angle. Data, like a photograph, indeed like any representation, has to be created, and that means it has to be created by people in response to their purposes. The fact that the local community sees itself so differently to how the data collected by outsiders portrays it affirms that there is no such thing as 'raw data'. One limitation of the Data:In Place platform is that citizen

data was generated by working with the system in co-located workshop settings. However, open data platforms like this should promote wider participation and inclusion of ideas and voices. However, there is potential here for enabling much broader participation through the online nature of the platform, enabling those not present in workshops to participate.

The mapping tool could also provide charities with a way of understanding the work that other organisations do by sharing some of their organisational data. Local Charity A worker CW6 talked about his experiences of finding out what work others do in the community:

> *So, it seems that when we go into meetings, or when we meet with different organisations, people are doing very similar things or things that can help each other out quite a lot of the time, and a lot of the time a problem comes up as, 'Uh, I did not realise you are doing that," or 'Maybe we could help out along the same thing." So, as kind of providers of services, it would be useful for us to know who is doing what, as well as the individuals who would be using those services as well.*

This kind of inter-visibility is important because, in the past, it has led to friction and competition between the groups. Not only would it remove the competitive element, it could lead to collaboration:

> *[...] and if somebody feels threatened that you are kind of stepping on their toes, then they can get maybe a little bit defensive about what they are doing. But if they would understand fully what you were actually doing by using some kind of tool to gain that information and that can help put them at ease and then maybe help them feel a little bit more welcomed working alongside you as well.* (CW6)

Working alongside each other, they could also use the platform to collaboratively make sense of this data, to make a stronger case for additional funding and ultimately provide better services to the whole community. Similar findings were obtained in a study by Erete et al. (2016) on the use of open data by NPOs, where they emphasised that *'effective technologies'* should help NPOs access, interpret and collaboratively make sense of data related to their goals and objectives. Their work described the collaboration between NPOs and external data experts; however, using Data:In Place and making the data truly open and accessible, NPOs could be more active in defining the issues they want to explore and collaborating with other NPOs to achieve common goals. This raises questions around the value of these different sense-makings and whether making data accessible and linked together with place will help people make sense of it and use it.

4.7.3. Civic Advocacy and Action

The third design principle embedded in the platform was to *integrate local knowledge and data to enrich collaboration and promote civic action*. In designing Data:In Place, the aim was to help citizens in their endeavour to effectively use data for gathering evidence to contest decisions in their community. Coming back to the seven-layer model of Gurstein (2011) for *'effective*

use', it was illustrated that Data:In Place did well against the first five, but more things needed to happen for the effective use of the data in civic advocacy. The somewhat naïve assumption that by tackling the engineering problem-solving agenda through co-creating a tool for accessible data would have a direct route to civic action was quickly proven to be premature. What was learned through the process was that technology can only go so far to make data accessible and help people interpret, make sense and visualise it; however, it takes social capital to transform data into activities for the local benefit (Gurstein, 2011). For data to work for community problem-solving activities, there needs to be individual citizens or community organisations that are willing to own the issues, work on them and organise others to help advocate for change. Through the co-designing process, and by seeing that there were others who were passionate about tackling the same issues and improving their communities, such community organisers started to emerge and step up. Furthermore, some of that needed social capital could come from outside the immediate community and could potentially be facilitated through the platform. Data:In Place could be extended to include help and expertise from outside to understand and explore issues using data science. This means adding a new layer of communication between citizens and experts.

Work by Voorberg et al. (2015) highlights that one of the influencing factors of citizen involvement in policy-making and co-creating of public services is poor communication channels, in addition to the attitudes of people in power in regard to not wanting to include citizens they cannot control or rely on. However, reflecting on the engagements with citizens and the interviews conducted with members of the local council, it was learned that decision-makers were interested in what the citizens were telling them; they just lacked the suitable means of communication to facilitate dialogue (more in Johnson et al. (2018)). Perhaps it is more about collaboration and working together with decision-makers rather than against them, in addition to using a system as a medium to get expertise and add new meaning and different types of data to ensure everyone's views is represented. There is scope for the platform to support similar collaboration and sensemaking in other contexts where decision-makers and citizen data-subjects are otherwise siloed. Contexts where the inclusion of, and collaboration between, citizens and decision-makers in processes of making sense of and interpreting data could be the key to tackling the problems of society.

4.8. Summary

This chapter presented the co-design and development of Data:In Place, an online platform for non-professionals to access and make sense of OGD in the context of (1) a neighbourhood plan, gathering evidence to back up their claims, and (2) local charities, understanding the communities they operate in and for and the impact they are having. Through the process of investigating people's relationship with data and how it might help them collaboratively make *'effective use'* of it so as to have a stronger voice in what happens in their community, a set of roles played by data in these processes were uncovered. With Data:In Place, it was illustrated that open data – when accessed correctly, presented fairly, tied to a place and contextualised – can be an important tool for helping people explore issues and understand their communities. The aim of the case study was to democratise data access, comprehension and sensmaking in the hope that the communities – who are the subject of this data – can become more than just those with the issues, but also active contributors of knowledge to tackle them locally. In addition, the case study explored the extent to which the expert can be taken out of that loop, mediated or replaced with automated systems. Furthermore, new challenges for the effective use of data by communities were discovered, which were more related to human engineering or SI than engineering problems. The next chapter will reflect on these challenges and the first two case studies in more depth in order to build a better understanding of the factors for the effective use of data by communities for civic participation, advocacy and action.

Chapter 5. CDI: A Model of Shared Data Interactions for CI

A novel solution to a social problem that is more effective, efficient, sustainable, or just than existing solutions and for which the value created accrues primarily to society as a whole rather than private individuals.
— James Phills et al. (2008)

This chapter reflects on the previous two case studies in this thesis, which aimed to support the *effective use* of data by citizens through active engagements. Up to this point, the thesis has looked at ways to engage citizens in the exploration of local issues through data production (Case Study I in Chapters 2 and 3) and through accessing and making sense of publicly available datasets (Case Study II in Chapter 4). Despite taking a community perspective, both of these case studies had a sharper focus on the engineering problem-solving agenda by responding to issues regarding the lack of specific technical tools, expertise or access to data, which included setting up a community commissioning toolkit to enable citizens to access resources for collecting or commissioning environmental data (SenseMyStreet) and designing and building an open data accessing and visualisation platform (Data:In Place) for communities. Drawing on these case studies and the wider research in Ubicomp, HDI and CI, this chapter starts off by analysing the technical challenges of *accessing*, *making sense* of and *effectively using* data to support civic participation and advocacy. In addition, the chapter begins by analysing the successes and failures of each project and initiative in the context of this research. It is important to note that the analysis of the case studies was conducted from the perspective of an experienced data scientist and information engineer designing and building usable data systems for communities. Furthermore, throughout the case studies, there was a constant dissembling of the role of the professional data analyst, exploring how far the expert's knowledge can be removed from processes of active data use by citizens.

After this, the chapter discuss the important aspects of human interactions with data and systems facilitating these interactions. However, these are not only limited to interactions between humans and machines, but also extend to the 'human friction' often rooted in particular contexts and the social dynamics within communities. Through this analysis and the reflection on the previous case studies, it is possible to outline the current open challenges that still exist for shared data interactions at a hyper-local community level. Based on these challenges, this chapter underlines the methods, strategies and practices and outlines a holistic approach through which usable data systems could be formed. Building on the critical approach in CI (Stillman and Linger, 2009), the conceptual framework of CI (Gurstein, 2012), a commissioning model (Garbett, 2017), an effective use framework (Gurstein, 2003), and extending the CAF presented

in Chapter 3, a model of shared data interaction as CDI is offered, which provides a definition of CDI and ontology within the context of the effective use of data for community action in relation to *roles*, *infrastructure*, *resources*, *constraints* and *capacities*. Going forward, the model provides an abstract set of considerations when designing systems of shared data interaction for the effective use of data by communities for civic advocacy and action.

Related Publications and Acknowledgements

- Acknowledgements go to Dr. Andrew Garbett and Dr. Emma Simpson for collaborations around open data in public health setting which provided further ideas for discussions and evaluations of the challenges for CDI.

5.1. Data on the Web

Technology and policy advocates have been calling for open data for some time, a call which, in some respects, has been answered by institutions publishing their data; however, the hopes and promises of open data empowering everyone have been short-lived (Davies et al., 2019). Data, often represented as machine-readable files with abstract descriptions and column headers, holds the values of the specific purposes of data generation or production. There is an abundance of data (or in some case metadata) available on the web, and more is being published daily, but there is actually a paucity of data and technology for helping to understand what kinds of questions could be answered by particular datasets. By supporting the Semantic Web agenda (Berners-Lee et al., 2001), data can be standardised and described for querying and retrieval by machines, a significant step in the right direction with regard to helping to remove ambiguity around understanding what a particular dataset represents and reducing development costs on data accessing, combining and merging. Although linked data standards[1] for the web have been around for over a decade now, we have not yet achieved the goal of having a 'Web of Data', as opposed to a sheer collection of datasets on the web. As a result, data science skills are becoming increasingly more important to aid the exploration of data and support evidence-based practices in decision-making and action. However, organising and linking data on the web is only the first step towards its effective use by communities for civic participation and advocacy.

5.1.1. Organising Data on the Web

There are a multiplicity of ways that data and information are organised on the web and can be shared between people. In the corporate world, it is essential to keep track of organisational memory (OM) (or knowledge) (Tuomi, 1999) in a way that it could be accessible and usable to all the members of the organisation. In addition to linking data, knowledge graphs are often used to connect data together and make its retrieval easier from a variety of different sources. The most notable knowledge graphs are those used by Google to enchance its search engine, while another well-known technique is the archival and indexing of resources. The retrieval of information from those archives follow a system of catalogues that are identified or tagged by topics or other common identifiers, examples of which are Archives Portal Europe[2] and, probably most recognised in academia, the Digital Object Identifier (DOI) system [3]. In addition to formal ways of organising data and information on the web, there are also systems for community curated 'linked data', the most popular source of which is unquestionably WikipediaÕs free online encyclopedia, which enables anyone who has internet access to use the resource and contribute to it. However, in community problem-solving activities that aim to use data through IS, attention to these practices often falls to the wayside.

The vast number of connections and links between individuals that are enabled by modern technologies and can be harnessed through the web have opened up new ways of information and

[1] https://www.w3.org/standards/semanticweb/data
[2] https://www.archivesportaleurope.net/
[3] https://www.doi.org/

knowledge sharing. On top of media sharing platforms, such as FB, Instagram, and YouTube, to name but a few, there are multiple forum-based Q&A systems that aim to connect and exchange information between individuals and groups, e.g. online platforms such as Stack Overflow for any question related to software development, or websites like Reddit and Quora, which encourage discussions on a variety of topics. Although there is an abundance of different tools available to retrieve information from digital systems, they often require interactions that are confined in the boundaries of specific technology or contextualised in a particular way. This is especially evident in interactions with systems that provide numerical big data, which is inherently complex in its nature. Such data is organised and accessed in ways that people would not organise and retrieve information themselves in their everyday lives, hence the expression 'data in "machine readable" format'. Furthermore, what motivates the work in this thesis is how to enable the effective use of data by communities for civic advocacy and action. Indeed, it is essential to have data stored in a way that is accessible in standardised formats in order to aid the creation and building of tools that would interface with data archives, but these tools also need to be able to guide citizens in contextualising and translating abstract datasets into helpful resources for civic participation.

5.1.2. Making a Formal Data Request

To investigate the potential uses of data in the context of civic participation and action, there is a need to look at the current ways citizens can gain access to open data from official sources. With the Freedom of Information Act [4], a new mechanism was set up for people to access data held by public authorities, enabling anyone who wanted to request data from public authorities by submitting an FOI request, which can be looked at as a formal query for data, where a person or an organisation deposits a question to the data provider (i.e. public authority). Submitting an FOI request can be a lengthy and complicated process, often requiring knowledge regarding legal processes. In order to streamline these processes, MySociety came up with an online platform called What They Know[5], which helps people submit their requests through an online form. While this has made the process of submission much easier, the actual requesting of data has not improved. In order to get the data needed, one has to know about the existence of some unpublished document or dataset being requested. Apart from using the platform to explore previous FOI requests, there is no way for one to make sure that particular data is not yet published somewhere else. Thus, as long as there is no unified 'Web of Data' (Section 5.2.2) that describes all the existing data sources on the internet and makes them universally accessible trough standards to any service or client, this remains an issue.

Another scenario which can be encountered, is when a request is formulated as a direct question without an indication to particular requested resource. In that case somebody has to make the association between the question and a particular dataset (i.e., describe the relationship). Now we have reached a fundamental problem with data available on the web: *the lack of links between questions and abstract datasets that might bear the answer*. The ways that these

[4] https://ico.org.uk/for-organisations/guide-to-freedom-of-information/what-is-the-foi-act
[5] https://www.whatdotheyknow.com

connections are made are largely unknown and often require specialist skills with combination of tacit knowledge. This means either understanding the organisational structures of a public authority or knowing the initial purpose for the data generation. It is unlikely that this knowledge transfer can occur from organisations to citizens, however by contextualising data in different environments, it might help discover new uses and knowledge enabling to re-purpose these datasets. Data publisher and designers of information systems should not only aim to achieve the open data and linked data standards, but also look to achieve the *'FAIR Data Principles'* (Wilkinson et al., 2016) and the model of *'Effective Use'* (Gurstein, 2003, 2011) to promote uses of data by different user groups and within communities.

Underpinning the current state of the art of web technologies for data use, drawing on related works and examples from the Case Study I and II, the following two chapters provide a reflection of use of data by communities. Through this, a better understanding can be formed of ways that people interact with data within a complex sociotechnical systems to underline and identify existing challenges. By analysing not only technical aspects, but also social and organisational impacts, will enable to make better design choices for tools and methods to be used for purposeful and effective uses of data in civic participation, advocacy and action.

5.2. Technical Challenges: The Information Engineering Problem-Solving Agenda

The promise of the third wave of computing, or 'calm computing', envisioned by Weiser (1991) has only been partly realised over the past 25 years. Indeed, there are Ubicomp devices everywhere and sensors are being integrated into the fabrics of our cities and homes, so much so that they often seem to be without any physical presence; however, this has not solved the problems of information overload and has not enabled humans to have seamless interactions with these interconnected computer systems. People have also argued that the vision has not been fully realised because of the loose boundaries of the paradigm and the lack of 'killer apps' that would make it clearer and set the focus (Sorce et al., 2017). On the other hand, the wide availability of computing devices, the internet and IoT promotes a 'do-it-yourself' and 'hacking in the real world' mentality (Abowd, 2012), which means that research is moving out of the laboratory setting to the real world and into the hands of non-'tech-savvy' people. A similar shift had happened in personal computing by the mid-1990Õs, when HyperCard was introduced, enabling more people to interact with computing systems; however, Ubicomp is still searching for its own 'killer app'.

Despite the pervasiveness of computer systems and the availability of different information and data sources, the possible ways of interacting with data are still limited to specificities of individual computing systems or proprietary protocols, which require humans to know the 'lingo' of these systems. This means that in order for one to make use of these systems and the data they hold, one needs to know how they are built and operate, or at least how to use the technology to interact with them. This section examines challenges related to information engineering from the

technical aspect of purposing data from the web. Additionally, these challenges are also related to barriers in dissembling the expert data professional. By looking at the current state of open data standards and technologies, the section reflects on issues that surfaced from the previous two case studies, such as missing links between datasets, the lack of machine-readable metadata, non-standardised querying methods or the lack of following the gold standards of publishing open data. Building systems for open data and trying to make data more accessible and usable for people enabled the author to take an in-depth look at how data is currently projected to people and helped identify the barriers to its use.

5.2.1. Open Data Standards

	Bronze	Silver	Gold	Platinum
Legal				
Openly licensed & legally reusable (= 'open')	✓	✓	✓	✓
Clear rights statement, detailing any copyrights		✓	✓	✓
Privacy issues addressed		✓	✓	✓
Machine readable rights statement			✓	✓
Practical				
Accessible on the web	✓	✓	✓	✓
Discoverable (linked to from other web pages)		✓	✓	✓
Data is timestamped or up to date		✓	✓	✓
Data will be available for at least a year		✓	✓	✓
Guaranteed timeliness (data always up to date)			✓	✓
Regular backups of data			✓	✓
Quality issues documented			✓	✓
Technical				
Data uses a machine readable format		✓	✓	✓
Data published in content appropriate formats		✓	✓	✓
Data uses open standard machine readable formats			✓	✓
Single consistent URL for downloading data			✓	✓
Machine readable provenance documentation				✓
URLs used as identifiers within data				✓
Social				
Data is documented		✓	✓	✓
Contact details for people to provide feedback and ask questions		✓	✓	✓
Machine readable metadata (documentation)			✓	✓
Social media accounts used to promote data			✓	✓
Forum or mailing list for users			✓	✓
Dedicated comms team building user community				✓

Source: https://certificates.theodi.org/en/about/badgelevels

Figure 5.1 Levels of commitment for open data

Open data refers to data that is freely available for everyone to use, repurpose and republish without any copyright restrictions (Chapter 1). According to the Open Definition [6], open

[6] http://opendatahandbook.org/glossary/en/terms/open-definition

data needs to be open both *legally* and *technically*. The latter refers to having data available in machine-readable and bulk form. To help keep open data consistent on the web, a set of agreements called open data standards have been developed that help people and organisations publish, access and use data [7]. Additionally, there is a distinction between different levels of open data publishing – *bronze*, *silver*, *gold* and *platinum* – specifying how much commitment publishers have made to make their data accessible to others (Figure 5.1).

Developing Data:In Place's map-based querying system (Chapter 4) for open data, the author was faced with the reality of having to adopt the code for each data publisher individually and having to use specific types of geographical representations for retrieving specific data and different query languages, which often did not conform to open standards of web querying[8]. Additionally, when geographies or data were updated or aggregated, they often lacked links that could map the changes to existing or historical datasets. This was the case with geographies published by the Office of National Statistics[9]. This meant that data converters, translators and adaptors had to be built to create these links between different datasets. However, in order to create these links, there was a need for in-depth knowledge about how the data was collected, aggregated, stored and published. Ideally, this information could be described alongside the data and also be retrieved in a machine-readable form, but it was often lacking in detail, even from the documentations on the data portals. These barriers make the integration of open data (i.e. available data) to *accessible* and *usable* data for communities far more difficult and labour intensive for developers of these systems.

5.2.2. State of Linked Data

When people talk about open data, there is always mention of linked data, which is one of the building block of the Semantic Web or 'Web of Data'. Originally proposed by Tim Berners-Lee, director of the World Wide Web Consortium (W3C), linked data refers to structured data that is interlinked and described by relationships (or vocabularies) between data and 'things'. Let us illustrate this through an example:

> Six cities in the UK have urban observatories, all connected through the City Observatory Research platfOrm for iNnovation and Analytics (CORONA) project[10]. All six are doing some type of environmental monitoring and data collection, which means that they all have different ways of deploying sensors and collecting, storing and publishing data. However, if that data would be described and made available through standard descriptions and links (i.e. linked data), then if one, for example, wanted to know the current outside temperature of the UK cities, one could write a single query that would return all the temperatures from each city observatory's data portal in a preferable format.

[7] https://standards.theodi.org/
[8] https://www.w3.org/standards/semanticweb/query.html
[9] http://geoportal.statistics.gov.uk/
[10] https://urbanobservatory.ac.uk/corona

This is only possible when all the data and relationships are presented through standardised links [11]. In practice, linked data itself can be looked up and represented using many different open standards, such as URIs, HTTP, RDF, SPARQL, OWL, and JSON-LD (see more [12]).

In theory, if all the data is described and interlinked using linked data standards, one should be able to build a single query that returns data from multiple data providers in a format that is useful in a particular context. Whether building a report to discover and understand the most pressing issues in the neighbourhood or gathering evidence to advocate for change in the community, this capability to request, access and merge datasets from different publishers is almost always needed when building an accessible and usable data system for local communities. However, will this also work in practice? Let us examine an example of such a query:

> What are the current demographics, health statistics and powers of governance in a particular hyper-local community defined by a specified geographic area?

In the UK, this query spans four different governmental institutions: Office of National Statistics, Ordnance Survey, The National Health Service and The House of Commons. Although some of their open data systems have capabilities to request data using linked data standards, they are often experimental and do not interact well with other linked data systems. The example provided here is a real-life example of building a usable data system for public health interventions around breastfeeding [13]. This project is especially interesting because it aimed to intersect community generated data (Balaam et al., 2015) through a commissioned technology (Garbett et al., 2016) with official open datasets and make it useful for local communities. After multiple failed attempts to leverage linked data to achieve this, the research team[14] had to build a bespoke system that made these links between different data providers and combined them into a visualisation interface. In the end, linked data was leveraged only to some extent, not in a way that would be envisioned by the Semantic Web community. This also meant that with any updates to the structure of the data, the links had to be rebuilt. This puts the long-term sustainability of these systems in question because of the effort required by the researchers to provide technical support to keep the system running and updated. Similar issues were raised in the second case study (Chapter 4) while aiming to make data more usable and helping people discover new available datasets that might be useful for evidencing or exploring a local issue.

Finally, the support offered by data publishers is not sufficient to fully leverage the capabilities of the APIs. There are few very good examples of advanced usages of data and rarely a forum maintained for the data users to submit issues. Looking at the levels of commitment for publishing open data (Figure 5.1) from the perspective of effective use (Gurstein, 2011) in civic participation, only *gold* and *platinum* providers focus on engaging with communities to enable the better use of their data. From experience, there are only a handful of data providers (Chapter 1 Section 1.4.3)

[11] https://urbanobservatory.github.io/standards

[12] https://www.w3.org/TR/?tag=data

[13] https://feed-finder.co.uk

[14] The research team consisted of Dr Andrew Garbett, Dr Emma Simpson and the author. Together, a system was designed based on the design requirements obtained through engagement work of the two collaborators; Dr Garbett and the author building a platform; and the whole team deployed 'in the wild', which was launched via a stakeholder workshop as part of Fuse - Centre for Translational Research in Public Health.

that could be considered to have gold, and even less platinum, standard open data. The current state of linked data amongst official open data providers does not extend itself to the ideals of Tim Berners-Lee's 'Web of Data'. In turn, this again puts local communities at a disadvantaged position – not getting the data they need in a form that would be beneficial for them without the technical know-how and the interference from a professional data scientist. Furthermore, when there is additional data that has been collected by the community (as in the case of the platforms: SenseMyStreet, FeedFinder, and, to some extent, also Data:In Place), it is becoming increasingly difficult to intersect and contrast that data with official open data without the existence of the 'Web of Data'.

5.2.3. Human-to-Human vs Human-to-Machine

This section describes the differences between human-to-human and human-to-machine interactions and identifies the challenges for human-data interaction (HDI) in relation to social and cultural constraints that govern the interactions. These constraints have a major role to play in how we design human-system interactions for IS and the guiding strategies we apply to enable people to effectively use them.

Human-to-human communication via speech is built on pragmatic linguistics and the principle of charity, which can also be explained through the *'Gricean Maxims'*, named after philosopher Herbert Paul Grice (1975). What this means is that the way humans communicate and understand each other does not solely rely on grammatical structures (i.e. written texts) and the things being said. Humans also pick up on utterances in speech that take into account the contextual and social dynamics of situation and relationships and operate on a principle of charity, trying to interpret speakers' intentions to the best of their abilities. All these trades are illustrative of human intelligence, which helps us to create conditions for an adequate interchange between each other through conversation.

However, upon human-to-machine interaction, these rules do not apply, which creates asymmetry between humans and machines. Early research conducted in human-machine communication at Xerox PARC illustrated how these asymmetries would lead to breakdowns in communication (Suchman, 1985). An example given with an 'intelligent' photocopier revealed that the level of abstraction drawn by the computer system about human plans did not convey the intended actions, and without computers understanding human trades of communication, breakdown is inevitable. Now, more than 30 years have passed from these first attempts to understand human-machine communications from a behavioural- and human- centred perspective, making way for numerous communication environments and devices to be built that interface interactions between humans and machines. On top of traditional interfaces, such as graphical user interfaces (GUI), there are now interfaces that use natural speech (i.e. spoken voice), tangible computing, gestures recorded by cameras or other sensors, using wearable computing, or a combination of them, delivering a multimodal interaction interfaces for communication.

We are now seeing multimodal interfaces that are connected to the internet, relaying users' interactions to artificial intelligence systems, which can take advantage of immense computing power and make use of big data and machine learning, enabling them to understand more about context and learn about the environment they are in. Such interfaces are called virtual assistants, with examples such as Apple's Siri, Google's Assistant, Amazon's Alexa or Echo and Microsoft's Cortana, to name but a few. Assistant technology has now been embedded to most of our 'smart' personal and home electronics – smartphones, smartwatches and other personal computing, home media systems, hubs and multiple home monitoring devices. The assistant software can perform tasks like schedule calendar events, set alarms, search the internet, pull up information, adjust settings on a particular device, and respond to individual-based commands and questions, enabling engagement in two-way conversations between humans and their computers. Giving more personal data to these systems and enabling artificial intelligence to 'profile' an individual can help these systems potentially learn about an individual and adapt their responses in a manner that is truly personalised. This could mean that computers are now able to mimic some of the trades that make up human-to-human communication. However, what has started happening is that rather than waiting for the machines to learn the context and the particularities of each individual to respond to them in an adequate manner, humans have started to preempt machines giving inaccurate responses. People are trying to learn how to interact with the machines to get them to do what humans want. In other words, people are learning and adapting to 'talk' to the machines rather than vice versa, showing again the intelligence of humans and how we operate on the principle of charity when communicating Herbert Paul Grice (1975). This principle could also be used as a tactic for improving the responses of machines, e.g. assessing the usefulness of the information retrieved by the machine as a response to a human request. By doing this millions of times using a diverse sample, we could start building links between requests and relevant responses.

In some regards, this is what Google's PageRank has been doing for websites for years now – counting the number and quality of (hyper)links to a page to determine their relevance. It is one of the tactics used by Google to respond to people's search queries on their platform. The more people reference (i.e. hyperlink) particular websites for a source of information, the more relevant it becomes. The know-how in this method is obtained through automation and crawling the World Wide Web for these links. This not only makes it difficult to assess the real value of the links to a individual use case, but also means that it can be manipulated (Gyongyi et al., 2006). Following PageRank algorithm, there have been more sophisticated algorithms implemented, such as HITS (first proposed by Jon Kleinberg) (Kleinberg, 1999), CLEVER from IBM [15], and Google's own algorithm updates, such as Penguin, Panda and Hummingbird, which better deal with link spamming and assessing the quality of the links. Particularly interesting is the Hummingbird algorithm, which Google claims focuses more on natural language, trying to understand context and meaning rather than pure links and keywords to reference things. In a way, this is closer to the idea of the Semantic Web (Section 5.2.2), which aims to describe

[15] https://www.ibm.com/blogs/research/2018/05/clever-adversarial-attack/

the concepts and relationships between 'things' on the web. In addition to algorithms that are deployed to work autonomously by scraping the web and providing relevant information to people, we need to look at cases where more human approaches are needed. Discovery of relevant datasets is one such case, where feedback loops from humans could be beneficial for helping others understand and make use of data.

The first obstacle to HCI has traditionally been to get the machine to do the things 'you' want. In the context of IS, these barriers to using the machine still exist, but they can be overcome by applying methods such as user-centred design. In this regard, the design and development of the Data:In Place platform (Case Study II in Chapter 4) is a good example of how one might tackle this human-machine communication problem. However, the focus then shifts to how humans can learn from the data retrieved by the machine, and more specifically what types of data could be useful and relevant to particular people and their situation.

5.3. SI Challenges: Community Problem-Solving Activities

Coming back to the initial vision of Weiser (1991) of 'calm computing', although computers can be found in just about any context of modern life, they still lack the contextual and social awareness of the situations and people who are using them. According to Weiser's vision, computers of the future are always on hand and augment human actions and interactions in everyday life. To some extent, we are starting to see this in context-aware or location-aware applications (Section 5.2.3); however, the challenge still appears to be understanding the human intent, i.e. the 'activity' that the human is trying to accomplish. This also echoes the stance of Ackoff (1999) on computers and their inability to produce wisdom (i.e. actionable knowledge) from data – 'something' that can be put into action.

Furthermore, what is the use of 'computers in the background' that collect vast amounts of data if there is nobody who interacts, understands and acts on that data? In this case, data just becomes noise. Instead, data needs to go from *available* (i.e. open) to *usable* to *useful*. This agenda is a central challenge for research in human-centred computing and is prevalent in HCI and HDI fields. Low engagement with data is often fully credited to the lack of or access to technical know-how, skills and tools; however, there are additional challenges involved that relate to the social ergonomics (Brown and Newman, 1985; Ferreira et al., 2020) of using these complex systems in different contexts. Some of these challenges come from the fact that communities are diverse, all having unique cultural peculiarities and imperfect organic ways of doing things that are difficult to interpret into intelligent machines and processes.

5.3.1. The Importance of Community Links

The case studies in this thesis engaged with communities that all had different economic and cultural backgrounds, social capital (i.e. the skills and resources to produce knowledge), existing proficiency to use data and existing community links. Without any engagement, this already puts each community group on a different level in terms of its capability to effectively use

data for actions for local benefit. Many have noted that the way to increase the capabilities of the community is to actively include it in the design process or practice co-design (Bernarda et al., 2017; Dantec and DiSalvo, 2013; DiSalvo et al., 2008; Ferreira et al., 2020). Both of the preceding case studies in this thesis where conducted through a user-centred design approach, where communities were actively involved and drove the design of the technical solution. The SMS toolkit (Chapter 2) and the Data:In Place platform (Chapter 4) were both direct results of a longitudinal engagement with community groups. However, as the findings of both case studies show, the tools and data they provided could only get people so far. The groups that could effectively use data either had greater social capital or larger community networks. For example, the advocacy group in Case Study I (Chapter 3) consisted of self-motivated (ex-)professionals who had connections in local government, local universities and public health services. Additionally, they were already involved with multiple groups, and they knew people outside their group who had similar goals to them but in different hyper-local areas, which meant that their connections extended further than their own community. This goes back to the ideas around social capital (Coleman, 1988) discussed in (Section 1.4.3) of the literature review. These were the 'ties' that Granovetter (1983) refers to as *linking* connections. Granovetter (1983) stresses the importance of an individual having multiple 'weak' *linking* and *bridging* ties, allowing them to tap into skills and resources outside their social network, close friends and colleagues. In practical terms, this means that communities are more likely to get access to skills and resources that are out of their own capabilities, skills such as community outreach and most importantly data science skills in the context of the effective use of data. An example of this was the Twitter bot developed by one of the community members that published easily digestible daily graphs about the air quality from commissioned monitors. It was developed by one group at first but then shared amongst and adapted by other groups. However, Case Study I (Chapter 3) also included groups that lacked these links, which meant that data was collected or commissioned by them but not acted upon. Similarly, in Case Study II (Chapter 4), the fact that open data was made available and usable for people through the Data:In Place platform did not mean that people found it useful and could use it right away in activities for local benefit. To do so needed more capacities, effort, skills and resources from local people and also from outsiders.

5.3.2. Taking Ownership

An important aspect of working with communities around complex socio-technical issues is the importance of community resilience (Aldrich and Meyer, 2015; Bernarda et al., 2017; Ferreira et al., 2020; Mark and Semaan, 2008). There are multiple definitions of community resilience (Aldrich and Meyer, 2015; Mark and Semaan, 2008); however, in the context of this study, it can be referred to as a community's ability or strength to adapt to changes and take action towards positive change. Designing technologies with communities is also about people *archiving control* (Vines et al., 2013), which can be seen as a part of the activities that contribute to community resilience. Bernarda et al. (2017) state that the designer has the ability to act as an 'enabler'

and facilitator to help move towards *'empowerment'*. However, people need to take control and ownership of the the *'factors and decisions that shape their lives'* (Bernarda et al., 2017, p. 903).

This is something that came through across the two previous case studies. For example, in Case Study I (Chapter 2), a toolkit was built that enabled citizens and community groups to collect and commission environmental data about neighbourhoods. However, people needed to take ownership of those datasets to inform their actions; otherwise, the data just became another source of available data on the web. This does not mean that that data is not useful for anybody; it just means that it is not used for the purposes for which it was generated or commissioned. A lack of ownership and control over decisions and factors can also affect and stall progress towards positive change. What follows is an example of the lack of citizen control that happened in the local area where the author was operating:

> A community group was trying, for several years, to get traffic calming measures implemented around their local school. Further down the line, a local charity secured funding for a project to do exactly that. However, the local community group became agitated by the charity because it was not consulted on the project, and this came as a surprise to them. Soon after, a local councillor got involved, and the project was postponed for another year so as to go though a formal consultation process.

What this story illustrates is the importance of engaging people early and including them into the design and implementations of solutions that affect their community.

5.3.3. The Usefulness of Data

The aim of the two previous case studies was to make data *usable* for citizens in civic action, as opposed to its current form of simply being *available*. Previous case studies have looked at how to enable people to collect and commission the right data (Chapter 2) and enable access to and interpret already available data (Chapter 4) so as to understand what data they need. What was learned along the way was that data is not self-evident (Gitelman, 2013; Tuomi, 1999) to people; it needs processes, often social processes, to help give data meaning. Data is often particularly situated in terms of the community that is trying to make sense of it, the place it was generated, and the kinds of questions that people want answers to. This echoes the arguments of Gitelman (2013) that raw data is an *'oxymoron'*, in that the same piece of data can be used and interpreted differently in different contexts and places. Data does not always make sense when it is taken out of, or put into for that matter, its particular context of both place and the community actually working with it. There are more steps involved in going from *usable* data to *useful* data. In fact, *useful* is not actually a property of the data because all data is useful – it is actually a property of the person or the interaction or activity a person is doing. Citizens need a purpose for what they use the data for, and then it becomes useful to them. This means that having data in the right formats is not enough; instead, people need to have the skills to be able to use it. Citizens need to be asking the right questions and planning out the actions, sometimes even before there is any data available.

5.4. New Challenges of CDI

Open data advocates often paint a romantic picture of data being available for everyone to consume, like the air in the atmosphere, and making it ready to leverage its powers in aiding decision-making, improving public services, fighting injustices and improving equality for all. However, at closer inspection, and from the findings of Case Studies I and II, things are not as they have been advertised. Depending on where a particular person sits in a given society, they may think that current systems for data access and use are either fantastic or under-serving their needs. In addition to the clear need of data science skills to interact with the abundance of data-rich systems, there are also a number of barriers to using data for purposes other than its intended collection: published data is often not standardised; there is a lack of machine-readable metadata; tools and systems for data access are disperse and often not fully functional; and there is lack of clear documentation on how to leverage the existing infrastructures.

On top of the barriers related to the technical capabilities of data systems and the skills and resources needed to leverage them (i.e. the engineering problems), there are also multiple challenges around the human aspect of information engineering and the production of actionable knowledge from data (i.e. community problems). These relate to the social processes of finding ways that data can be made relevant and useful in different contexts of the everyday lives of citizens on individual, community, city and state levels. All this makes up a wide set of systems and processes that all link to processes of CDI.

5.5. HDI as Shared Data Interaction

Digital technology and the availability of data have created new opportunities to engage in civic participation. New channels of communication could potentially provide a stronger (and more diverse) voice to citizens and support them in active participation in democracy. Appropriation of data as a means of engagement and participation can take many different forms (Chapter 1). Similarly, information engineering has examples of big data analytics systems built to algorithmically harness data from citizens in order to streamline civic participation (Tenney and Sieber, 2016). The tendencies are that these two modes – active and passive, or direct and indirect – are kept separate from each other because of their origination from different disciplines, difficulties around mathematical representations of context or deeper ontological conflicts. Furthermore, there have been criticism of big data analytics and the way open data advocates portray big data as providing citizens with information and resources for capacity building and empowerment but are actually shifting towards datafication, i.e. instead of serving the needs of citizens (particularly the marginal and vulnerable), they are catering to agendas driven by large industry (Chandler, 2015; Michael and Lupton, 2016; Tenney and Sieber, 2016) and increase in 'digital divide' (Gurstein, 2003).

Alongside increases in the capabilities of technology and data, and new modes of participation, new research fields are emerging from HCI and other disciplines that aim to understand the impacts of these developments, e.g. HDI, a research field that is interested in understanding

how humans interact with data in this complex ecosystem involving ubiquitous and pervasive computing, big data and IoT. For example, Mortier et al. (2014) place the individual at centre stage and focuses on issues around *legibility*, *agency*, and *negotiability* in the context of the rapidly growing data-driven society, looking at how individuals can understand what data about them is being used and for what; what can be done to give people better mechanisms to control their data; and how the management of dynamic relationships arising from data use would look. In this regard, HDI focuses on analysing *'individual and collective decisions we make and the actions we take, whether as users of online systems or as subjects of data collection'* (Morton and Lurie, 2013). Without a doubt, these issues are becoming increasingly important in a world where data about people can be used to influence the behaviour of individuals (Bond et al., 2012).

However, in the context of community data use, HDI does not only apply to interactions with one's own data that is being exchanged through pervasive computing devices. The interactions with data that are the subject of this thesis move away from only looking at personal data to also consider data that has been collected about communities and cities at large – *civic data*, e.g. data collected about the environment during a school run at the local community level. This interaction may include data about the individual (e.g. location traces), but it also includes data recorded about the local air quality. Although the data may be valuable for an individual, it becomes much more valuable to the particular collective trying to instrument change in their neighbourhood. Such an example of shared HDI played out in the first case study (Chapter 3). In addition, HDI for shared interactions concerns data that has already been generated by other organisations and the government, which can also turn out to be valuable to citizens advocating for positive change in their local communities.

Through the course of the previous two case studies, the aim of the research was to explore ways data can be made usable to support civic advocacy and action. This was reflected by the designed digital tools and systems, the interactions that supported this type of *active engagement* around data and the ways that citizens could use that data to take local action (the CAF model in Chapter 3). In addition to the technical and conceptual issues around big data analytics systems, there is a fundamental issue that inhibits their effective use in civic advocacy – they mainly support interactions that are individual rather than group or community focused. It can be noted that when people talk about public engagement or interactions within the wider concept of data and smart cities, it is often referred to as citizen engagement or citizen participation in the city; however, these interactions are always connected to (either directly or indirectly) a wider social context. The ways people learn, form understandings, make decisions and decide to act on them are always linked to environmental and social contexts (Dourish, 2004; Suchman, 2007), which also applies to interacting, interpreting and making use of data and information. Therefore, interactions with data and the systems that mediate it should be taken as *shared interactions* instead of individual ones, particularly in the context of communities using data to support civic advocacy and action. This requires providing resources and support that fosters the community and ensures sufficient conditions for translating data into activities for the local benefit:

> *The capacity and opportunity to successfully integrate ICTs into the accomplishment of self or collaboratively identified goals* (Gurstein, 2003).

Extending research in HDI to incorporate the *shared* and *social* elements of interacting with data in the community context requires a formal definition of what constitutes *shared data interaction* and an understanding of the influential components and factors needed to be considered when designing technologies and processes for CDI. The following two sections extend the CAF presented in Chapter 3, formally define CDI, and present a model of shared data interaction that helps designers discuss and understand how to design for the effective use of data by communities. The concept of CDI and relevant terminology are derived from the findings of the two case studies presented in Chapter 2, Chapter 3 and Chapter 4 and the existing gaps and challenges set out in the literature review in Chapter 1 and this chapter (Sections 5.2 and 5.3), in addition to drawing on existing critiques in ISD, SI, CI, social design (Ferreira et al., 2020) and community commissioning (Garbett et al., 2016).

5.6. Defining CDI

The concept of CDI merges fields such as ISD, SI, and CI to focus on the interactions between people and data in a complex sociotechnical network. In a way, CDI can also be considered to be a 'fragmented adhocracy' of IS (Stillman and Linger, 2009) with a more focused agenda. Similar to CI, CDI seeks to address both technological and social issues around the effective use of data. The aim of CDI is not to create a new theory but to provide a process or design practice for shared data interactions for the effective use of data in community contexts and for purposes of civic advocacy and action.

In this thesis, 'the search' is for the *social value of data* that has been previously collected or will be collected in the future, seeking to understanding the influential factors that dictate shared data interaction and the methods and tools that could improve interactions with data around complex social and environmental issues. What has been evident from the previous case studies is that when the social ergonomics (Brown and Newman, 1985) (e.g. how the community operates, how the systems can adapt to different contexts, what the values and norms of the community are, what the relevance of data is, and what the aims of people's actions are) around the use of data are not addressed properly or are ignored, the value of data and the ability for citizens to use and make use of it decreases marginally.

5.6.1. Models in CI

Hirschheim et al. (1996) proposed an object system class framework to the design of IS, which was adopted by Stillman and Linger (2009) for guiding the design of CI systems. As pointed out in the literature review (Chapter 1), this model has flaws in terms of its one-dimensionality and an inability to recognise the role of technology in other orientations except *instrumentation of control*. First, the findings of the previous two case studies in Chapter 2, Chapter 3 and Chapter 4 illustrate that technology cannot only enable the community to achieve control but

can also help organise, coordinate and make sense of situations, build knowledge and argue for and communicate ideas and issues. In the case of the SMS toolkit (Chapter 2), it had almost all the steps (i.e. orientations) (Hirschheim et al., 1996; Stillman and Linger, 2009) supported by technical processes or artefacts, whereas in the case of Data:In Place platform (Chapter 4), it provided technical scaffolding for citizens to access and interpret (i.e. sensemaking) data from expert systems and also add their own opinions (i.e. argumentation) of that data and issues to the platform through geo-indexed audio snippets. Furthermore, technology is never neutral nor present in particular isolation. Instead, it acts as an *'instrumentation of power'* across domains of change (Gurstein, 2012, p. 55).

Second, because of the overlap of similar orientations across the domains of change (Stillman and Linger, 2009, also), the model should not be taken strictly as cross-relational. According to Gurstein (2012), who has been credited as one of the founders of the CI field, the framework for CI is more iterative, dialectic and multi-faceted, acquiring more of a *'linkages, feedback and "feed-forward" mechanism'* (Gurstein, 2012, p. 52) between stakeholders, constantly moving from one orientation to another across domains depending on the context and situations currently present. These views are also shared by Moor (2007), who adds that because of the ever-changing nature of the communities, it is very difficult to have a definitive and replicable framework for designing SI systems for community use. Finally, the model for designing CI tools needs to incorporate critical social theory and aid in the dual agenda of solving engineering problems in relation to community activities Stillman and Linger (2009), borrowing concepts from IS and SI (Kling, 2000; Stillman and Linger, 2009) in order to achieve this holistic approach.

5.6.2. Community Commissioning

Another relevant framework for the effective use of data is the Community Commissioning Model of Garbett (2017), which is a linear model consisting of five stages: *express needs*, *establish demand*, *design and resource*, *produce* and *adopt and sustain*. This is highly relevant and could also be applied to Case Study I (Chapter 2) in this thesis, where citizens were commissioning environmental sensors to collect data about the local areas. Although the process has a bottom-up approach, this model only extends to the act of commissioning the infrastructures. Once the data is collected, the framework is exhausted. One could argue that the last step of the commissioning model of Garbett (2017), *adopt and sustain*, would guarantee the extended uses of the commissioned technologies. However, the act of 'sustaining' something often requires rethinking, redesign and even more cycles of development or commissioning. When the processes stop supporting the improving and changing of things to cope with the changes of the community (and the outside changes), the technology is not sustained any longer but is degrading. In the last step of the model, *adopt and sustain*, Garbett (2017) states that the model continues beyond this, putting emphasis on the importance of engaging supporters in the adoption of the commissioned infrastructure to establish a community that would sustain it. What this highlights is the importance of building *capacity*, new skills, resources and links that would successfully transfer ownership of the technologies and solutions to the community. With this in mind, the

model for community commissioning, and for the effective use of data in that matter, should not be flat and linear, driven by a particular process. Instead, it should take into account the changing nature of communities, issues and situations, and incorporate building capacities into every step of the process. Furthermore, these findings also merged from the evaluation of Case Study I, where a CAF framework was posited to discuss and understand the factors influencing citizens taking civic action.

Source: Dr. Andrew Garbett's PhD Thesis

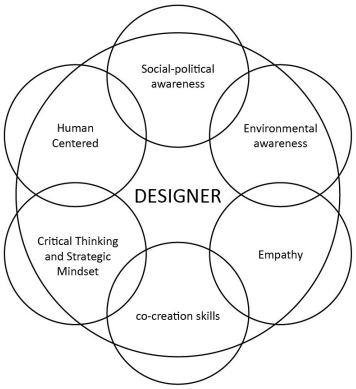
Figure 5.2 Community Commissioning Model (Garbett, 2017, p. 176)

5.6.3. Social Innovation Models

Participatory design, participatory action research (PAR), and its extension PADRE have links to social design and innovation (Bilandzic and Venable, 2011; Björgvinsson et al., 2010; Bradley, 2006; Ferreira et al., 2020; Haj-Bolouri et al., 2015), wherein designers working with communities usually apply any, or a mix of, these design principles with the aim of promoting social change, whether through small participatory efforts that could be scaled through networks or whether relating to building trust and capacities within the community (Ferreira et al., 2020). However, within the social design, it is often up to the designer or the social innovator to be the 'hero'; have the skills and knowledge to 'wield' the technology; be able to engage in critical thinking; have the awareness about social, environmental and personal factors; and be able to engage people in the design process (Ferreira et al., 2020). This is what Bradley (2006) called *'a new type of engineer'* or Bernarda et al. (2017) termed a *'designabler'*, i.e. those who have all the skills of an engineer and also of a social scientist. Figure 5.3a illustrates the skills of such a designer, whereas Figure 5.3b shows some of the factors, roles and contexts these skills fit in the social economy. Although a designer can certainly be an 'enabler', the change has to come from the community itself in order for it to be sustainable (Bernarda et al., 2017). However, the designer can become part of the network, either supporting, connecting, or enabling communities to instrument positive change. Furthermore, there should not be fixed roles in the new model of social change for CI; instead, it should be an ever-chaining mix of different types of collaboration patterns (Moor, 2007) that help build the capacities of community.

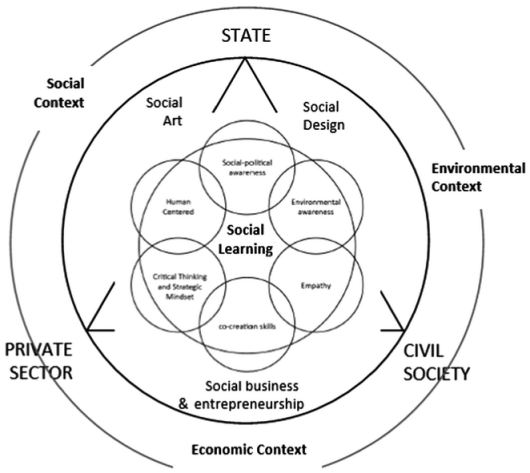
5.6.4. CDI Model

Analysing the models in CI, community commissioning and social innovation, and correlating them with the findings from the previous two case studies in (Chapter 2, Chapter 3 and Chapter 4) has identified the shortcoming of these models to effectively address communities' needs. Therefore, building on the CAF framework presented in Chapter 3, the model of CDI moves away



Source: Rocha et al. (2018)

(a) New skills for designer



Source: Ferreira et al. (2020)

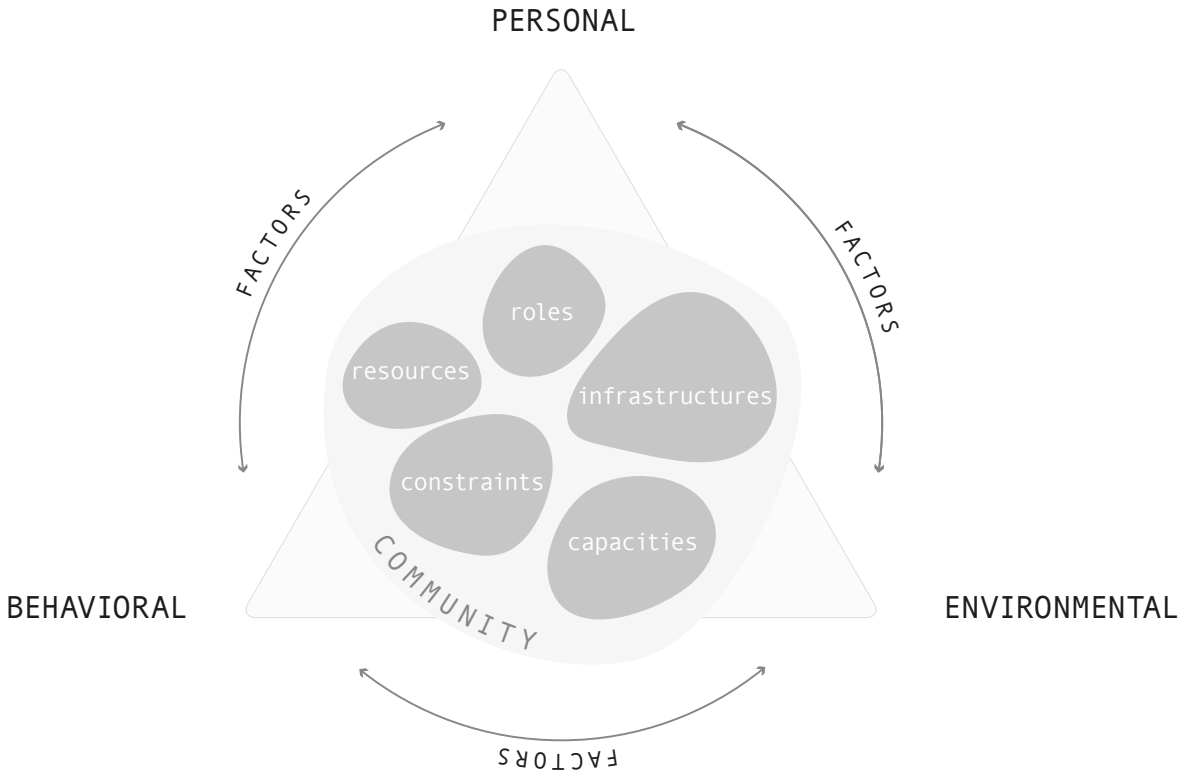
(b) Social economy

Figure 5.3 New skills for social learning and innovators in a social economy

from having a linear process or being one-dimensional to not having fixed roles or stakeholders, constantly changing orientations and adapting like the community itself, at the same time as being influenced by the three factors (i.e. personal, behavioural and environmental) of influence (and social learning as posited by Bandura (1977)), which will also become the domains of change that the community wants to influence and take action towards (Figure 5.4). To translate this model through the CI object class model (Hirschheim et al., 1996; Stillman and Linger, 2009), the objects are *roles*, *infrastructures*, *constraints*, *resources*, and *capacities*; at the same time, *personal*, *behavioural* and *environmental* are the domains of desired change and also the orientations (i.e. influencing factors) of change. However, the importance of each object and its role is defined by the particular context the community is in (i.e. orientation) and the changes that the community wants to implement. Furthermore, technology is not only present in just one orientation but is omnipresent across the CDI model, helping citizens achieve control and empowerment.

5.7. Processes of Shared Data Interaction

Having a formal model of CDI enables discussion on how designers could refer to this model to design for the effective use of data by communities for community action. Furthermore, it is now possible to look at the different components of the model and discuss what forms they could take, providing a lexicon to discuss these forms and their realisations in action. Within the CDI model, the key elements are identified as *roles*, *infrastructures*, *resources*, *capacities* and *constraints*. Table 5.1 illustrates the different forms key elements can take in CDI to achieve the effective use of data by communities. This table has been created through re-examining and adapting elements of the commissioning model of Garbett (2017) and the ideas of Gurstein (2003) relating to the active and 'effective use' of ICT and open data (Gurstein, 2011), which builds on the *'Access Rainbow'* of Clement and Shade (1998) and by analysing the findings of the previous two case studies. The remainder of this section focuses on expanding these key elements and providing



Source: Author

Figure 5.4 CDI model

some examples that arose from the research in the previous two case studies. It is important to note, however, that some of these concepts contain overlapping elements, which illustrates the interconnected nature of this research and the synergic way of working that is needed for shared data interactions and the effective use of data.

Roles	**Resources**	**Constraints**	**Infrastructures**	**Capacities**
Community Member	Time		System	Community links
Community Volunteer	Skills	Changes	Service/Carriage facilities	Connections
Community Activist	Assets	Actions	Social Facilitation	
Community Organiser	Knowledge	Agenda	Configuration/Appropriation/Unplatforming	Mobility
Local Charity Worker		Service access/provision		
LA Officer		Commitments	Governance	Diversity
Elected Member of LA		Complexity	Content and Formatting/Content services	Cohesion
Professional Data Analyst, Researcher, Instructional Designer		Advocacy		Social Capital

Source: Author

Table 5.1 Five elements of CDI

5.7.1. Roles

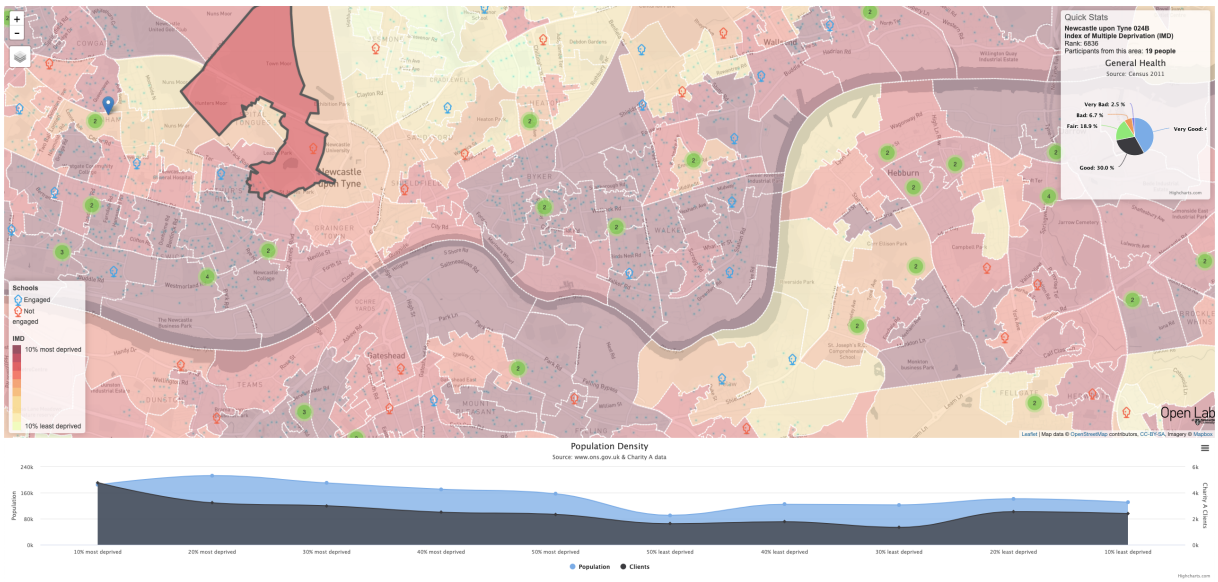
Chapter 1 outlined the different stakeholder groups involved in the development of open data ecosystems. However, when looking at the development of the effective use of data within communities in the wider context of CDI, a more focused look at different stakeholders, their importance and the roles they take is needed. Although these roles are defined explicitly, individuals can adopt multiple roles and even transition from one role to another depending on the state of the developments, the actions of other stakeholders and people, outside developments and also the influences of ongoing trends. For example, a member of the community can become a volunteer or an activist if they find an issue they think needs addressing or a cause worth fighting for in the community. Similarly, an organiser can be a member of that community and also part of or work for a local charity that aims to build social capital in the community. The first column of Table 5.1 shows some of the roles that members of the community took and were given by others in the previous two case studies within the context of CDI and the effective use of data by communities.

Professional Data Analyst, Researcher, and Instructional Designer

An underlining aim of this thesis is to look closely at the role of the data professional in the context of CDI and the effective use of data by communities. In this context, the notion of a professional data analyst is a person who effectively makes data usable for people beyond just data visualisation (although it might involve visualisations) to also come up with the right questions, know where to look for data and present data in a way that is comprehensible and actionable. This involves a great deal of technical expertise relating to information processing, while also requiring cognitive skills that take years of training. This is the reason why data science skills are in high demand and why companies are willing to pay a large amount for them.

Through the case studies in this thesis, the author acted as a data professional for these communities, providing technical support and skills for working with data. This often also involved making an initial analysis of the data and building a story around it to illustrate the power of using data. Such an example can be seen in Figure 5.5, where the author used aggregated client data from a local charity to illustrate how they tackled inequality in the community by carrying out targeted engagements. In order to create this application, the author had to develop a hypothesis, find the relevant data, process and merge it with the data from the local charity and use visualisation technologies to create a data-rich narrative, tasks that require skills both in software engineering and information processing (i.e. the skills of a data scientist).

Indeed, it appears that, at the moment, people need professional analysts or engineers to mediate these processes for them. Going forward, what is the role of a professional data analyst in these processes? Is there a process that makes citizens smarter and able to participate directly in decision-making that involves data at the city, county or even country level? To make it truly democratic, the challenge here is how those skills can become transferable to ordinary citizens to do it for themselves or can be built into systems so as to automate these processes for them. Designing and developing the Data:In Place platform (Chapter 4) was the author's first attempt



Source: Author

Figure 5.5 Data narrative application for the local charity

to build a system that would help automate and facilitate some of these processes; however, additional issues surfaced that were not solely linked to engineering and information processing challenges but were more linked to community problem-solving activities (Section 5.3), such as coming up with questions for data and then making that data useful in the context of community issues.

5.7.2. Infrastructures

In the context of CDI, infrastructures include the building and configuring of technologies and/or services to enable and support communities to make use of data. The obvious examples of infrastructures are **systems** and **services** that are usually developed for specific purposes and aims. For example, Data:In Place (Chapter 4) can be considered as a system and SMS (Chapter 2) can be considered as a service for environmental sensor commissioning. Furthermore, SMS is made up of multiple systems but also other smaller services and different **configurations** of existing services. In some sense, the tools that SMS used had already been adopted by communities and how they operate; the toolkit just added them as additional resources. This way of configuring has also been termed *'unplatforming'* (Lambton-Howard et al., 2019) or design *'appropriation'* (Dix, 2007) – identifying an existing community on a particular platform or tool and reconfiguring the system to serve the agendas and actions of the community. For example, SMS initially used people's smartphone's GPS (or any GPS) to attach locations to hand-held monitor readings and used Google Forms and Calendar to automatically schedule senor loans. Although, later, a bespoke Android application was developed (and phones added to the toolkit) to record GPS locations for the hand-held monitor, it still supported any type of GPS device to attach locations. Gurstein (2003) also calls this type service infrastructure *'carriage facilities'*, which refers to making sure people have the required supporting infrastructure to take up the use of applications. Similarly, there needs to be availability of **'content services'** to support

the application use. This type of infrastructure refers to the *'usability and locally contextual requirements'* (Gurstein, 2003), which means working with the community to understand what the local context would imply for data and tools. An example of this work was the user-centred and iterative design process of Data:In Place (Chapter 4) to make open data available and usable for the community. Finally, there is a need for **governance**, i.e. the legal, regulatory and policy infrastructure that would enable citizens to become active data users. An example of this kind of structure is the neighbourhood planning law part of the Localism Act 2011[16] in the context of what the community engaged in Case Study II (Chapter 4) was working on.

5.7.3. Resources

Resources involved in CDI and the effective use of data by communities relate to the skills, knowledge, time and physical and monetary assets to deploy, use and sustain community technologies or services. **Skills** are communitiesÕ abilities to perform particular tasks or master the use of some type of technology or process. **Assets** are physical and non-physical items and resources that the community owns. Both skills and assets can be individually owned and maintained, but if they become part of the community agenda, they can be categorised (but are often not transferable) as being community owned. A skill might be, for example, operating a hand-held monitor and developing a website, whereas assets could include personal smartphones, environmental monitors, money for printing flyers, community meeting facilities and webservers (Chapters 2 and 3). **Knowledge** can refer to the tacit knowledge of each member of a community and the focal knowledge of a community group within the context in which it works.

5.7.4. Constraints

All the other elements are bounded by constraints in regard to enabling the effective use of data by the community. In the context of CDI, constraints can be discussed in terms of the community trying to make use of data by leveraging resources in order to address a particular **agenda** and take **actions** towards civic **advocacy** (i.e. translating into activities or changes for local benefit). However, the community is always developing and has to adapt to constant **changes**, at the same time as bearing in mind that the community's other **commitments** are not linked to this specific agenda. The constraints are also applied to all the actions people take related to the **complexities** of those actions, for example, a community group organising a 'stand' and drop-in session at a community festival to engage the wider community with the data and issues (Chapter 3).

5.7.5. Capacities

In the context of CDI, capacities are probably the most important in terms of the availability, adaptability, sustainability, and longevity of the technologies and solutions for the effective use of data by communities. Both of the previous case studies illustrated the importance of **community links** and **connections** to making use of data in civic action, for example, knowing people in the

[16] https://www.gov.uk/guidance/neighbourhood-planning--2

local council or university to better organise efforts and target the use of data (Chapter 3), or the ability to bring together multiple charity organisations and local councillors to coherently (**cohesion**) work on a combined neighbourhood plan for the community (Chapter 4). However, there was also examples of people not being able to leverage data or sustain engagements due to a lack of **social capital** within the community and links to the skills and knowledge needed. Such examples of community groups can be found in Case Study II, where after the stationary monitor was commissioned and data collected, the engagement did not continue due to the lack of resources or an actionable plan for the data (Chapter 3). Additionally, it is important to note that **mobility** and **diversity** are also important factors that contribute to increasing the capacities of the community and its ability to reach different links and connections to tap into outside resources.

5.7.6. Overlapping Components

As mentioned before, there are components or objects spanning multiple elements of CDI, which can be seen as aspects that have the most value to communities trying to make effective use of data.

Time is probably the most precious *resource*, in addition to being the biggest *constraint*. Community advocacy work is mostly carried out by highly motivated individuals who volunteer their free time to bring about positive change for the community. Although the aim is to establish the effective use of data by communities, advocacy efforts often fail or succeed because of the individual (and combined) efforts of community volunteers and activists.

In the context of CDI, **social facilitation** can be referred to as the *capacities* and *infrastructures* that support the adoption or failure to adopt technologies and solutions to effectively use data by communities. Social facilitation means having the infrastructure for training, skills transfer and increasing social capital for the community. In terms of the dual agenda of CI and CDI, social facilitation aims to improve the community problem-solving agenda by facilitating, supporting, training, linking, and enabling the community to achieve control and empowerment.

Service access/provisioning spans across all the elements of CDI and both social and technical infrastructure. It is connected to access to appropriate tools (digital and physical) and training facilities; physical infrastructure; links to networks and other communities, often professional in nature; and sometimes even organisational structures. However, a lack of access to these services can become a constraint of CDI and can restrict the effective use of data by communities.

5.8. Practical Approaches for CDI

The remainder of this chapter builds on the existing challenges outlined and discusses ways that they could potentially be addressed using the model of CDI and the design of tools and technology, in addition to the use of transdisciplinary approaches and practices that come together to help unlock the social value of data for citizens and communities.

In the case of CDI, there is a need to take a more holistic approach to implementing strategies that would create a system capable of providing social value from data and enabling people to enact positive change in their communities. As research in CI already has a dual agenda (Stillman and Linger, 2009), there is a need for a dual approach. Furthermore, being a transdisciplinary research effort in the interdisciplinary field of HCI, a more suitable approach could perhaps be found by looking to a discipline outside of engineering and computing science. Looking at practices and methods of implementing strategies for influencing systemic and complex social issues led the author to the *dual approach*, which is widely adopted in the context of public health interventions (Bauer et al., 2014) and increasing health equity (Graham et al., 2016). The dual approach uses a model where strategies are applied to improve the health of the general population while, at the same time, using targeted interventions to address issues encountered by a specific targeted population. It follows the principle of *targeted universalism*, which states that alleviating disparities of the marginalised populations is essential for developing contextually relevant strategies for achieving universal goals and improving the wellbeing of everyone (National Collaborating Centre for Determinants of Health, 2013).

In the context of CDI and the effective use of data for community action, the dual approach means using contextualising to improve the ways people make requests, access and interpret data using computing systems and implementing strategies to help local communities deal with specific isolated issues. However, it is not impossible to work within the context of one group without working on interrelated issues and groups. Interventions are affected by other groups, actors, solutions, stakeholders, trends, etc. Therefore, when working on 'wicked problems', there needs to be a holistic approach that considers and bridges different disciplines, stakeholders and issues, or as Asad and Le Dantec (2017) put it:

> *Designers and researchers then need to identify where and how to bridge those diverging interests. One strategy, as described above, is through designing added transparency and feedback into advocacy support practices – not just the tool – so that different affected groups develop their own capacities to act* Asad and Le Dantec (2017).

This also relates back to building capacities through the *bridging* and *linking* of social ties discussed in Section 5.3.1 that are vital for the community problem-solving agenda of CI.

5.8.1. *Contextualising Data: Making Data Accessible and Usable*

The biggest technical challenge for CDI from the perspective of the effective use of data is making data relevant to people and their situations. Most of the available open data is collected to provide higher-level statistics for governing and administrative purposes. Although context-aware computing exists, it does not yet extend to humans interacting with abstract datasets though computer interfaces. Data needs to be anchored, grounded and situated in every facet of life; in other words, without contextualising structure, there is no value in data. According to Hakken (2003), in order to create or transfer knowledge through IS, they should aim to embed as much context as possible in order to counterbalance the abstractions needed to represent data through computer systems. Linking data with issues of importance and places of interest for local communities, however, sets out a series of challenging technical tasks. The following sections explore the ways data can be contextualised in IS by anchoring it in *place* and to *issues* and *inquires*.

Linking to Place

Chapter 4 focused on the ideas of making data accessible and usable through places identified by communities. This became the main premise of the Data:In Place platform and the *map-based querying system* that was developed to interface between people and abstract open datasets. Geographical data is, in most cases, represented in computer systems using the ISO 6709[17] standard to express it in coordinates describing geographic point location. Linking data to place is done by connecting metadata (i.e. other data that provides information about this data) to coordinates, or in the case of the Semantic Web, described through the relationship of that collected data to geographies (e.g. for example, UK official administrative geographies[18]). For example, census data in the UK is often published on the output area (OA) level (at least 40 households and 100 people, the target size being 125 households), which is the smallest official geography unit used by the Office of National Statistics. Table 5.2 illustrates different standards, techniques and methods of linking data to place. Although multiple different standards can represent a particular place, they need to be associated with something that communities can relate to in order to be effectively used by them. Working with different communities in Case Study II (Chapter 4), it became apparent that the things people relate to are often connected to personal stories or locations that have particular memories for them (e.g. streets, points of interest (POIs), and postcodes) (Manuel et al., 2017). The challenge is to translate these locations that people refer to into something that computers can store and reference against available datasets. This can be done through: *conversions*, which just convert one format to another; *best fit*, which tries to find the closest official geographical representation that it fits into; or translators, which maps conversions on the fly. The Data:In Place platform (Chapter 4) used all of these methods when matching boundaries to ONS Codes, converting them into ISO 6709 and mapping postcode data on the map by translating it using Javascript mapping libraries on the fly.

[17] https://www.iso.org/standard/39242.html
[18] https://www.ons.gov.uk/methodology/geography/ukgeographies

Additionally, having data in a standardised format also enabled people to use the Data:In Place platform to discover data from the commissioned sensors from the SenseMyStreet (Chapter 2) toolkit. Data:In Place used these standards, techniques and methods to contextualise data and make it available and useful; however, these techniques could also be used to improve the discovery and linking of data.

Standards in Place-Based Systems	Techniques	Methods
National Grid	Conversions	GIS software
ONS Geographies	Best fits	Javascript libaries
ISO 6709	Streets	WebGL
Projections and coordinate systems (WGS84)	POIs	
Representation of geodata (GeoJSON, WKT, KML/KMZ, Shapefiles)	Post Code	
Linked Geo Data	Translators	

Source: Author

Table 5.2 Standards, techniques and methods of linking data to place

Linking to Issues

Another way of contextualising or anchoring data is to link it to a particular issue that the community is facing. Issue mapping was used as the first step of the SenseMyStreet toolkit (Chapter 2) to have a purpose for the data once it was collected. In the case of the Data:In Place platform (Chapter 4), the first set of datasets incorporated into the platform directly translated from their relations to the issues pointed out at the community meeting. These links, however, were made by a professional data scientist (i.e. the author) who had knowledge of the available datasets and knew where to look for them. Future developments of issue anchoring should also consider ways these links can be made automatically or through the help of extended community links.

Linking to Inquires

Finally, data can be anchored with inquires, which can be linked to concrete questions or even to curiosities that people have. For example, when building the data narrative application (Section 5.7.1), there was a need for a hypothesis that would dictate what data was needed and in what form it was needed. Additionally, there was also a need for an understanding of the context in which the local charity worked (i.e. the data they collect) in order to produce something that would be of value to them. However, this part of the process of inquiry was undertaken by the author by translating the needs of the charity into hypotheses for data. In the context of CDI and building capacities, new ways and processes of inquiry should be developed that would actively engage citizens in these processes and enable them to set these agendas for data use.

5.8.2. Actioning Data: Making Data Useful

What was learned from the previous two case studies (Chapters 2, 3 and 4) is that data, regardless of its form, is not particularly useful for the community if it does not find a purpose. Without an actionable plan for data, the commissioned data was not utilised by a community (Chapter 3), and

without next steps of action, the open data accessible to people did not translate to any activities for the community (Chapter 4). On the other hand, when a community has an actionable plan, the data finds purpose and becomes useful (Chapter 3). This was also expressed through the analysis and presentation of the CAF in Chapter 3. The CDI model illustrates that there is a need to move away from the assumption that data is the 'silver bullet', meaning that if there is data, then information, knowledge and wisdom (i.e. knowledge that can by applied in action) reveals itself without additional steps, i.e. automatically. Furthermore, CDI has identified the key elements of the effective use of data that may lead to action and activities for local benefit. The CDI model for the effective use of data could also be used to help understand the gaps, needed resources and capacities to improve the usefulness of data for communities and aid in developing strategies for actively participating in decision-making processes.

5.8.3. Applying the Model of CDI

This chapter has discussed the dual agenda of CI and examined it through the model of CDI, in addition to extending it to examples of practical applications. Understanding how people were interacting with data through the built digital systems provided a deeper contextualised knowledge of CDI. Reflecting on the previous case studies in this thesis highlighted new challenges for CDI that revealed the importance of targeted work on the community problem-solving agenda. There is a clear need for not only technologies that would work for communities but also processes that would help build the social capital and capacities of the community. Currently, there is still a significant reliance on the work of the expert data professional to make effective use of data by communities. The model of CDI has been presented in terms of roles, resources, constrains, infrastructures and capacities, which are all influenced by the factors (e.g. personal, behavioural and environmental) that are simultaneously the domains of change. The model of CDI, however, has highlighted that there are far more social processes needed to guarantee the extended engagement by communities and the use of data for local benefit. Rather than building a new system or tool, researchers may consider focusing more on the ways we think about 'data as a resource' in general, as well as its use and reuse, to enable a paradigm shift to happen. Furthermore, applying the model of CDI to the design of CI projects enables us to further understand these concepts and configure the projects to achieve the desired outcomes for the community.

5.9. Summary

This chapter has presented a model of CDI for the effective use of data by communities for civic advocacy and action by reflecting on the previous two case studies in this thesis from the perspective of the dual agenda of CI (an engineering problem-solving agenda and community problem-solving activity), in addition to drawing on the literature of CI and social innovation. The CDI model was presented through five key elements, which are influenced by factors that are also the domains of change within the community. Expanding upon the conceptual model of CDI and the key elements within it, the chapter provided an ontology of components and discussed them through examples from the previous two case studies in this thesis. Lastly, the chapter provided examples of practical approaches for the CDI model that can be applied when designing for the effective use of data by communities. Although examples of components of CDI in action have already been identified from the previous case studies of this thesis, the next chapter will present Case Study III, which was conducted to apply the CDI model in full in order to provide empirical evidence for its usefulness and any potential issues arising from it.

Chapter 6. CDI in Action: Applying the Model of CDI for Civic Advocacy and Action

If you do not know how to ask the right question, you discover nothing.

— W. Edwards Deming

This chapter focuses on understanding the challenges of making data *useful* for the community and the methods that may help guide people through the whole process of achieving the effective use of data to support civic advocacy and action. Using the model of CDI established in the previous chapter, this chapter presents a case study where the model was applied in practice. Building on the practical approaches set up in the previous chapter (Section 5.8), a dual approach was taken to work on the engineering problem-solving agenda as well as the community problem-solving activities of CI, which included a redesign of the Data:In Place platform; the use of hypothesis setting and leveraging Right Question Institute's (RQI)[1] Question Formulation Technique (QFT) to aid people to better articulate their questions for data; and the use of challenge creation to help people come up with strategies towards action that make effective use of data. The case study was conducted through running data consultations in focus groups with key stakeholders from an NPO and through a Community Action Day with the local community in a workshop setting, where a combination of digital tools with supporting methods were trialled to address the dual agenda of CI. All the participants involved in the case study also took part in the initial design processes of Data:In place, as described in Chapter 4. The present chapter begins by highlighting the challenges and motivation for this study and explaining the dual approach in the context of CDI to apply this in a two-phase study, where two communities are trying to make effective use of data. In addition, the chapter presents empirical evidence regarding the use of CDI and the effectiveness of these methods to define actionable steps towards purposing data, in addition to starting to map out the supported processes needed for the effective use of data in advocacy and action, processes that have often been neglected by designers and information engineers building automated information technologies (AIT) for data discovery.

[1] https://rightquestion.org/

Related Publications and Acknowledgements

- Acknowledgements go to Dr. Jan Smeddinck for helping with the research design of the study and Thomas Maskell for helping to run the community workshop activities.

6.1. Motivations

Developing and deploying tools that promote better *access*, *comprehension* and *usability* of data does not mean that the data will find its way into actionable results. Data itself is not evident or inherently useful (Gitelman, 2013) to a person or group who is able to access it. In both case studies, SMS (Chapters 2 and 3) and Data:In Place (Chapter 4), the technologies built aimed to make data more accessible by anchoring it with place and local context so that people could make sense of it. With SMS, it was a case of democratising *data production*, whereas with Data:In Place, it was a case of democratising *data access*. Both studies made good progress towards helping people access and explore different datasets on their own terms. However, they might also have created an overload of choice for people rather than helping them come up with questions, queries and strategies for data use. From the case studies, it became evident that it is not only up to engineers to build better tools and interfaces for data, but there are also other, more human, factors that need to be explored to combine a more holistic approach for the effective usage of data in civic action. As a result, a model of CDI was proposed to illustrate the key elements of effective use of data by communities.

The case study builds on previous work conducted in this thesis related to community data production using the SMS commissioning platform (Chapter 2) and exploring the uses of OGD for local decision-making using the Data:In Place platform (Chapter 4), combining the knowledge gained from designing, developing, deploying and evaluating these data systems. Extending on the previous case studies and using the model of CDI, this chapter discusses what makes people go from 'I want to see all the data' to 'I want to see data that relates to a question I have', thus providing data for an actionable purpose. Such issues link to the open challenges set out in the previous chapter (Section 5.4) that need to be addressed around CDI to enable the effective use of data by communities. Building on the previous chapter's practical approaches (Section 5.8), this chapter explores how we might start linking data to questions that people have and actions they want to take. Through the use of CDI and the dual approach, this case study aims to improve the processes of inquiry for data by applying digital technologies and cognitive processes to use data in support of civic action in complex sociotechnical settings.

6.2. Dual Approach for CDI

The case study follows the overall approach of PADRE utilised in this thesis Introduction; however, to address the open challenges of CDI arising from the previous case studies, a more holistic approach was taken. In this regard, an adopted version of the dual approach (Section 5.8) was taken to improve the ways people make requests for data and use strategies to help local communities deal with specific isolated questions. This was realised by working closer with communities to implement a set of *mutually supporting strategies* (both technical and social cognitive) within the scope of and across two components, which included *Component 1* – systems approaches to promote the usage of data and build support for data discovery and *Component 2* – capacity building to make data actionable and increase a community's social

capital to take action. This approach simultaneously focuses on technical implementations that can be used by everyone who has access to the internet and social cognitive strategies that focus on specific communities. In terms of CDI, the dual approach aims to address issues in all five key elements of the model. Table 6.1 illustrates the usage of dual approach strategies applied concurrently for improving general approaches to working with and using data, focusing on specific targeted communities to make data actionable. The general interventions are linked to new technical features integrated into the Data:In Place platform, and targeted interventions are methods to help local communities understand their data needs and plan actions for data. Both also include ways that the community can increase its social capital and create new links outside its immediate community. The following sections will go into more detail about the different types of interventions shown in Table 6.1.

Component	Strategy	General Interventions	Targeted Interventions
System Approaches	Build support for scalable accessing and discovery of datasets	Anchored Data Requests	Hypotheses, QFT
Capacity Building	Improve effective use and actionability of data	Data Challenges	Challenge creation method

Source: Author

Table 6.1 Strategies for a dual approach in the context of CDI

6.2.1. General Interventions: The Redesign of Data:In Place

The initial prototype of Data:In Place (Chapter 4) was built to enable community groups and charities access and make sense of open data related to their local area. The design and development of the platform was largely driven by the questions and issues of a specific public, and the datasets that were first integrated into the system were a direct response to issues and questions posited by the local Neighbourhood Planning Group at their meetings. Furthermore, additional datasets were added through engagement with local charities and looking into ways their work could be targeted and evidenced. In later stages of the study, a *Data Request Feature* was added to the system (Chapter 4) that enabled people to put in requests for datasets digitally, without having to meet face-to-face with a researcher, making it more generic for everyone to request adding datasets. Although that solved the issue of physical presence, it did not make it easier for people to come up with requests or for the researcher to reply to them. All these issues added to the core issue of the open datasets being underutilised by communities. The Data:In Place platform was facing the same issues that were pointed out in the case of the What Do They Know website (Chapter 5).

Working on the dual approach's *general interventions* meant improving the platform to help unpack the challenges around connecting people's questions and issues to abstract datasets. Additional features were implemented into the platform that automatically added contextual data to requests, with the aim of improving data responses to people's issues and questions. These took the form of *anchored data requests* and *data challenges*, which will be described in the following subsections.

Anchored Data Requests

The Data:In Place case study (Chapter 4) illustrated that to make data accessible and usable by people, it needs to be made relevant to place, local issues and local activities in order for it to become useful. These can be seen as anchors for data that make abstract datasets more comprehensible to people. However, using anchoring can also be beneficial when requesting new datasets to be added to the platform, providing contextual information to better understand the context where data needs to fit and be made use of.

The platform's *data request* form provided people text entries for describing the issue they were exploring, linking the actual dataset or public authority who might posses the data and describing the issue by means of a story (Chapter 4). This works well if a person has an idea about what type of data they are looking for and how to access it. Then, it was a case of connecting the database with the map-based querying of Data:In Place. Furthermore, if the particular data provider used standardised vocabularies and descriptions of the data (such as linked data or JSON-LD) exposed through an API, these processes could be further automated. However, if a person had no idea what data might be useful for exploring a particular local issue, it made it much more difficult for the researcher to find those datasets to add to the platform. Previous work undertaken in this thesis on developing the Data:In Place platform (Chapter 4) within the context of the neighbourhood plan illustrated that people had difficulty making connections between particular issues and datasets that might inform them. They had an understanding that there might be some data available, e.g. by making a statement that 'the council has the data'; however, the type and format of the data was unknown to the community. Working closely with the community and attending their neighbourhood plan meetings enabled the author to make these connections because of the added context obtained from these meetings. This again points out the heavy reliance on the work of data professional to make these connections and find the data needed.

To improve data requests on the Data:In Place platform and add additional context alongside the submissions, a metadata collection was implemented that included map *boundary data* (Chapter 4) drawn by the request submitter, the submitter's *geographical location* (provided it was allowed by the person) and the previously *explored datasets* within the submission session. This additional metadata, which was automatically attached to the requests, provided some of the context that would have originally been obtained through the community engagements but was missing when people only interacted with the system, and it could potentially help respond to requests with more accurate and useful datasets for the community.

Data Challenges

An issue encountered in both the previous case studies (Chapters 2 and 4) was the lack of action in regard to using the data once it had been obtained. This was often an issue of not possessing the skills needed to work with the data in order to transform it into something usable (Data:In Place Chapter 4); however, it also related to the lack of planning or mapping out routes to actions that would be of local benefit. This meant developing attachments to the data and committing to

activities that would use it. These issues are tightly coupled with building community capacity and reaching out to supporting networks for additional support and skills, which is a big part of the CDI model. Some of these processes could be automated, but it was still evident that, in some cases, people would need support from professionals.

As a response, a *Data Challenges* feature was added to the platform, which enabled people to post challenges about *finding*, *accessing*, *producing* and *making sense* of data, in addition to then *taking action* with it. This served two purposes: first, to serve as a plan for taking action with data (i.e. making data useful), and second, as a way to increase the capacity and widening the community network by reaching out to people externally. Technical features that were added were: a Markdown[2] form and tagging functionality for submission forms (Figure 6.1), resulting in a data challenge info display (Figure 6.2). These types of technologies are commonly used for issue reporting on software development platforms such as GitHub[3]. Furthermore, the same metadata was collected and added to the *Data Request* forms, enabling a direct link to be made to the challenge (Figure 6.3). Additionally, a *focus bar* was added to the platform to always display the aim of the investigation when a particular challenge was selected. This focus bar also served the purpose of providing an issue focus in the *question-asking* activity in community workshops (Section 6.3.1).

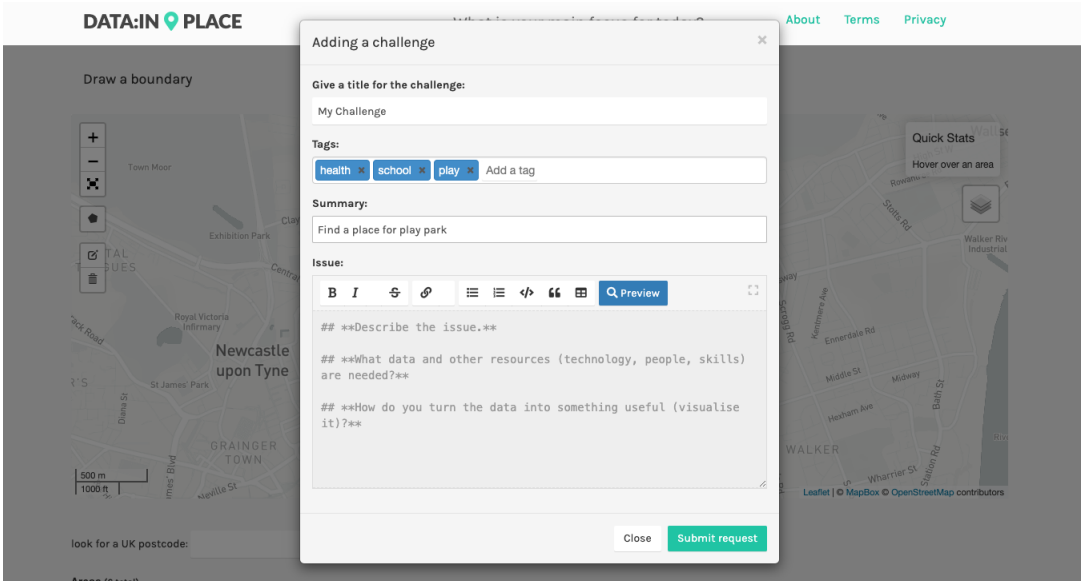


Figure 6.1 Screenshot of data challenge form

6.2.2. Targeted Interventions

In addition to focusing on the technical implementations by adding features on the Data:In Place platform, a parallel more targeted approach was also taken to aid people in making data useful for their purposes. This consisted of engagements with different groups of people interested in using data in community settings to explore or evidence issues and take action. This was carried out by guiding people in positing better questions for data and assisting in the creation of

[2] https://www.markdownguide.org
[3] https://github.com

Figure 6.2 Example of a data challenge

Figure 6.3 Example of a data challenge link

steps for actions based on the data. In addition to directly working with these communities and providing them professional support, the aim was also to develop methods and resources that could be used to help other people outside these communities use the new features of the Data:In Place platform and find actionable use for data.

Hypotheses

In order to know what types of data are needed or might become useful for a particular case, there needs to be a clear aim for the investigation. The case studies described in the previous chapters, although both working with different communities and datasets, had a recurring criterion for the successful use of data in action – data was found useful in civic action only when people knew ahead or planned out what they were looking for in the data, i.e. when they were asking the right questions from the data. Having a predefined agenda for the use of data turned out to be particularly important for community groups who were hoping to use data for civic action. This was also illustrated by the CAF in (Chapter 3). However, setting that agenda was a difficult cognitive task, which often required the knowledge of a professional data analyst. In a sense, it was a 'chicken and egg' dilemma that needed further investigation in order to depict the variables that influence the process of discovery and queues that would aid people in coming up with these agendas.

A common practice in scientific research projects is to first start with hypotheses that the projects are trying to prove (alternative hypothesis) or disprove (null hypothesis). This approach

is used in a variety of scientific communities and fields, from engineering to social sciences, which forces one to come up with a set of statements to explain a phenomenon or process. Hypothesis setting and investigation was one of the approaches trialled in the case study with participants from Charity A, who were not professional data analysts, to develop statements for discovering and making use of data.

Question Formulation Technique

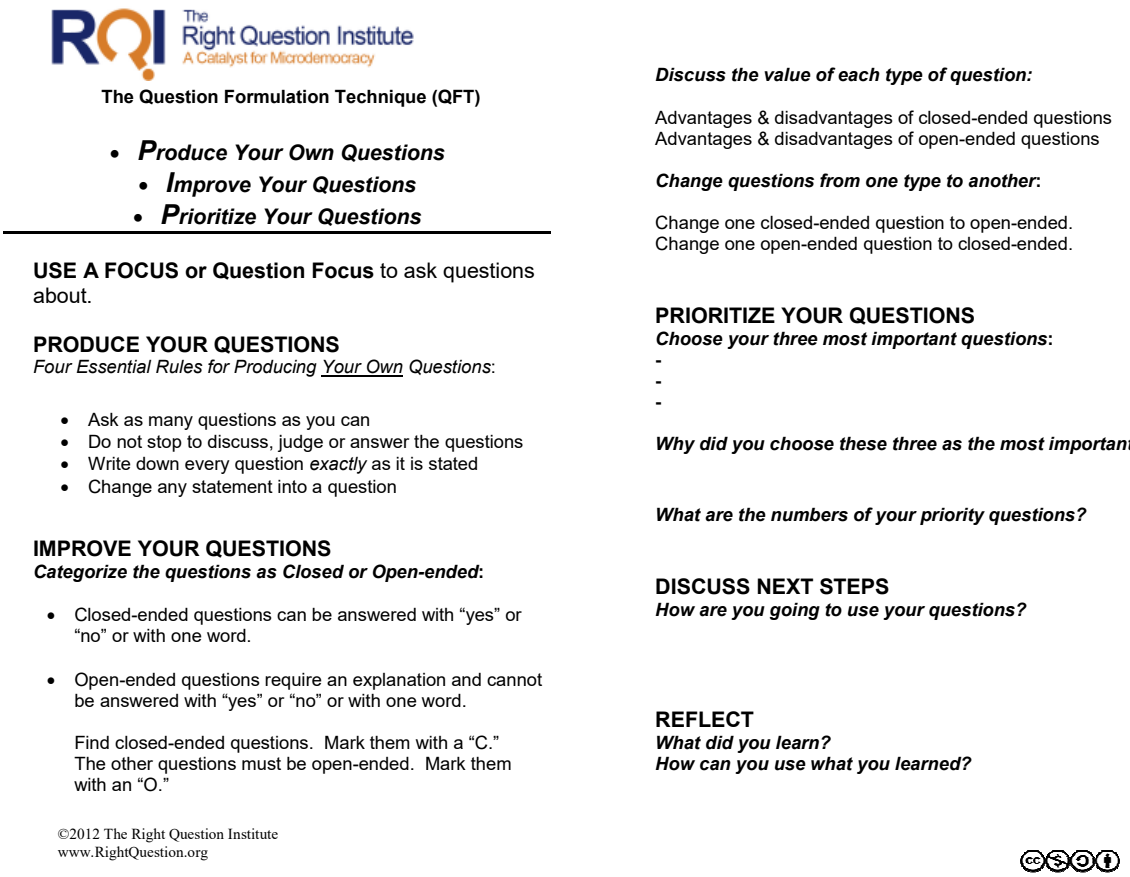
The scientific hypothesis approach or scientific approach may work in professional settings with people who are familiar with these concepts. However, the aim of this thesis is to help truly democratise these practices and make them work in community contexts where the scientific approach is not commonplace and most probably would not work. To make the process more inclusive and not serve the values of some stakeholders over others, there was a need to look at additional, alternative approaches. The closest things to hypotheses are questions, which people ask every day. The difference is that the questions are unknown and highly dependent upon the context and the way they are formulated.

The challenge was to come up with questions and issues that could be mapped to and explored by means of abstract datasets. There was a need for a better process of guiding people in asking questions that could be explored and answered using data. Fortunately, there was already an existing process, the QFT, which was initially developed by the RQI[4], to assist parents in better participating in their children's education process so as to prevent school dropouts. Furthermore, RQI's methods have now been used in a variety of contexts and settings, such as education, health and social care, community-based organisations, and in local democracy. In general, the aim of RQI strategies is to *'help all individuals learn how to ask better questions, participate more effectively in decisions, advocate for themselves, their families, and communities, and hold decision-makers accountable on all levels of democracy'*[5]. The QFT strategy, in particular, is a guided step-by-step process to helps people formulate, improve and use their own questions by applying multiple ways of thinking: (i) divergent thinking – linear learning and following instructions; (ii) convergent thinking – generating one's own problems and questions; and (iii) metacognitive thinking – reflecting on one's own cognitive processes. The detailed process of the QFT is summarised in Figure 6.4.

The Challenge Creation Method

Asking the right questions may help people better *access*, *interpret* and *make sense* of the data used in decision-making processes and perhaps even collect their own data to inform an issue. However, to effectively participate in these processes using the same data, people need to be able to come up with plans or strategies for civic action that *make use* of these datasets. A more in-depth approach had to be taken to explore and map out the steps that people need to take to move from questions and statements to planning out actions for data. Creating plans for action

[4] https://rightquestion.org
[5] https://rightquestion.org/what-we-do/



The Right Question Institute
A Catalyst for Microdemocracy

The Question Formulation Technique (QFT)

- ***Produce Your Own Questions***
- ***Improve Your Questions***
- ***Prioritize Your Questions***

USE A FOCUS or Question Focus to ask questions about.

PRODUCE YOUR QUESTIONS
Four Essential Rules for Producing Your Own Questions:

- Ask as many questions as you can
- Do not stop to discuss, judge or answer the questions
- Write down every question *exactly* as it is stated
- Change any statement into a question

IMPROVE YOUR QUESTIONS
***Categorize the questions as Closed or Open-ended*:**

- Closed-ended questions can be answered with "yes" or "no" or with one word.
- Open-ended questions require an explanation and cannot be answered with "yes" or "no" or with one word.

  Find closed-ended questions. Mark them with a "C." The other questions must be open-ended. Mark them with an "O."

Discuss the value of each type of question:

Advantages & disadvantages of closed-ended questions
Advantages & disadvantages of open-ended questions

***Change questions from one type to another*:**

Change one closed-ended question to open-ended.
Change one open-ended question to closed-ended.

PRIORITIZE YOUR QUESTIONS
***Choose your three most important questions*:**
-
-
-

Why did you choose these three as the most important?

What are the numbers of your priority questions?

DISCUSS NEXT STEPS
How are you going to use your questions?

REFLECT
What did you learn?
How can you use what you learned?

Source: https://rightquestion.org

Figure 6.4 Question Formulation Technique (QFT)

needed to integrate questions and statements as the starting point and get people thinking about the practical steps that needed to be taken to answer them. This also included figuring out what type of data might be useful and in what form and what kind of resources were needed to make it happen. This guiding process was called *challenge creation* so as to link it with the general intervention approach of submitting data challenges (Section 6.2.1).

The next section describes the study design taken to investigate the effectiveness of these approaches, accompanied by the general intervention strategies described in Section 6.2.1. This includes planning data consultation focus groups for hypothesis setting, leveraging *Ambit* (Johnson et al., 2020), a tabletop game for mapping and reflecting on issues, creating a modified approach to the QFT method to help people come up with questions that can be answered by exploring different datasets, and integrating challenge creation guides with QFT to be trialled in community setting. The aim of applying these methods in a workshop setting with a specific population was to enable people to make better use of Data:In Place's added features for data use (Section 6.2.1) that were part of the general interventions of the dual approach. In terms of CDI, these interventions are part of *social facilitation*, enabling an increase of capacities and utilising infrastructures to use data effectively.

6.3. Study Design

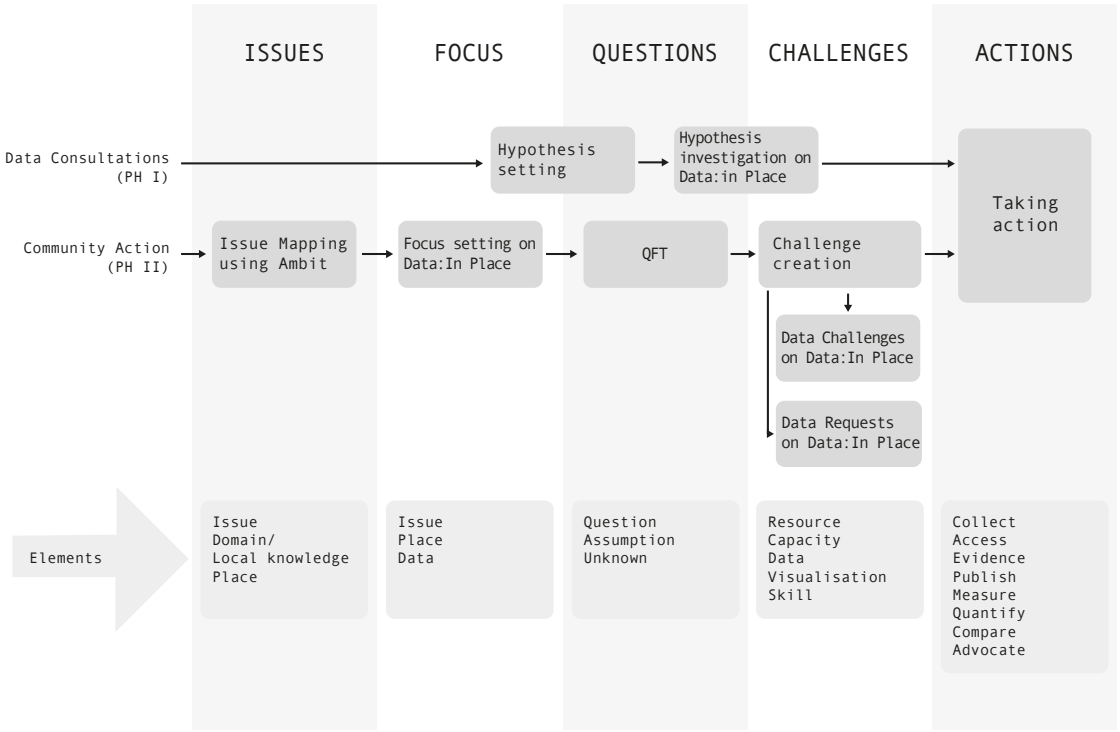
The evaluation of the case study methods and processes were undertaken in two subsequent phases:

PHI. (*Data Consultations*) involved running focus groups with key stakeholders from a large local charity (Charity A) organisation funded by many local organisations and lottery funds. Charity A worked on programmes and interventions around the health and wellbeing of young people, and they were also involved in the design of the Data:In Place platform (Chapter 4). Data consultations were run in a format where the author acted as professional support for the participants in regard to exploring and purposing datasets. This phase served as initial scoping to understand the cogitative processes that people go through in order to come up with actionable strategies for data use. Additionally, focus groups provided requirements for the further design of the Data:In Place platform.

PHII. (*Community Action*) focused on the evaluation of methods for helping people formulate better questions for data and deriving challenges from those questions. This phase built upon PHI and was conducted through a workshop where multiple methods were coupled to help participants explore and identify issues, posit problems and questions, and plan actions for utilising data in community context for civic action. Participants involved in this study were residents from the same local community who were participating in the neighbourhood planning meetings and the initial design of the Data:In Place platform (Chapter 4).

Figure 6.5 illustrates the two phases of the case study and the digital tools and methods used to explore ways to help people make effective use of data. The process consisted of five stages: identifying *issues*, setting *focus*, asking *questions*, creating *challenges*, and taking (or planning) *actions*. Furthermore, the figure outlines the important elements in each stage of the process that either acted as anchors or were considered as objects of change. Through these study phases, the following questions were explored:

1. How can people articulate the questions they are asking about the data?
2. Dual approach for CDI:
   (a) How to citizens transform data into something valuable (i.e. make data useful)?
   (b) What is the real value for citizens accessing these datasets?
3. Roles in CDI:
   (a) What role will non-professionals and (semi-)/professionals play in this?
   (b) What would the collaboration between professionals and non-professionals look like?



Source: Author

Figure 6.5 Tools and methods used in different study phases across the five stages of making data useful

6.3.1. Apparatus

Issue Mapping: CC in Community Action

CC is a method developed by Johnson et al. (2017) that uses *Ambit* (Johnson et al., 2020), an augmented reality audio capture tool build in Open Lab at Newcastle University to support deliberation in consultation processes for local decision-making in community settings. It has proved to be a valuable tool to structure the process and document the voices of people who traditionally would have not been included in these processes, making the process more inclusive and democratic (Johnson et al., 2017, 2018). A modified version of CC (CCmod) was trialled[6] in Case Study II (Chapter 4) to map local issues and gather people's opinions about them (Johnson et al., 2018). The location-based opinion data gathered through this process was then fed into the Data:In Place platform (Section 4) as another datasource generated by citizens themselves open for interrogation, side by side with official open datasets.

The workshop made use of the data gathered in previous workshops with the participants (Chapter 4) to start off with a set of issues that had been previously identified by the community. Figure 6.7 shows the CCmod map used in the workshop, with pre-marked locations from previous sessions. The issue marking activity followed the mechanics of CC (Chapter 4.6); however, a new set of prompt cards (n=2) were created to give more focus to the community action workshop: (1) *wild card*, which enabled a new location or new information about a previous location to be added and (2) *second that*, which enabled people to prioritise markers that had set a focus on

[6]CCmod was designed in collaboration with Dr Ian Johnson and the author and built by the author.

Figure 6.6 Modified CC map for Ambit technology with pre-marked locations

issues. Additionally, participants in each group had a limited set of turns to add things to the map due to the time constraints of the workshop. This was determined by additional stickers on the cards that they had to place on the map, as shown in Figure 6.6.

Focus Setting: Data:In Place in Community Action

The second activity used in the community action workshop following the *issue mapping* was *focus setting*, which built on the mapping exercise to help the group focus on a specific issue. Groups had to choose an issue from the map that had the most support and consolidate that issue using the Data:In Place interface. In addition to discussions in the group, people could also use the Data:In Place platform to listen to location-based recordings from previous sessions recorded using Ambit (Chapter 4). After the group reached consensus through discussing and listening back to previous opinions, they used a mapping functionality to mark the issue on a map, similar to how they had done previously on the paper map. Issue mapping was a new feature that was specially built and added for running the community action workshops.

One other added feature integrated into the Data:In Place platform for the workshop was a *live map view* using WebSockets[7], which enabled the participants to see what other groups were doing on the map in real time, thus promoting reflection and collaboration. This also guaranteed that the groups would work in different areas and on different issues, covering more ground. A workflow of focus setting in a two-group session is illustrated in Figure 6.8. All the data inserted

[7] https://developer.mozilla.org/en-US/docs/Web/API/WebSockets_API

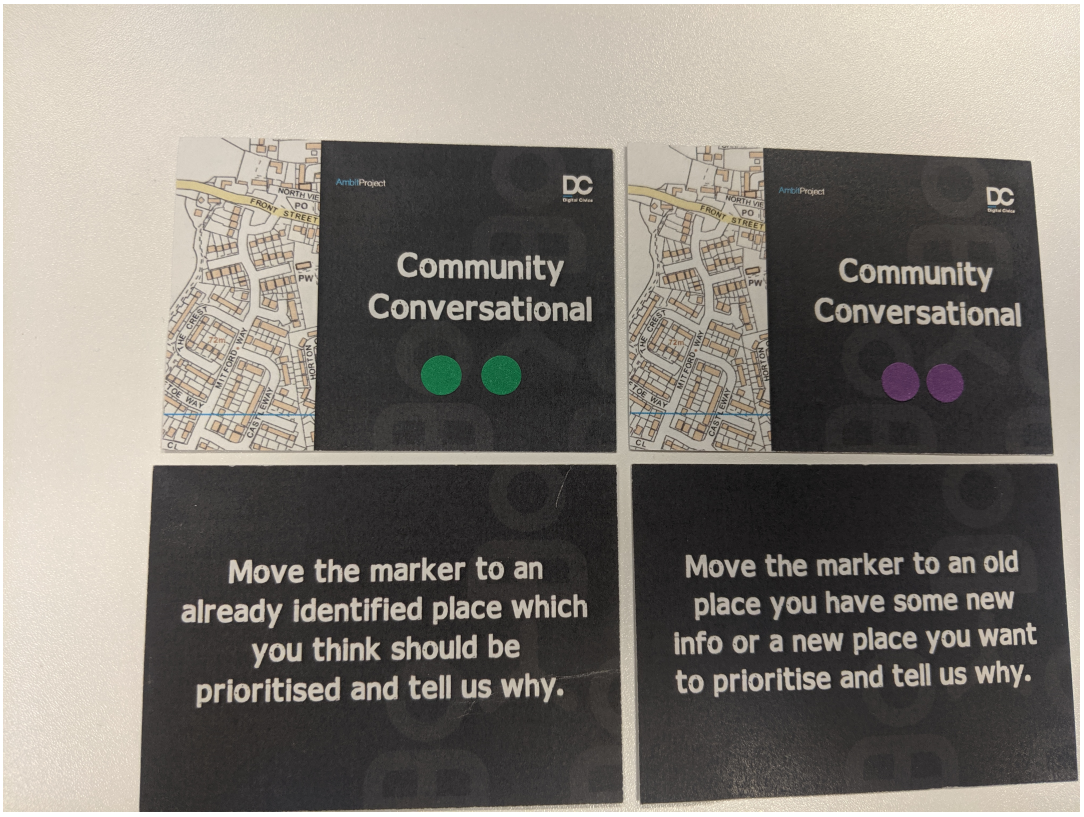
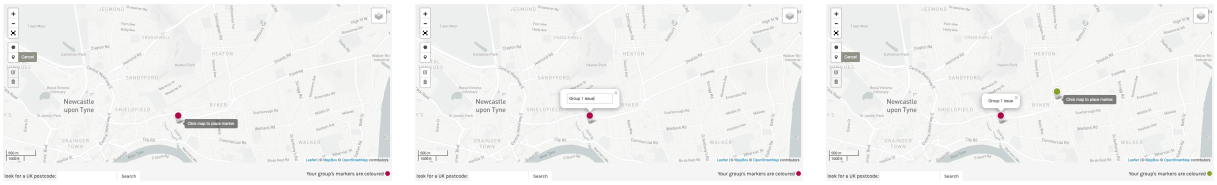


Figure 6.7 Modified CC cards



(a) Group 1 locates a place for an issue

(b) Group 1 maps and describes an issue

(c) Group 2 sees Group 1 issue and can map their own

Figure 6.8 Issue focus workflow

into the Data:In Place system was saved for further analysis and recall, which includes boundary data, mapped issues and their geographical coordinates.

Question Asking: QFT in Community Action

The third method used was the QFT, which was appropriated as a design interaction method in a workshop setting for helping people find uses for data in community action. Previous activities helped to feed into the QFT method by setting a focus for the question asking (Figure 6.4). The QFT had already been used in multiple contexts; however, it had never been trialled as a design method in PADRE and was a novel engagement method in HCI. This part of the community workshop was run as a guided exercise, where the QFT was explained step-by-step using an example of 'rock climbing' (chosen by the thesis author). Additional worksheets were made and given to the participant on each of the steps (Appendix B), which were new configurations of the original QFT worksheets (Figure 6.4), to make them more focused on community action.

This guaranteed that the participants followed the structure of the method, keeping participants focused on only one task at a time and making the assessment on each stage easier. As a result of this activity, people had a top priority question that they selected as a group from a pool of questions created by leveraging the QFT method.

Challenge Creation in Community Action

The final method used in the community action workshop was *Challenge Creation*. In the context of the effective use of data, a *data science challenge* falls under categories such as data access, collection, data processing and analysis, augmentation, visualisation and contextualisation. The starting point for the challenge creation activity was the top priority questions chosen by each group in the previous activity. The activity itself was set up for people to transform these questions into challenges for data through thinking about what data and other resources (e.g. technology, people and skills) were needed. People were given additional worksheets (Appendix B) with prompts on them to help them think about different aspects of creating a *data challenge*:

- What do we already know?
- What data/resources do we need to answer these questions:
  - What is out there/What exists?
  - How to get access to it?
  - Do we need to collect something ourselves?
- How to turn the data into something useful?
- How to make data or the product of data actionable?

Participants needed to think about and document things they already knew, resources and data needed and how to access them, what they could do themselves and what kind of help they needed externally. Similar to data consultations, people also had to think about how they would then bring that data to life through visualisations and how they would use these visualisations, or the knowledge gained from them, to take action and advocate for change in the community. This was done through open discussion within the groups, where participants also used provided worksheets to sketch their ideas. The final step was for participants to document their challenges in the Data:In Place platform using the data challenges functionality (Section 6.2.1).

Some of the challenges could be taken up by participants themselves without any external help, but some needed help from, for example, a data analyst or policymaker. However, the challenge creators would still be involved as the owners of those challenges, providing local insight and collaborating with the people who take the challenges up. Finally, the participants were given an opportunity to be involved with the project going forward and indicate whether they wanted to collaborate or take up any of the challenges. They could do so by providing information about skills that contribute and their preferred communication channel (e.g. email, post, face-to-face, FB, WhatsApp, and video calls).

6.4. Data Collection and Analysis

This section describes the data included in the case study and how it was collected and analysed. Data collection and analysis is broken down by each phase of the study, where the second phase of the study, although run with different participants, built on the lessons learned from the first phase.

6.4.1. Data Consultations: Phase I

As mentioned previously, throughout the research of this thesis, the author acted as a professional data analysts for the communities that were engaged. One of the aims of this was to understand the complexities of the cognitive processes involved when making use of data. Another objective was to investigate what role the data analyst should play in processes of the effective use of data by communities and how far the professional can be removed from these processes or replaced or mediated by digital systems, such as the the Data:In Place platform (Chapter 4) being designed and built to make open data accessible and usable by communities.

Data consultations related to continuing engagement around the Data:In Place platform and the effective use of open datasets, involving key stakeholders who were included in the design processes of Data:In Place from the early days. Data used in this phase was collected from three focus groups over a period of four months. Participants (n=3) involved in the data consultation focus groups were members of Charity A (Section 6.3). Each participant had their specific role in the organisation – outreach officer (OO), learning and skills manager (SM) and programme officer (PO). The outreach officer (OO) took part in all three focus groups, while the others joined a particular focus group at specific stages of the data consultation process.

Focus groups were set up as guided ideation activities, where participants followed a worksheet (Appendix C) to explore how could they use data in their practices. Data consultations consisted of three parts: (i) *Hypothesis Setting*, (ii) *Hypothesis Investigation*, and (iii) *Taking Action*. Activities were run using a *think aloud protocol*, where participants actively provided feedback on the processes. For each of these processes, the author acted as a professional data analyst helping to find datasets that could be explored for their hypotheses and helping participants make use of the Data:In Place tool to access and visualise data.

Data in this phase were collected in the form of audio recordings from the three focus groups, which were fully transcribed using a professional transcription service and participant worksheets that were filled in during the activities. Additionally, a screen capture was used to record participant activities using the Data:In Place platform in the session, producing visualisations out of the data that correspond to their hypotheses. The transcribed audio was used in conjunction with the worksheets as content for constructive reflection to map out interactions and report on the focus group findings.

6.4.2. Community Action: Phase II

This section describes the methods used in the *Community Action Day* workshop (Section 6.5.2), the data collected and the analysis undertaken to report on the findings of the workshop. This case study was part of a longitudinal engagement undertaken in this research that involved participants from the same group of people from Case Study II (Chapter 4).

A three-hour workshop was run with participants (n=8) who were all members of a local neighbourhood planning group, which included local residents, volunteers and employees working in small local charity organisations (Charity B) set up to improve community cohesion and resilience. Participants divided themselves into two groups of four and sat down at two tables, which had been set up to run the activities. The workshop consisted of four activities: (1) *issue mapping* that made use of *Ambit* (Johnson et al., 2020) technology; (2) *issue focus* that used redesigned the Data:In Place platform with special workshop features for collaboration added to it; (3) *question asking* guided by a modified version of the QFT method; and (4) *challenge creation* for coming up with actionable uses for data. Data in this phase was collected from multiple sources, including audio recordings of the workshops, interactions with two digital systems (Ambit and Data:In Place) and paper worksheets filled in by participants (Appendix B). Each system and its method of data collection will be described in more detail in subsequent sections.

The aims and goals of the workshop were the following:

- Understanding how the QFT method works in community action (i.e. as a design workshop tool in a community setting) and with a mixed group of people.
- Enabling people to achieve empowerment through better question formulation and by understanding how they can take action.
- Understanding the roles within CDI and how links can be brokered between professional data analysts and the community.

Audio recordings were fully transcribed using a professional transcription service, and qualitative analysis was conducted on the transcribed audio using thematic analysis (Braun and Clarke, 2006).

6.5. Findings

The following sections present the findings from the analysis of transcribed audio recording and observations from data consultation focus groups with members of Charity A (6.3), and from a Community Action Day workshop run with local residents in a community. Findings are represented following the different parts of each study phase, as illustrated in Figure 6.5.

6.5.1. Citizens Taking Power of Data: Data Consultation Focus Groups

This section presents the findings from three data consultation focus groups with stakeholders from local Charity A. Findings are presented using the three parts of the focus groups (hypothesis

setting, hypothesis investigation and taking action), which are taken as the main concepts and reported upon using pseudonymised quotes from the dataset to illustrate discursive processes in temporal order. The sections that follow will describe these concepts and the themes around them that surfaced from the think aloud activities, helping to chart out the cognitive processes that people go through to drive actionable results.

Hypothesis Setting

The hypothesis approach required participants to come up with a null hypothesis (HO) and two alternative hypotheses (H2, H2) as the initial part of the data consultation (Appendix C). Hypotheses are the preferred way of conducting experiments or investigations in the scientific community at large. However, what was observed right away was that the participants found it difficult to come up with well-formed hypotheses, and they fell back on asking simple questions, which was more natural to them. For example, when trying to come up with H2, the programme officer (PO) suggested following:

> *Has the wellbeing of residents improved in the city, I suppose, or people that have interacted with the campaign somehow?* (PO)

To which, the outreach officer (OO) added:

> *What is their mental and physical wellbeing like now? [...and] What is it like after we've finished?* (OO)

Although all of these were valid questions, they were not in the form of a hypothesis, and they lacked detail for starting the investigation, e.g. simple details like a time period, place and population, but also a specific measure of, for example, wellbeing that could be looked at using data. In order to move into the next phase of the consolidation, there was a need to narrow down on such details. Through the process of guiding questioning from the author (later replaced by the RQI method), the participants were able to come up with all three hypotheses (Appendix C).What also surfaced was the fact that hypotheses were often derived through assumptions or were linked to local and domain knowledge obtained through years of doing work in the community. Here is an example of the learning and skills manager (SM) narrowing down to potential metrics to investigate school performance measures relating to Charity A's work.

> *I can't imagine we'd have much impact on maths, but I suppose you could argue that if they're improving their attendance.* (SM)

Or another example of local knowledge from the OO:

> *[...] a hypothetical example. If we know that, if we already know [people's health in one part of a city] is great, it's perfectly fine. Let's go target [another part of the city], for example, instead.* (OO)

This shows that the proposed data-information-knowledge-wisdom (DIKW) hierarchy of Ackoff (1999) does not apply here. In this regard, data is not put into use to reveal new knowledge but is used to enforce arguments that are already known. It is also notable here that these snippets of local knowledge were essential for coming up with hypotheses that would provide actionable insight for Charity A and retain the ownership of investigation results.

When discussing different datasets that might be used to investigate the proposed hypothesis, the participants were convinced that data must exist somewhere amongst the official open data sources:

> *I know the data is probably out there. Attendance, homework [...] Attendance, especially, will definitely be out there. We'll definitely be able to get hold of [it]. Behaviour and homework will probably be recorded, but might just be kept to the individual schools.* (OO)

Additionally, there are organisations that participants were aware of that have been specially set up for collecting this type of data:

> *Maybe it's an Ofsted thing. Ofsted might have hold of all these things because Ofsted go in and they scrutinise schools and look through all the paperwork.* (OO)

However, knowing that there might be data available did not help them access and use it. Participants felt overwhelmed with the amount of effort and expert knowledge needed to find the datasets that would be of interest to them. SM shared experiences of trying to gather information for funding bids and navigating all that data:

> *How do you find it, and how do you access it, and how do we then understand some of them? I've looked at datasets before, trying to find information for bids, and they're just a nightmare. Some of the stuff you get from the Department of Education is horrendous. For some of the stuff, you need do need to be a data scientist to unpick it.* (SM)

However, putting down concrete hypotheses made participants think about how they operate as a charity organisation and whether they should also try to collect some of the data, i.e. reflecting on their own practices and trying to get data more relevant to them. Current practices of the organisation were limited to a couple of surveys they asked the teachers to fill in, but thinking about the future, participants started to consider new datasets that could link to their hypotheses.

> *It's something that we could record when we're in schools. If we're looking at the first hypothesis [H1], where we're saying that [Anon Programme] improves these things, we can definitely ask these schools, 'Can we have your attendance information, your behaviour information, and your homework information while we've been here?' [...] There might be a standardised way in which we collect that or we ask the schools to give us that data.* (OO)

In addition to thinking about how to standardise these practices across the different programmes the organisation was running and perhaps also across the charity and voluntary sector to promote data sharing and collaboration, participants were also conscious of not loading the responsibility of this onto teachers, who are already overworked, and were leaning towards automated systems and processes.

> *If there is some sort of online portal that all schools have to upload to, then great. [...] The teachers don't want any more work. If they need to start telling us behaviour, attendance, and homework of each single child in the school, that's a big pain.* (OO)

The hypotheses helped participants think about the bigger picture and all the different data sources required. For example, in order to compare and relate the data that the organisation produces to the schools that they do not engage with, they would also need data about the general population:

> *I'm just thinking, for the second hypothesis, obviously we can't go in and ask a school for that if they don't want to buy into our primary schools package in the first place. That second hypothesis would have to be collected from data that's already out there.* (OO)

This also made the participant reflect back to the original rationale of the project:

> *I suppose my original idea with data and place was: 'How can I compare what I capture to the general population?' If I know the measures of the general population, I can relate my data to there and say that my programme is working, it's great.* (OO)

The statement above illustrates that there are notable distinctions between *open data* (i.e. available data), *accessible data* and *data in context* (i.e. usable data). The latter two should be the preferred forms of data and the ones that could potentially be made useful for communities to take action. Being able to think about data in different formats got the participants thinking about the ways that it could made useful for them through visualising it.

Data could provide Charity A with the confidence that they are doing good work and having an impact in the community, but in order to show that to others, it needs to be curated. Participants felt that the types of visualisations needed would depend on the target audience and purpose. There are differences in the way you communicate these concepts to different people. For example, for official documents, the participants would need numbers and statistics, but for engaging the public, they would need 'easily digestible infographics'.

> *Obviously, I know we discussed who to show the data to. You've got certain people which you'd want to show a bar chart to. They're probably the stakeholders and the people who are funding the project, for official documents. People who are more visual, it might go nicely into presentations and things for schools and showing teachers.* (OO)

It was important to the participants that the visualisations would have all the facts but would be communicated in a clear way: 'Like I say, something that people look at and they go, "Ah." They can instantly see what it is, anyone can understand it, and it's got the facts.' (PM). Probably the most important aspect of the visualisation for the participants was that it needs to incorporate the identity of the organisation and the things they feel strongly about, 'so it's identifiable that it's us that's presenting that infographic, not just anyone else.' (PO). During the consultation activities, participants used the worksheets (Appendix C) to provide an idea of how it would look. Here is OO commenting on the sketched visualisation:

> *I think it's quite nice to incorporate the [Anon Sports] club into a lot of this information that we portray. It might be the amount of seats in the stadium or it might be using a percentage of the pitch, which has been [...] Taking the data, or taking the stats, and putting it into something that's [anon sports] related and digestible.* (OO)

Setting hypotheses and thinking about the data and resources needed to investigate them and the types of outputs people would expect created a basis for the second consultation. In order to investigate the hypotheses, some of the pre-work had to be done by a data professional (i.e. the author). This included finding the relevant data sources, pre-processing them (if needed) and making them accessible thought the Data:In Place platform's map-based query system. These tasks could have been fully automated through data requests built into the system or carried out by any data professional leveraging the collaborative open-source system; however, for the purpose of this case study, it was done by the author, who acted as the data consultant for Charity A. As the author had also been working closely with the organisation for a long period of time, a level of trust had been established that enabled the charity to share some of the operational data needed to carry out the pre-work.

Hypothesis Investigation

Touching on the role of the data professional in these processes, a question arises as to how far the professional can be removed. Can people leverage technology to do it themselves or is there need for human experiences? At the start of the engagement with Charity A, a data application was created by the author without any input from the organisation (Figure 5.5 (apart from them sharing some of their organisational data). This data narrative was the output of a professional's work and represented an insight into the way the organisation operates; however, the usefulness of it to the charity was not necessarily guaranteed. The aim of data consultations was not only to replicate these results using participation methods and digital tools, but also to seek ways in which these processes could be democratised to represent the voice of the charity (and citizens as a whole).

Going through the process of hypothesis setting, coupled with ideation activities fostering data use, provided crucial input from participants to set a precedent to guide the investigation in the second data consultation, which aimed to prove/disprove the hypotheses set by the participants. In this stage, only the Operation Officer (OO) was present at the consultation. With

guidance from the author (who acted as a data professional), the participant used Data:In Place in a think aloud activity to access official statistics and merge them with their own organisational data in order to produce actionable outputs. This section describes the investigation process through activities and representative quotes from the session in chronological order. This helps to illustrate critical moments that helped the participant understand and use data to respond to the hypotheses.

The participant started off by drawing the boundary around a place of interest, *'Say we're looking at Newcastle [...]'* and using the map query system (Chapter 4) to narrow down the request: *'Yes. That's [Hebburn north]. We'll get rid of Hebburn. [...] Camperdown, get rid of that. I think Camperdown is in North Tyneside, as far as I'm aware. Is this in alphabetical order?'*.

Once the desired area was selected, the investigation continued by selecting the official data of school locations – *'We'll look at schools for now, yes?'* – to get them to appear on the map. To prove/disprove the set hypotheses, the participant needed to look up and mark schools in which they had already done engagements.

> *Newcastle, let's try Hawthorn Primary. [I know we've been there]. H-A-W-T-H-O-R-N. [...] Oh, it turns them all to blue. [...] Byker, is that in that area? Yes, cool. St Lawrence's will be [...] It'll be S-T and then there might be a dot, St Lawrence's. Or even just Lawrence. Oh, there it is.*

The above interaction marked the first time that the participants were able to replicate the same results that was initially achieved programmatically by the author (Figure 5.5), thus opening up possibilities for further investigations in the future. However, schools often had similar names, which made this process more complicated. This promoted the participant to go back to discussing the ways the organisation records their own data. One of the options could be to use unique identifiers to improve cross referencing the schools.

> *I suppose we could. I don't know if the school will have those readily at hand. [...] I'm just thinking, if we've got our coaches there in the school, and they ask the secretary what their unique reference code is, they're probably not going to know. [...] When schools sign up, to whatever, we just ask them for their unique ID as they're filling out the form. [...] If there was a tool that I could use to input all of those schools and find out all the reference numbers, then that would be great.*

The next steps were to select official data that included Ofsted[8] statistics about school performance measures (e.g. reading scores, maths, and attendance) to retrieve data about selected (engaged and not yet engaged) schools in the area. Looking at examples of the engaged schools, we could already see that there were some differences between the ones they had not been in. This was encouraging for the participant; however, in order to use that data, it needed to be visualised. The participant chose *absence trends* to visualise using the RAWGraph (Chapter 4) graphing

[8] https://www.gov.uk/government/organisations/ofsted

functionality built into the platform. Being able to easily graph and see the visualisations made the participant think about new alternative representations of the statistics.

> *If you were looking at the whole UK, it would be 88% are steady [absence trend], 6% are increasing. Then, what's that? 6% are decreasing, or something like that. Then, compare that to our engagement, you've got 90% are steady [absence trend], 8% are increasing [and] 2% are decreasing. Then, we can say that, on average, our engagement has got either steady or improving schools.*

The following thread represents the conversation between the participant and the author after making the above visualisation, which can be seen in Figure 6.9:

> Author: *Their absence rate is decreasing.*
>
> Participant: *Oh, excellent, right. So they're coming to school more?*
>
> Author: *Yes, they're coming to school more because they are interested in the things that are happening. [...] That's already something we're flagging.*
>
> Participant: *Yes, definitely. That was the point, wasn't it? That was the hypothesis. [...] Yes, wicked, yes, I mean I'll show that to whoever is interested back in the office. I'm sure that will please some people.*
>
> Author: *In terms of simple statistics, we've proved it. We didn't run any specific tests, but kind of [...conversation-starter].*
>
> Participant: *Yes, we're not bothered about p-levels [p-values] or anything like that or whether or not it satisfies the criteria. [...] That visually, to us, looks good, that's all we care about.*

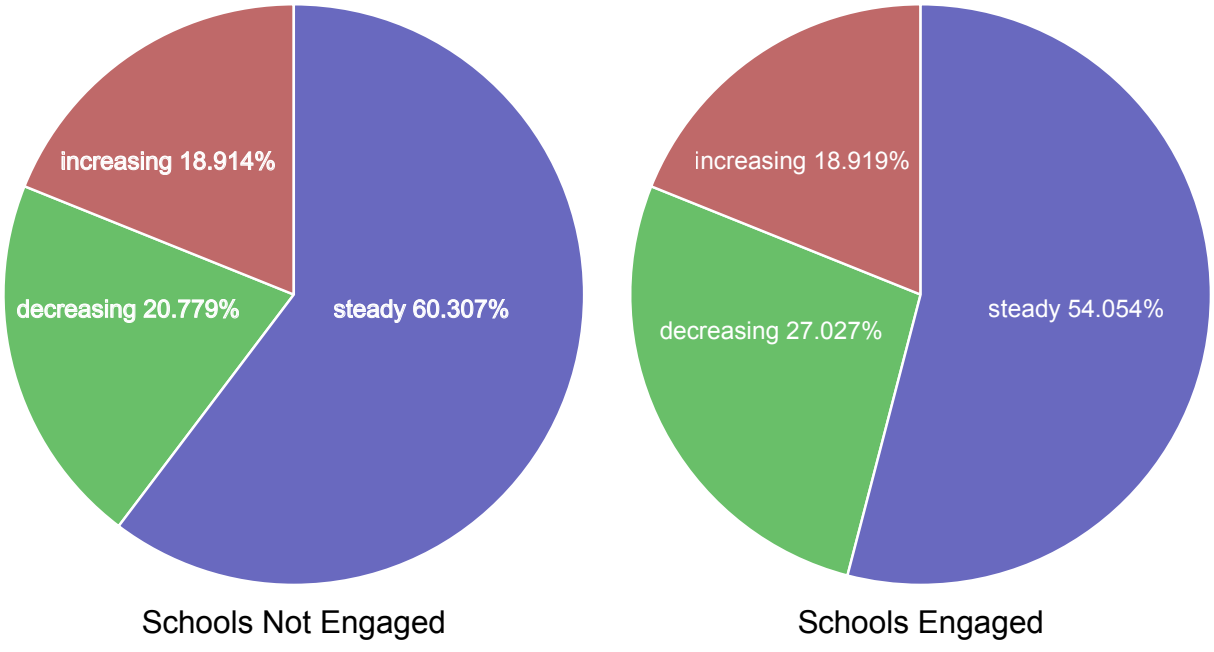


Figure 6.9 Absence trend graphs made using Data:In Place

The above marks another milestone towards democratising data science practices by enabling citizens to make sense of and visualise data in context through the use of digital tools. The Using Data:In Place platform participants were able to utilise data to give answers to the hypotheses set by themselves, similar to the professional work done by the author previously (Section 5.7.1). This was also a step closer to removing the researcher from the equation and letting the citizens do it by themselves.

> *That's the idea, isn't it? There is no point you sitting here showing me, you doing it for me, because when I go look, I won't have a clue. So I agree, definitely. It will be worth me going away and having a play myself. [...] As I say, I'll try to figure it out. I'll have a play and just give you my feedback and see if I can figure something out. It looks [...] it's coming on. It's definitely getting there. It's exciting, it'll be great. Those graphs that you showed me, very simple stuff, will make some people in the office interested and they'll be happy with what we're doing. I suppose that's the main objective, isn't it? Making people happy?*

New skills learned by the participant and additional design requirements gathered for the Data:In Place platform marked the conclusion of the second data consultation activity. The participant was happy with the initial results of the investigation; however, discussions about the actual use of them did not get further than 'making people in the office happy', indicating the fact that steps were still needed to be taken for the potential use of the data in action.

Taking Action

The final part of the data consultation with Charity A (6.3) concentrated on discussing how the output of data could potentially be used for action. Throughout the process of data consultations, participants identified multiple utilities for using data, such as self knowledge, evidence, measuring and recording behavioural change, getting funds to continue delivering interventions, and a tool for quantifying change.

> *We would want to know that personally. So, that stuff we're doing with [anon researcher] is great, but it's' backed up what we knew was happening. If he came back and said, 'Look, the kids step counts [physical activity] is going down,' we wouldn't publish that fact, but we would do something about it. So, we would have to be changing intervention.* (SM)

> *We've just got to try to do something that evidences what we have done has made a difference to mental wellbeing. For example, we could do a match day campaign. As a result of that, 1,000 people could ring Mind for some support. How do we evidence that's what we did that made people ring Mind and not that just suddenly loads of people decided to ring Mind? We're in the very early stages of looking at all of this. We're going to get another company to help us figure out how we record the behaviour change, basically.* (PO)

> *One of the things I'm working on at the moment is going to be an awareness campaign, city-wide, through match days and our social media, amongst other things, trying to improve health and wellbeing. We want to be able to prove that what we've done has made a difference.* (PO)

> *That's where our health and wellbeing department is at the moment. We've got that aspect; we've also got the bit that I'm doing in terms of the targeted mental health intervention. If we're trying to relate that to data and place, for me, I know that I'm going to have a 12-week programme in which they'll be sessions which try to improve the mental and physical wellbeing of the guys that take part. Looking at the long-term behaviour change and the long-term measures of those that I take at the beginning, over 12 months would be great.* (OO)

Those aspirations for data were guiding the participants in the hypothesis setting and investigation parts of the consultation, aiming to come up with outputs that serve those needs. Participants were focusing on the two main aims for the data output: (1) official documents and (2) presentations.

> *If we were reporting... say, if we had funding, and we were reporting back to the stakeholders or the people that funded us, then a graph would be a good visual tool to use. In terms of writing a funding bid, it's the facts that we would need, percentages or official figures.* (SM)

> *We want the facts and figures for [...] yes, official funding documents or reports to stakeholders. Then we're going to settle on easily-digestible... What are they called?. Infographics using the [Anon Sports] club as a metaphor, I suppose.* (OO)

When presenting the data to people outside the organisation, participants felt that the data outputs needed to be converted into something that represented their identity, which also came up in the first part of the data consultation. These custom infographics could potentially be used to reach different audiences through a multiplicity of channels.

> *I think for your average-joe consumer, it's plain, simple facts. As an example, we've got an event on the 22nd of March when we're inviting a variety of schools that we've worked with and schools that we've never worked with [...]. This is going to be a room full of teachers who are in charge of budgets. We want to sell, to them, our product. If, at some point, we can say, 'Up on the screen now, you can see that these are the schools that we haven't been in, their attendance, behaviour, and homework. These are the schools that we have been in, their attendance, behaviour, and homework. As you can see, there is a significant change between those two.' When we go into a school, we now know that these things improve. That would be a visual thing. It wouldn't be, 'Let me read from a paragraph and give you loads of statistics and loads of numbers.' People wouldn't get that.* (OO)

> *They [organisation marketing officers] might go to that school, knock on the schools door if they've never been in before, and now might take with them this infographic. They might go, 'Look, this is a simple way of us showing the impact that we've got, the impact that we've done in the region. The schools that we've worked in with [Anon Programme] have beaten the away team, the team that hasn't had [Anon Programme].' It would probably be the same infographic as what was used in the presentation held at the stadium. That would be taken and put onto some sort of booklet or brochure or easy-digestible material that they can take to schools. That's probably our two ways in which we sell, I suppose.* (OO)

> *Infographics are great because they're also shareable. If you put that on social media, people are more likely to go, 'Oh, I'll share this.' Then, it just getsâ€¦ It's more awareness of the foundation as a whole, isn't it, rather than it just being just a graph.* (PM)

However, making these customed infographics was something that could not have been achieved by the participants themselves, even when using the Data:In Place platform; in this sense, they would have needed support of a data visualisation professional. Throughout the consolidation process, that person was the author of this thesis, who assumed the role of the data professional. Going forward, there was a need to find new ways of providing that expertise to people that does not involve support from the researcher.

> *I think what you're saying is you know the answers, you know what the data are, you know all those different things and know the questions with those answers, and it's creating those pathways which I get that [...].* (SM)

> *There needs to be somebody who's not you, who's got your skill set but is able to do that. I suppose you probably know more than we would whether or not there are people out there that would be incentivised enough to answer the questions and challenges.* (OO)

Participants here were starting to think about how these connections between data professionals and issue owners could be created and mediated through digital systems.

> *It would almost be like a message board that challenges one to start with. [...] I guess a lot of the stuff we always ask is often the same, so it will just be like first run through and asking for data and it's usually general.* (SM)

This was significant because it would open up new opportunities for potentially expanding the skills network of Charity A and reaching out to external parties. Although this was promising, the participants saw multiple barriers due to the temporalities of these 'data challenges' and ways to incentivise individuals to take them up.

> *I mean, we can post as many challenges as we want in terms of we want to know something in a certain area. A lot of the time, it's last minute as well. It'll be like,*

> *right, we need to know who's doing what in here within the next two weeks. So say, for example, we post that on this system here who's going to pick that up. Why would they pick it up? Is it because they want to be socially doing more things for the society and things? You'd probably know more than me because you're sat in that position and representing that individual, whereas I personally, when I go home from work, I don't want to go back to work. So whether or not that guy picks it up on the other end and whether or not he can turn that around within two weeks will probably be the task.* (OO)

Additional barriers that surfaced included unwanted competition, safe ways of sharing data and establishing trust and credibility between issue owners and challenge takers.

> *I suppose there might be a bit in there in terms of, say, you've got a chat board or a chat room or whatever it is on here, everybody can see that [the organisation] has just asked, does anybody know what the crime rate is like in the west end of Newcastle? Other people might go, 'Why are they asking that?' Then, that could suddenly increase competition for certain bids as well.* (OO)

> *I suppose it's what you were saying before – how do we then trust that person, how do we know who that person is in a way. So how is that person verified through the system or through a different web because it could theoretically be, it could be then walking down the hill.* (SM)

> *That could be a 12-year-old kid in his room, and he's just got computer (OO). Get some data and just likes chucking it all over the internet somewhere, which hopefully it wouldn't be, but you never know really, do you?* (SM)

Indeed, it was important to establish trust between both parties to work together exploring and using data. It was not only an issue of asserting confidence in challenge creators but also of making sure that questions and challenges posted were coming from valid sources.

> *[...] if I'm asking questions on here, 'Hey, does anybody know how many overweight and obese men there are in the north east?' then that might make somebody go, 'Well, hang on a minute, why is he asking that?' If it was just at random, I think if nobody's got their name attached to it, people would be more willing to... [pick it up]* (OO)

The participant here went on to discuss the ways challenge takers could be verified on the web and trusted by challenge creators:

> *Bio almost, with even links to other stuff that you've done in the past.* (SM)

> *I'm going to trust our Eddie, who's got x, y and z or hundreds of publications on data science more than I am Jimmy, who's 12 years old.* (OO)

> *And our Ed works at Newcastle University part of Open Lab at Newcastle University as opposed to Jimmy, who's 12.* (SM)

Data science professionals spend years honing their craft, undertaking a range of projects to obtain the skills needed to work with data and produce valuable outputs from it. Participants saw an opportunity here to potentially give learners a chance to take up projects to improve their skills and solve real-life problems.

> *I suppose your future pool of people could be undergraduates or post-graduates who you want to build a portfolio as a data scientist.* (OO)

> *[A] real-life programme and challenge for a company while studying, that's probably one of the issue for students.* (SM)

One of the ideas participants proposed was also to build an ecosystem around data challenges that would enable a community of data scientists to grow and sustain the service.

> *I suppose if you want starting off small with your smaller community of data scientists in order for them to want to do this, incentivise them with a potential stake in the company or whatever it is or in your page or a stake in the interest of it as opposed to suddenly become they're your panel members and your decision makers of who is awarded certain statuses as future data scientists.* (OO)

The data consultations guided people through the complete process of data use and provided an in-depth first look into the cognitive processes that people go through in order to find, access, make sense of and use data in action. The findings presented here enabled further understanding of the model of CDI in terms of how people posit questions for data and the ways data can be accessed by people, transformed into a useful form and leveraged for actionable results. The data consultations provided new design considerations for both the methods of inquiry and the digital tools that would aid and mediate them, both of which are crucially important for the *social facilitation* (Chapter 5) of the effective use of data by communities. Lessons learned from the data consultations were applied in the second engagement activity with the local community.

6.5.2. Community Action Day

Building on the findings of data consultations, a workshop around making the use of data in community action effective was undertaken. This section presents the findings from that community workshop, where a set of methods and digital tools were trialled in order to explore their usefulness in guiding people through the process of using data in community action. The reporting on this phase follows the process described in the Study Design (Section 6.3) and illustrated in Figure 6.5. Because of the nature of the workshop setting and having a lot of cross talk in the transcriptions, the community is represented as a united voice and quotes are presented without pseudo identities. Distinctions are made, however, between the two groups formed in the workshop.

Issue Mapping

Issue mapping through the use of CC provided people an open space to bring any issue to the table. Participants were asked to use prompt cards (Figure 6.6) to provide more information about

issues flagged in previous sessions (Chapter 4.6) or to add completely new issues to the map. Although people had the freedom to do so, throughout the longitudinal engagement with the local community, a particular set of issues kept coming up and being identified by the community, which were linked to things like a lack of facilities for kids and green spaces that people in the neighbourhood could enjoy. Participants turned back to discussing these issues, some of which had also been marked in a previous Community Action Day where CC was used. The utility of tools such as Data:In Place or Ambit (the technology behind CC) was not only to document these processes for further use, but also to provide anchoring of these issues by situating them.

Issues that people were identifying were mainly observations about different places in the community that they wanted something done about. As mentioned before, many of the issues were linked to underused fields or green areas that had been neglected and were in need of maintenance:

> *It's on your left side and down the road, isn't it? Where Ronsons was, and then they got all the derelict ground.*

> *You got a lot of area out there on [Anon] Road, haven't you, that's just like lying derelict.*

Participants felt that these spaces could be utilised better, for example, by providing activities for young people who did not have any facilities:

> *Nothing, nothing. Absolutely nothing. Sometimes, it does get used for kids drinking on the field, hiding in the bushes, [and] so forth. So, I think a lot more could be done with that area there. They could have a lot more for the kids to play on there.*

Although participants were noticing these issues in the community, they felt that others, e.g. the local government, needed to do something about it. There were also ideas of establishing new housing and businesses or allotments on those 'no man's lands'; however, participants shifted the actions back to the local council:

> *I mean, the council could put into plan a cheaper rent for people to start their own businesses, or anything like that.*

> *Isn't that because the council have a long list of people who would like an allotment and the council are so slow at informing the people when the allotments become available.*

Using the prompt cards and taking turns flagging and discussing issues, both groups ended up with a set of marked issues on the map. The activity was set up in a way that people could express support or agreement with others through the 'second that' prompt cards, helping to filter out issues that were most important to people and preparing for the next activity of prioritisation.

Issue Focus

To move onto exploring an issue, participants had to prioritise flagged issues by discussing them amongst each other and listening back to recordings of other people's opinions before finally deciding on which issue was the most important to them. The mechanics of the CC already helped bring some of the issues forward that multiple people felt important for improving in the community.

As mentioned before, an example of 'rock climbing' (chosen by the thesis author) was provided to guide participants through the next steps of the process and the question formulation. However, having already gone around the map in the first activity and identifying issues provided the participants with a better, more anchored focus. One group took ownership of an issue regarding the lack of facilities for children – *'[We] want to focus on [Anon] Way, because there is nothing there for the kids'* – while another group focused on green areas and wanted to *'repurpose [Anon] View playing fields'*, where half of it was used by a local football club. Once the focus was agreed, each group used Data:In Place platform's mapping function to document their focus and make it visible to the other group. This provided each group with an issue to focus on and a location to anchor it to before moving on to exploring it further through questions.

Question Asking

At this point, participants had identified and flagged issues, listened to other people's opinions about them (both live and recorded), discussed them within the group, and chosen a focus for their investigation. They had then to use that focus and apply it within the QFT that was modified for (Section 6.3.1).

When people started asking questions about their issue, they were not constrained only to what they wanted to see in those places or *'what it is used for currently'*. Instead, people had a more exploratory attitude, promoting curiosity and discovery, coming up with questions like:

> *What would we like to see there?*
>
> *How much would it cost? [and] Where would the money come from?*
>
> *I wonder if there are any people who have already seen the places. What have you seen in the parks that you think we'd love to have?*
>
> *Would a sports day be – I don't want to say popular – community sports day be something you'd value?*
>
> *Also, how much would it cost to maintain? It might be like an ongoing thing.*

Asking questions about the issue also shifted the focus from 'what the council can do' to 'what the people themselves or the community can do':

> *[The] first one would be what can **we** do with a bit of grass?*
>
> *How would **we** go about changing it? The likes of who would **we** have to see?*
>
> *We can only try. If **we** get the funding on it.*

When people worked on their questions according to the QFT method (for example changing the openness of them), they also reflected on their own role as community leaders or spokespeople for the community, perhaps using the same questions to give voice to the wider community and get more opinions forward:

> *Actually, between them two and these two [questions], what I've changed is you're not giving people the option on what they would want there as in what- what was it then we changed, number four? We're talking about the green area behind a specific thing, and to me it would be a park, but on my first question, what would you like to see there? Now that gives everybody and anybody a chance to say what I would like – a bike park, skate park, this park, that park. My next question is would a park benefit your children?*
>
> *So, I probably shouldn't even use the word 'park', because that limits it a bit, doesn't it?*
>
> *How do we go about changing from what it is now to what we want it to be? Or what the community wants it to be? It's not what we want it to be, is it?*

Additionally, questions sparked discussion around the way participants could figure out the needs of the wider community, make them visible to everyone and get the community involved in delivering those changes for the neighbourhood. These points were the main considerations when prioritising questions for investigation and challenge creation for both groups:

> *So, our priority questions were: What are the options for the use of the space? Quite broad. What is lacking in terms of the existing local provision, facilities? Then, what is the space already used for?*
>
> *Our three questions were, because we would like to see more things on there for kids, so our three questions were: How do we get the community involved and maintain their involvement? Because if not everybody wants it, then it is not plausible really, is it? It's got to be everyone on board. Our second question was: How do we go about changing the space to make it what the community want to see? Our third question was: who would fund this?*

Applying QFT on the issue focus set the groups up to have concrete questions at the end of the process, which set the tone for the challenge creation activity.

Challenge Creation

Challenge creation aimed at breaking down each question using the prompts provided with the worksheets (Appendix B). Because both groups were looking at green areas, it was important for people to figure out who owns those lands first. People always assumed that land was owned by the council, but nobody knew that for certain. This was a clear challenge in terms of *data access* to obtain the information of land ownership.

> *I always thought, 'Who owns the deeds of that land?' You know, because obviously I think they should be...[owned by council]. Then, of course, the football team comes into it but owns that last 50 percent of the field or whatever is left of it. Is it [the] council or is it private land?*
>
> *So, local knowledge, council. Presumably, do the football club own their little patch?*
>
> *Yes, it's finding out the ownership, yes? So, what I was saying about the idea of finding out what the options are for how to use it.*

Another challenge identified by both groups was to figure out the current uses of those two spaces: *'I wonder if the idea of the, what's the space already used for? That's quite a key question, isn't it, really?'*. People had some ideas of the uses based on the observations of *'there is walking access, [...] there are a lot of dog owners that use it [the paths on fields]'* and *'there is a metal seat there, and there are metal goalposts, and a hoop [...] kids do use that [...], but sometimes it is full of glass and that'*. Additionally when asked if there was any data about it, participants said that there was nothing they were 'aware of'.

In terms of data challenges, this was a challenge for *data production*. This could have been done by observation or by compiling all the local knowledge of the workshop participants, but people thought this could be a good opportunity to collect data from the wider community – people's current uses of the fields and their ideas for the future. This was a challenge that resulted in suggestions for immediate actions that people themselves could take it rather proposing it to someone externally.

> *Right. To collect the data, or what people think, we would do a questionnaire. We could go door knocking in the local area. Yes [to figure out what people think], and in general if people would be interested and that.*
>
> *Actually, I think you would have to put it into a different thingy. Not what would you want to see there? Because you could have one person wanting that, that and that. If you said, 'Would you be interested in a park being put on?' [then] it would be a closed question, as in yes or no. Then your piece of paper would just be two. Or it could be three. You might have someone who says, 'Maybe.' (Laughter) Then you would only have two, and then it's pros and cons, isn't it? Yes and no.*
>
> *Do more people want a park than not?*
>
> *So, I think the collecting stuff ourselves thing, that's going to be what people would want to see it used for, isn't it, and what they're already doing. So, there is potential there for collecting data from existing usage and that, potentially.*

This would also extend to not only collecting the current uses of the fields but also getting people involved in thinking about what can be done with the spaces.

> *So, as in, what do they know that's elsewhere that they'd like to have on their doorstep [...]*

> *I know this sounds like Eden, you know that big project down in Cornwall? A local communities thing called the Eden Communities that helps local communities to repurpose green spaces and that sort of thing or to organise themselves.*

This also sparked conversations around compiling data of things people do in their free time and figuring what the current available pastime facilities are in the area or neighbouring areas, creating multiple challenges for data in terms of *data access* and *data production*.

> *We could do with all this, knowing what the council provisions [...and] a little bit of just local knowledge that we could use there.*
>
> *Yes, so actually even knowing that exists there is going to be useful, for just knowing what the options might be for knowing about the other facilities [...]*
>
> *There was one over the, I think it was Station Road in Newcastle. If you come off the [crossroad], you go down over the bank, there's actually a massive park which has everything, obviously swings for kids and I think one looks like a full-sized basketball – a tennis court, whatever, but they have full-sized court there, which is – but I don't know if it's still there or not, but it looks like there's a park for kids, swings [...]*
>
> *I can't remember what information is in the census because, you know, you can get the anonymised data from the census from 2011, whether that would have anything about what the pastimes... I don't know if that's in there. You don't happen to know?*

Data challenges not only focused on the first part of getting the data (i.e. data access and production) but also got the participants thinking about challenges that would follow once the data was obtained or collected. These challenges were linked to *data visualisations* that prompted people to think about how they could turn that data into something useful. Participants thought that it would be best to present that data back to the community to gather support and get feedback on ideas.

> *I think you could put it into a leaflet. We could put one through the doors or if you go on the computer, then that's all that data on there.*
>
> *With the computer or just talk, have a chat sort of thing.*

This would potentially include data collected from the community using questionnaires, which are process as *'[...] information in simple numbers, or graphs'*, *'then you would probably put them into categories. Would people want it as this, this or this'* and merged with data from official statistics *'to give a simple answer everyone can understand'*.

> *The questions would... What would you like to see on the field? What would you like to see the field used for, blah, blah, blah? When the questions are all done, stacked up.*
>
> *So, from any data that we would collect from door knocking or talking to people, or whatever, that could be presented in a... Fifty people said, wouldn't it be fab if*

we had a kids' playpark. A hundred people wanted a...' So, there could be some graphical representation of [statistics].

Could we also put in something there about the census pastime stuff as well? So, you know, we've got 300 people from the [X place], who love to cycle.

However, not only were people interested in representing people numbers and official statistics, they also wanted to seek feedback from people and start contesting these numbers:

If we stuck in a bike park, it's a bit of – they don't quite [add up]. So, just [If you point your finger with that sort of thing there], then there is a load of cycle tracks, but if that's the place you mean?

Participants were interested in presenting data about the spaces – what exists, current uses and potential ideas – in a way that was more interactive and representing the community identity.

Yes, I guess the visual representation of the space, we're just using a map of the space and saying, 'That's already the football thing. That's already this. This is the space we're talking about' which represents the landownership bit.

So, if you have an idea of what the council provision of their facilities have already got somewhere else, there's potential, isn't there, to stick some pictures in of the fact that they've got a skate park here. The likelihood is we're not going to put something there, but just to say, 'That's there. There's a bowling green there.' There's whatever.

I guess you can, sort of, mark bits out, so people can go, 'Could this be this? Could it be that? Could it be that?' I wonder as well, if you were going to do a leaflet for the [Blue Sky Way's] bit, then there's potential to just say, 'Oh, this place in Sheffield has this and green space like this place. There has that...'

Perhaps the biggest challenge people identified was how to fund the development of these spaces and where they would start looking for such funds.

Well, it's got to be the big one, isn't it? Where would the funding come from? Where would the money come from?

But we could put it to the council, but then you could also put it out to... Oh, I don't know. I would probably ask where could I go to find funding for it?

Yes, so I guess presumably different funding bodies will have very specific things they want to do and stuff.

Participants were not only thinking about how they could convince the local government to fund these changes, but perhaps there were alternative ways of funding it, looking at how they could collaborate with other organisations in the community and even outside of it

Maybe get some businesses on board. There are local businesses around here. There are B&Qs. There is GTF. Actually, there are also another two community centres. There is the [Anon Organisation] and the thingy, so maybe they could help.

Oh, yes. Presumably, there might be someone that can say, 'Yes, okay, we've got a specific pot of money, and we would build a kids' playpark,' or, 'We've got a specific pot of money that'll do this or that.' That might influence what you end up doing.

For the last activity, participants were brainstorming actions they could take with the data. Their initial idea was to compile all the data and then seek feedback from the community to build support:

Yes, I mean, so in terms of making the information actionable, are we going to say to people from this leaflet, 'Give us a tick to say, 'Yes, we'll want to do that,' or a cross to say, 'No, I don't want to do that.'

Yes, or do we say, 'Just come along to the meeting and we'll talk about it at a meeting,' and present a much more concrete ideas about things.

Getting all the community together to brainstorm something like that. I know you can't keep everybody happy.

I guess, in this exercise, we're assuming that if we're doing this, there's a certain level of buy-in, we're happy, as people, doing this to galvanise a bit of... You know, if we're talking about doing something with here, we're happy to lead a bit on saying, 'Come and get behind this, we're going to do something with it.'

and then taking that information to the local government:

You would have to get a meeting with the council [...]. We would have to contact the council, wouldn't we? You would take your findings from that to your council.

If everybody agrees, then you go to [the] council and just say, 'This is what we'd like to happen.'

[saying to the council] 'This is what the community wants. We have done that door-to-door. We have done whatever is needed'

We sort of show them the data, do we, to say that this has got community backing?

Yes. Ask the council if there is anything they can do, even if it's half.

Participants were not really convinced that the council could do anything and they would probably respond with '[...] It's gone past our budget for this year, sorry. (Laughter) That's what they'll say.' However people started to shift into the mindset that perhaps the council did not have to do it all and there were ways to do it together with the community.

Yes, but maybe it is not all about them doing everything. It's making them aware that the community are willing to help as well.

So, I guess, almost then using that as the start of trying to gather people together, affirmation of it, so people say, 'Yes, we would do that,' or, 'Yes, I want to do that, and I'm also willing to put the time and effort into making it happen.'

Are you [member of community] painting? Are you digging holes? Whatever it may take.

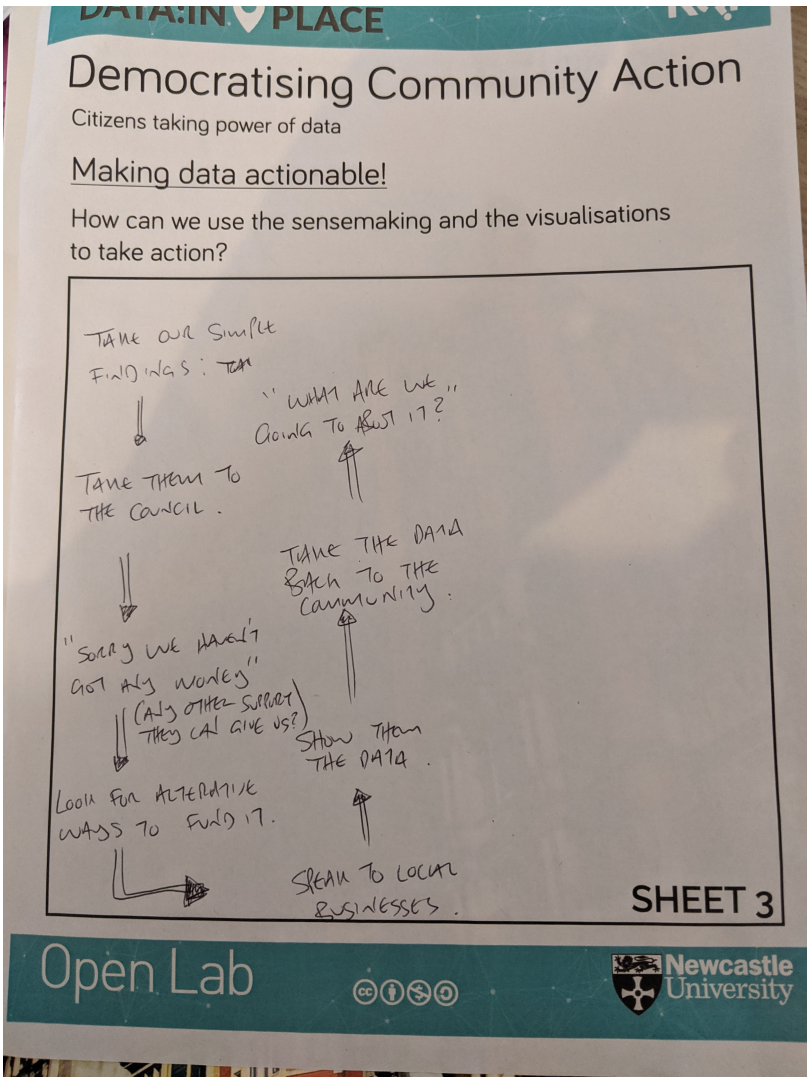


Figure 6.10 Screenshot of proposed steps for community action

Here, the participants used the challenge creation activity to think about not only what others could do to make things happen for them, but also to look at what actions they could take themselves, finding skills within the community and also using the Data:In Place platform to post data challenges to get access to professional expertise externally. Figure 6.10 illustrates the proposed action plan of one group. Furthermore, statements such as *'it would have to be a community project'* represented the realisation that it has to be a collective effort of the community, where data could potentially be used to foster dialogue between different stakeholders and bring the community together to take action.

6.6. Discussion

The dual approach for CDI facilitated the process of discovering ways that data could be made actionable. A process of inquiry and asking the right questions enabled people to form a better understanding of what data they might need and plan out steps for actionable results. The findings contest the ways that data is currently projected *to* and *about* local communities, in addition to opening up new possibilities for better knowledge making and more meaningful *CDIs*

to take place. The following sections discuss the findings of the case study, reflecting back on the questions and aims of the case study (Section 6.3) as well as the model of CDI in practice.

6.6.1. The Power of the Right Questions

Many scholars in HCI have made a strong case for the importance of curiosity, discovery and inquiry in the participatory design process of exploring issues (Balaam et al., 2015; Balestrini et al., 2014; Dantec and DiSalvo, 2013; DiSalvo et al., 2008; Marres, 2007; Moor and Cindio, 2007). However, there is an inherent difference between observation and investigation of matters of concerns. The case studies in this thesis have shown that flagging issues is useful for providing new observations about things people are concerned about. These are useful in many regards, especially when mediated through digital tools, providing a way for people to express their concerns, resulting in a new and interesting dataset. This type of crowdsourcing of data is common in large-scale participation projects (Dantec, 2014; Liu, 2014; Thompson, 2016). Archived collections of citizens' opinions serve as interesting datasets for researchers working at the intersection of qualitative and quantitative data. There is a risk, however, that the mode of citizen participation is often abstracted to pins on the map (see literature review Chapter 1 for different modes of participation) and not engaging with citizens to become the active users of data.

Unfortunately, in a lot of cases (also illustrated through the case studies in this thesis), these observations do not necessarily always follow any actions. The reasons for this may be linked to factors such as a lack of understanding of the issue, complexities surrounding decision-making processes that bring upon or instrument change, and a lack of community support or perhaps a lack of evidence. Although data might be 'out there' or there are ways to generate it (for example, using the toolkit described in Chapter 2), people often feel powerless and suffer from an inability to take action. Early in the process of defining issues, people were pointing out places (using the Data:In Place platform and functionalities of Ambit) in the community that they thought someone needed to do something about, for example, places where youth engage in anti-social behaviour, rundown housing estates or neglected derelict grounds. The pressure was put on the local council to 'sort these places out', and people assumed that they cannot do anything about these issues themselves. However, in many cases, the community problem-solving activities needed issues to be better articulated and dissected by community members.

Introducing QFT into the process of inquiry by choosing to focus on something particular and starting to ask questions about issues opened up alternative views and possibilities around taking action, whether it was connected to bringing the issue to light in the eyes of the wider community or even 'rolling up the sleeves' to put in the work themselves. This also extended to a better understanding of what data (and other resources), capacities, and infrastructures are needed and what actions need to follow. Furthermore, having concrete questions helped people to understand the roles that needed to be filled and the constraints they were working against in a given context, all promoting actions that extended from getting and understanding the data to what was equally important: How that data became interpreted and presented to others, thus shifting the end goal

from *getting access* to the data to *making data actionable*. This also challenges the ways citizen participation through data should be perceived, recognising broader interactions and the extent of how data could be used by communities.

6.6.2. Holistic Approach to CI

The practical approach for the CDI was put forward though the dual approach method described in Section 6.2 in order to better understand how CDI works in practice. Hakken (2003) points out that the main aim of informatics, as an engineering discipline, is to design AITs that work. This often lends itself to building abstractions that are distanced from the social and/or contextual dimensions important to knowledge-making processes. In the context of CDI (Chapter 5), these are the key elements of the model that are bounded by the factors that influence these elements and also become the domains of change. Using an approach that facilitates the functionalities of digital tools and data enables not only an understanding of these dimension but also helps accommodate the use of AIT to already existing social contexts. In the case of the participants in the study, these were linked to the historical relationships between residents and decision-makers or current practices of the local community organisations. These contexts became part of the process of inquiry, helping bring issues to public attention and making people reflect on their own practices by thinking of the ways things have been done before and how they could be done in the future. Digital tools that accompanied these processes acted both as enablers and historians of sensemaking by creating digital trails of people's interactions. Studies in HCI and participatory design have demonstrated the benefits of capturing different voices and situated understandings, enabling a multiplicity of narratives and social interactions to come to the surface (Bartindale et al., 2019; Manuel et al., 2017; Rainey et al., 2019). In many ways, the processes engaged in data consultations and Community Action Day served as means to bring attention to relevant contexts so as to interpret and make use of abstract open datasets available through Data:In Place. Although Data:In Place and its functionalities were generically built to enable better access to and exploration of open datasets, their appropriation was highly related to the communities using them and their identities. This related to the datasets needed and was relevant to not only the issues people were faced with but also the way data had to be represented to include the visual identity as social context. For Charity A, these were things that associated them with the local sports team, while for the neighbourhood planning group, these were things that represented community identity, making data relatable to the local residents. However, this context needed to be brought out by people and intertwined with the data and its formal representations.

As information and knowledge are situated and constrained by interpretations (Suchman, 1985, 2007), this makes it difficult to use IS to convey that needed context. According to Devlin (2001), however, it is impossible to embed context into information – it should be extracted from information, and the most we can do is to provide an indication of what the context might be. This formal abstraction has been criticised by people in SI because it is contradictory to the social understanding of information and knowledge (Hakken, 2003). However, both sides agree that in order to recover knowledge from representations of information, a social interaction

is needed. While face-to-face interactions are the preferred way of knowledge networking, alternative approaches such as embedding the context in representations of AITs and supporting knowledge-making processes around them should be considered.

With regard to social readings, Hakken (2003) provides an understanding of why such approaches are more likely to have the desired outcomes. In this sense, he states that *'different understandings of functionality in software would provide a better key to developing a more satisfactory approach to the problem of mediating knowledge networking via AIT'*, and the *'goal should be to find AIT-mediated forms of knowledge representation that functionally equal the sociality of existing knowledge networking'* (Hakken, 2003, p. 156), thus making the main goal of AIT system to implement ways that enable to encode as much context into formal representation of data as possible in order to support knowledge creation and action. He calls these *applied social informatics of knowledge (ASIK)* systems, which combine ideas from User Participation in Systems Development or Participatory Design (PD) (Dahlbom and Mathiassen, 1993) and CSCW (Winograd et al., 1986), acknowledging that most of the work take place in groups.

Drawing on Hakken (2003) and others in PD and CSCW, an instance (or perhaps a manifestation of the author's experiences and PD process) ASIK was also designed and built in this particular case study following a longitudinal engagement with stakeholders from the local community. Moving away from purely information engineering approaches to social and community informatics, and through the use of CDI, enabled a more holistic approach to how data is accessed, interpreted and made useful by communities to be taken. The findings suggest that the dual approach for CDI, using Data:In Place with QFT and Challenge Creation, helped facilitate the process of contextualising data and provided conditions for actionable knowledge to be created. The process of inquiry often revealed that the required data was not as complicated as initially perceived, and through contextualising, it could be presented by means of simple visualisations that people could do themselves or by using the Data:In Place platform. Additionally, the future directions of the platform may consider ways that knowledge networking can be expanded by brokering the communications between different stakeholders, both professional and non-professional.

6.6.3. CDI and Collaboration

The findings from this case study (and from Case Studies II and II) highlight that when it comes to data science skills, there is still a need for expert professionals to provide guidance or lend their skills. Even if some processes can be mediated by intelligent systems – such as was illustrated in the case study when local Charity A was using the Data:In Place platform to access school performance data and correlate it with their engagement data to produce visualisations Ð data often still has to be found, processed, mediated and/or put into these systems. Without a strong existence of a *Semantic Web* to represent all the data on the web, a lot of the pre-work needs to be done by expert professionals. In the context of the research conducted in this thesis, the author took this role and provided training, tools and professional support to effectively use data by communities. Going forward, solutions such as Data:In Place's *Data Challenges* become

important for mediating wider collaborations between communities and expert professionals whose help is needed. Reporting issues and challenges through a digital interface enables adding important meta- and paradata to adequately respond to them. However, these connections can be better achieved when the community knows what its agenda is, what the influential factors are, what capacities and resource are needed and what it wants to change. This can be done through conducting design and engagement work together with communities. Although some of the skills and resources could be outsourced, such as data science skills, it is also necessary for the community to take ownership of the issues and tap into local resources and skills to develop a strong community network. These realisations were also echoed by the participants, who knew that a handful of community advocates was not enough to take action; instead, it needed to have the backing of the whole community to instrument change. Furthermore, the building of the capacities of the community is a step closer to the successful transfer of ownership of these technologies and the solutions built in this research to achieve their sustainability.

The case studies in this thesis have shown that a wealth of networks (Benkler, 2007) and capacities play a huge role in making use of data to support civic advocacy and action. Community groups and individuals who possess larger social capital can make use of links outside their own communities to get access to the skills and knowledge needed to make use of data. On the other hand, people who are marginalised or living in more deprived areas are less likely to 'tap into' resources and infrastructures that enable them to use data in any meaningful way. However, through *social facilitation* and *service access/provisioning*, which are the key components of the CDI model, it is possible to work on the community problem-solving agenda and nurture these connection, enabling the whole population to achieve control and empowerment. Working together with different communities through methods that include direct interventions (e.g. training, facilitation and skills transfer) opens up new opportunities for appropriating digital tools and creating a social ecosystem capable of having an positive impact on the community. Going forward, it is essential for CI systems, and also big data systems, to integrate mechanisms for community interaction, feedback and collaboration built upon the contextualisation of abstract datasets.

6.7. Summary

This chapter presented a case study that applied the CDI model in practice. This was done through a dual approach method, consisting of system approaches and capacity building, both including general and targeted interventions to help implement strategies. The case study consisted of two phases: (1) Data Consultations, where the author worked with stakeholders from a local charity to understand the cogitative and social processes surrounding the effective use of data and (2) Community Action, which related to applying social facilitation methods with digital tools in a community workshop to help people better formulate their agendas for using data in civic advocacy and action. Using this holistic approach to CI enabled further identification of the practical forms that the key elements of CDI could take, giving insight to applying better strategies for the effective use of data by communities. The next chapter reiterates the findings and discusses the research conducted in this thesis, reflecting on its aims and objectives, before concluding with outlining future approaches to and potential issues of the CDI model.

Chapter 7. Effective Use of Data: CDI and the Effective Use of Data by Communities

For these issues to be successfully addressed through the use of ICTs, attention will need to be paid not simply to 'access' but also to an entire range of supports for 'effective use'.
— Michael Gurstein

The primary aim of this research has been to explore tools and methods for the effective use of data by citizens in civic advocacy and action. Taking a PADRE approach, investigations were conducted through the design, development, deployment and analysis of two CI platforms (SMS in Chapter 2 and Data:In Place in Chapter 4) for the effective use of data, which are being further sustained though research facilities and continue to be used by communities. Reflecting on the iterative design processes and the use of these systems by communities enabled the author to produce a new model for shared data interactions as a CDI that could help designers discuss and design for the effective use of data by communities. This chapter begins by providing a brief overview of the research and revisiting the initial objectives presented in the Introduction to see how the study has responded to these objectives and how it contributes to the wider research field. Furthermore, the sustainability of the solutions resulting from the research and how this work can inform further research in HCI, CI and wider fields are discussed.

7.1. Overview

Advances in ICT and an increase of smart city technologies being deployed have made it possible to use data for the measurement, control, optimisation and management of cities, as well as mediating and supporting citizen participation, which can take different forms depending on the level of engagement from citizens. At the far end of the spectrum, there is passive participation through data that supports the algorithmic smart city agenda (i.e. citizens as a source of data). However, at the other end of the spectrum, there are smart citizens – citizens as active producers, consumers and users of data. These latter modes are supported and advanced by citizen sensing technologies, open movements (e.g. open knowledge, open source and open data) and new ways of engaging in and creating community networks. These different modes of participation, highlighted in Chapter 1, can help citizens get more engaged in, and in some cases use and produce, data; however, more work is needed for that data to become useful for the community. Furthermore, the data in question is often big data at the scale of cities, which is inherently complex, i.e. large in scale and/or velocity. To work with such data requires specialised tools, together with the skills and knowledge of expert data professionals (i.e. data scientists) who ultimately make that data useful for people. However, skills relating to information processing and cognitive processes take years of practice and training, which puts them outside the grasp of most individuals and communities. The challenge is then to understand how these skills can be transferred either to ordinary citizens or into a system that automates processes for them. The field of CI has emerged to respond to these challenges; however, it is a relatively new field that still has gaps and lacks empirical evidence relating to how IS could be utilised by communities in practice. This thesis explored such challenges through the design and development of CI tools (Chapter 2 and Chapter 4) and proposed a new approach for the effective use of data that aimed to utilise technologies in all aspects of CDI (Chapter 5). Furthermore, the research went a step further and provided an example of the practical use of the CDI model (Chapter 6) in a community context. In this sense, the research not only designed and built the types of tools that non-professionals would need to interact with data, it also extended beyond this and looked at methods of *social facilitation* needed for the effective use of data by communities.

The Limitations of CDI for CI Research

The research in this thesis was conducted through 'in the wild' (Rogers, 2011; Taylor et al., 2013) deployments with particular communities, dealing with explicit issues in specific locations at a certain point in time – which makes the repeatability (or reproducibility) of the research virtually impossible. Furthermore, the research was highly dependent on the efforts of the author, who positioned himself as part of the communities, providing expert skills and knowledge, to become one of their supporting capacities. In turn, this enabled a better understanding of the context and identification of the key elements that contribute to the democratisation of data science practices and the effective use of data by communities. However, by learning from the processes

documented by this research and applying the CDI model on their own practices, researchers will be able to produce valuable outcomes in similar contexts.

7.2. Approaches to CI Research

This section looks back at the aims and objectives of the research conducted in this thesis, reflects on the contributions of the study and points out the contribution of the thesis within the wider research field.

7.2.1. Open/Available, Accessible, Comprehensible, Usable and Useful Data

> (*Conceptual objective*): Through a critical lens of engineering (i.e. Information Theory), epistemology and social studies, develop a conceptual model and understanding of 'what it means to make data usable for the purposes of civic advocacy and action'.

The majority of critiques around the use of data by citizens point to the lack of resources, access, knowledge and skills to interact with expert systems that are used to produce or make data available; in this sense, the statement that data, in principle, is 'open' for everyone to use is not true in many cases. For example, it is not true in relation to active citizen data production and accessing open datasets such as OGD. For the vision of a *citizen as an active data user* put forward in Chapter 1, these barriers needed to be overcome for this positioning to become a reality. Hence, the initial investigations of the research looked into ways that data production could be democratised through establishing a commissioning toolkit (Case Study I in Chapter 2) and how available (i.e. open) data can become accessible, comprehensible and usable data for citizens when it is contextualised using a CI system (Case Study II in Chapter 4). One could argue that, in terms of the dual agenda of the CI (Stillman and Linger, 2009), these problems related more to the IS (or engineering) problem-solving agendas, which deal with the technical design of information technologies and systems. However, the user-centred integrative design process undertaken in the design of the technical solution was dictated by community problem-solving activities (the second agenda of CI). These efforts resulted in the development of SMS (Chapter 2), the first sensor commissioning toolkit for communities that enables people to use and commission data from scientific-grade environmental monitors from their local neighbourhoods, and Data:In Place (Chapter 4), an open-source platform that facilitates accessing and making sense of open data from expert systems by communities. These systems reflected the different layers of challenges involved in democratising practices around data, i.e. data needs to be (i) accessible, (ii) comprehensible, (iii) usable and (iv) useful, which the research in this thesis systematically worked through.

Although the systems built addressed the dual agenda of the CI, working on both the technical and social design of IS, the implementations on their own did not make the data useful for communities. Reflecting on the first two case studies (Chapter 5), it became evident that *usefulness* is not actually a property of data. Essentially, all data can be useful, but usefulness is

actually a property of the person or the interaction. People need a purpose they can use the data for, and then it becomes useful to them. This contradicts the notion of Ackoff (1999) of data's ability to produce information and knowledge and wisdom (i.e. knowledge that can be applied in action). Data, often in the form that it has been collected and presented, is not self-evident and does not become useful right away. It needs multiple processes, and often social process, to help give it meaning. Furthermore, when data is taken out of a particular context in relation to both its place and community, it often looses its meaning (Gitelman, 2013; Hakken, 2003; Taylor et al., 2014). These are, however, things that cannot be implemented into algorithms but relate much more to human processes of facilitation and trying to work with people to help them understand what their needs and challenges are. Additionally, the findings from the first two case studies revealed that the key to being able to make use of data is social capital and often being able to access capacities and skills outside the community group. However, it was illustrated that digital tools can still be configured to be generic and robust but have processes attached that help facilitate or infrastructure the creation of community networks. These findings, while somewhat contradicting existing research in the design of community technologies (Dantec and DiSalvo, 2013; Ehn, 2008), have been proved to be valid by research in community commissioning (Garbett et al., 2016). Contacting research in these spaces and by moving from data access and use (including production) to start looking at what makes data useful for people in communities, the research took a shift towards a more collective and collaborative approach and started looking at the wider factors that influence the effective uses of data by communities, one of which was the shift from active citizens to communities, and also at the roles and supporting capacities around the uses of data for advocacy and action.

From Active Citizens to Communities

From the beginning, the aim of this research has been to understand the processes around active citizen participation in the smart city. The term *citizens* has been used throughout this thesis to refer to *'people who inhabit, work and play in cities'*. The official understanding of a citizen 'as a legally recognised subject or national of a state or commonwealth, either native or naturalised', however, does not apply to all the citizens referred to in this thesis. When embarking on this research around active citizen participation though data, the objective was to support active forms of using data (Chapter 1) by citizens to produce actionable knowledge. This was also supported by positing a CAF in Chapter 3, a novel model to help understand, discuss and plan for civic advocacy and action by citizens. However, the context of this research was not about individuals using data and IS to create value for their own purposes (although this can be a part of it, see Chapter 3); instead, it was more about community sharing, creating knowledge and courses of action. Furthermore, the knowledge that is needed for civic advocacy and for taking collective action is collective knowledge and distributed information of the community. Hence, throughout the research, there was a shift of focus from active citizens to communities. While still promoting the agenda of citizens as active users of data, the focus shifted to the ways communities can make data useful for activities of civic advocacy and action. The findings

related to civic advocacy and action from Case Studies I and II also showed the importance of social capital and building community capacities. For that reason, the CDI model was built around community and focused on the key elements that promoted the creation of community networks that benefit the achievement of the *'knowledge society'* (Hakken, 2003; Jennex, 2017). For Jennex (2017), the knowledge society is an approach that uses knowledge and intelligence to create processes and products that better serve citizens. This idea of creating processes and products for citizens to use aligns well with the research agenda of SI and also CI (Stillman and Linger, 2009). However, thinking about the CDI within CI, the knowledge society is more about the process of facilitation and support through the use of methods and digital technologies to help communities create and exchange knowledge for the purposes of civic action. This aligns more with the approach of Hakken (2003) to knowledge networking and the consideration of Gurstein (2012) that *'development, deployment, and use of ICTs for local benefit' (p. 36)* should be the central concern of CI. If the focus shifts from just designing and building better SI and CI tools to accessing, making sense of and using data to thinking about the interactions the community is having around the use of data, we can then start incorporating new ways of researching, designing and building tools that support the processes of CDI and the effective use of data by communities. Achieving the 'effective use' of data by communities (Gurstein, 2003) is a far more desirable outcome than a very well built tool for data access, comprehension or visualisation.

This type of thinking and designing was put into practice in Case Study III (Chapter 6), where the aim was not to build a perfect CI tool for data (i.e. a perfect version of the Data:In Place platform), but instead to help communities better understand their challenges and the resources and capacities needed to use data for activities of local benefit. The inclusion of *Hypothesis Setting* and the *Question Formulation Technique (QFT)* to the processes of inquiry and the *Challenge Creation* methods for taking action are some of these examples that where trialled in Case Study III. Whether the community's aim was to improve and optimise its working practices, as was the case for the local charity, or to get better service provisioning and facilities for the community, as was the case for the neighbourhood planning group, the indicative findings from Case Study III (Chapter 6) suggest that the ability to understand the challenges, recognise the capacities needed and be able to plan out the actions were far more empowering for people than having access to open datasets. Rather than giving communities products and services, and coming in and trying to solve their problems, the approach taken was more about helping communities increase their capacities and resilience to resist economic and social disparities.

Dissembling the Role of the Data Scientist

One of the important capacities around making data useful for local benefit is the ability to manipulate data, which does not only mean having access to data and the tools to work with it, but also possessing the skills to be able to use it: being able to ask the right questions and knowing what needs to be found and what to do with it. These are all the skills of a professional data analyst, more widely known now as a data scientist. Often, the type of data in these

contexts, such as in the smart city, is inherently complex (i.e. big data). Quite often, therefore, due to the complexities of data, there is a need for the help of a data scientist. Throughout this research, the author acted as a data scientist supporting communities with the skills and knowledge needed to work with data. However, the challenge was to figure out ways that the role of the professional could be reduced in order to truly democratise the processes. This was done, for example, by using intelligent CI systems such as Data:In Place (Chapter 4) to interface with expert data accessing systems (i.e. APIs) and, through the process of QFT, to come up with the right questions for data, or using Data Request on the Data:In Place platform to find data needed by the community. As a result, the role of the professional did not disappear from the picture but was reduced and accompanied by automation and also, to some extent, skills transfer to ordinary citizens. Furthermore, the combination of *Challenge Creation* methods and *Data Challenges* on the Data:In Place platform trialled in Case Study III (Chapter 6) opened up new opportunities for capacity building and the creation of community links between citizens and data professionals. This illustrated a new way of replacing the role of the researcher (who acted as a data professional for the community) with automated tools to help other data professionals interface with the community.

7.2.2. Digital Civics Approaches to CI

> (*Technical objective*): Design and build tools that enable and promote the use of data in civic participation and advocacy. This involves understanding the context, culture, values and practices of different community groups working with data in order to adequately respond to their needs and the issues they are tackling. This also includes establishing processes that will help build capacity within the community so as to sustain the support for community groups working with data.

The case studies conducted in this research resulted in the design and development of multiple digital tools (SMS in Chapter 2 and Data:In Place in Chapter 4) for data production, access and use by communities. Some of them were built as bespoke platforms and tools, such as: the online GIS issue flagging system[1]; an Android application[2] for recording GPS and personal reflections; an online proposal collaboration tool IdeaBoard[3]; a sensor data analysis tool[4] for the SMS toolkit; and the Data:In Place platform[5] for accessing and using open data. Others, however, were already existing platforms that were configured to serve different purposes for the SMS toolkit: automatic on-boarding and monitor loan forms[6] and an automatic monitor loan scheduler[7] configured on Google Apps and automatic sensor mapping using Carto[8] for mapping hand-held monitor data from community sensing. Furthermore, all these technologies were the

[1] https://sensemystreet.uk/map
[2] https://play.google.com/apps/testing/uk.sensemystreet
[3] https://sensemystreet.ideaboard.co.uk
[4] https://github.com/aarepuu/fidasfrogparser
[5] https://app.data-in.place
[6] https://goo.gl/forms/7A9eVtB4KwyuM0AA2
[7] https://goo.gl/14wrgM
[8] https://sensemystreet.carto.com

result of an iterative design process, where the solutions and features implemented were the direct result of community challenges and the issues they were tackling.

The initial exploration for establishing the SMS toolkit (Chapter 2) was conducted through ethnographic work, engaging with active local community groups, followed by a community workshop to further explore ways to promote the usage of data for raising awareness about the issues in local neighbourhoods and for civic action. The toolkit was then set up trough a user-centred integrative design process that evolved through attending and taking part in community meetings with local community groups advocating for built infrastructure changes in their local areas. All this resulted in the development of SMS, a sensor commissioning toolkit for communities that enables people to use and commission environmental data from scientific-grade monitors from their local streets and neighbourhoods. The research not only tried to provide the needed resources for the communities but also worked together with the communities to understand best practices for utilising resources and helping citizens effectively use the commissioned data (through the use of CAF), often designing user-friendly manuals to replace the ones accompanying the scientific-grade environmental monitors[9]. Constant feedback loops with the community provided valuable input for further improving the technologies, in addition to also enabling the author to understand the key elements of CDI and successful uses of the data in action by communities (Chapter 5).

Similarly, the Data:In Place platform, presented in Case Study II (Chapter 4), took data access and active use by communities under consideration by exploring the value of OGD to help inform local decision-making. Although the case study was structured by means of three design principles for accessible data systems based on the current state-of-the-art of open data systems for non-expert users and critiques from the literature, the platform itself was designed and developed through an 18-month iterative co-design process involving a neighbourhood-planning group and multiple charity organisations from the north east of England. This involved the author attending and taking part in neighbourhood planning meetings and having regular meetings with local charity organisations to understand the contexts the communities were operating in. The detailed iterative design process that was described in Chapter 4 showed how the features of the platform (and the data) were direct responses to the community's problem-solving activities (Stillman and Linger, 2009).

Taking a Digital Civics (Olivier and Wright, 2015; Vlachokyriakos et al., 2016) approach to designing and building CI technologies provided an opportunity to expand on the ISD framework of Hirschheim et al. (1996) to include more social, community and activist agendas, in addition to also enabling a better theoretical framing of CI through the model of CDI (Chapter 5), expanding the uses of technology further from just being the instrumentations of control, as presented by Hirschheim et al. (1996) and extended by Stillman and Linger (2009), to enabling people to find uses of technology in other orientations, such as *strategising*, *sense-making* and *argumentation*. The findings from Case Study III (Chapter 6) highlighted that technology can also find use in discussing and deliberating around issues by using *Ambit* in an issue mapping activity

[9] https://sensemystreet.uk/sensors

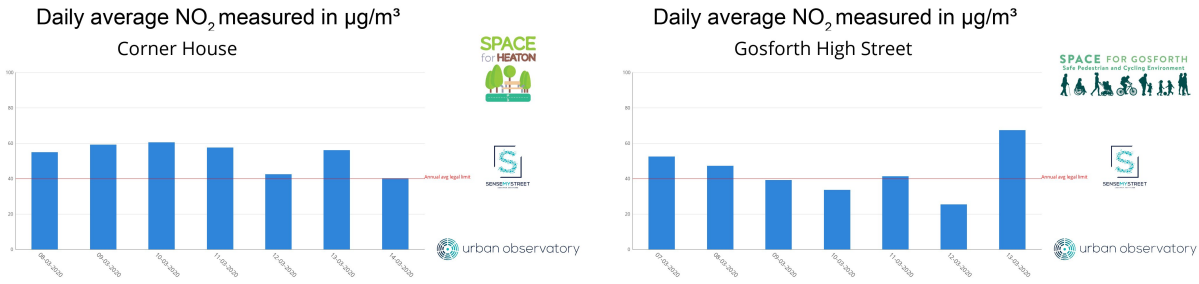
or communicating issues to others by using Data:In Place data requests and data challenges. Additionally, technology aided in capturing some of the context (e.g. metadata and paradata on the Data:In Place platform) needed to understand the challenges of communities in order to adequately respond to them without necessarily engaging in face-to-face talk. Furthermore, according to Hakken (2003), finding ways to use AITs to facilitate interactions that redeem face-to-face talk should be the main task of knowledge engineering by communities.

The Sustainability of CDI for CI Research

Having the capacities of the researcher on hand to constantly provide support to communities puts the sustainability and longevity of these resources and systems at risk. This issue around sustaining a community-commissioned information resource was also raised by Garbett (2017), emphasising the need to find a community that will *adopt and sustain* the asset in the future, keeping the information resources up to date and relevant to the community. However, the research in this thesis went a step further to look at how not only to enable citizens to produce community-commissioned information resources, but also to make effective use of those resources. Positing CAF, a new model for civic advocacy and action, enabled a better understanding of the key factors that influence citizens effectively using commissioned data sources, and provided a framework to help plan focused advocacy efforts in the future. Similarly, the findings of the Case Study II evaluation (Chapter 3) showed that in order to sustain the engagement around the commissioned data resources, there needs to be a transfer of ownership to the community; however, it also has to be supported by building capacities in order for the community to make effective use of the data. For some of the groups involved, these capacities were found through existing links and networks that enabled them to tap into new skills and competences, such as data science skills for making easily digestible visualisations for the community (Figure 7.1). Another solution proposed in Case Study III (6) for cases when the community lacked these links was helping to establish connections through the use of an online platform (e.g. Data:In Place's Data Challenges) that could help interface between communities and data professionals. Similarly, there have been initiatives that 'rent out' data scientist[10] to help communities leverage these skills. However there would always be a reliance on the researcher (or the software developer) to support and sustain these technologies (e.g. the Data:In Place and SMS toolkits).

The SMS toolkit was officially launched in the summer of 2017. Since then, over 60 people have signed up, seven groups of people have carried out environmental sensing with hand-held monitors loaned from the toolkit on their commutes to explore issues in their local areas, and four communities have successfully commissioned stationary monitors to be deployed in their neighbourhoods, with more people signed up and waiting in queue to get access to a monitor. However, the research for this thesis was coming to a conclusion, putting the future use of these community infrastructures and resources in jeopardy. In order to continue to sustain these community resources, there was a need to transfer ownership to the community or

[10] https://futurecities.catapult.org.uk/cityx/data-science-fellowships/



(a) Daily air quality reports @AirHeaton **(b)** Daily air quality reports @AirGosforthHiSt

Figure 7.1 Daily reports of local air quality from commissioned sensors

another supporting facility. As a result, the SMS toolkit is now being transferred to the UO[11] research facility, and it will continue to be sustained as a community resource in the future. Furthermore, there has been increasing interest from other national observatories linked to the UK Collaboratorium for Research in Infrastructure & Cities (UKCRIC)[12] project to adopt similar models to their local contexts.

For the Data:In Place platform, which was developed for more generic uses of CDI, there is not a particular facility that could adopt and sustain the technology in the future. However, the technologies built in Data:In Place (Chapter 4) and the lessons learned have already been applied to create a new community information resource[13] that is of interest for a multiplicity of stakeholders – citizens, policy officials and public service providers. Furthermore, the nature of these technologies leveraging open-source data[14] and being published as open resources has the potential to be picked up by others and expanded upon[15].

7.2.3. Guiding Principles for CDI

> (*Pragmatic objective*): Provide a model of practice for interactions with data in knowledge networking at the community level that can be applied when designing and building tools for the effective use of data by citizens for civic participation, advocacy and action.

The concept of the *'effective use'* of ICT and data put forward by Gurstein (2003, 2011), which later became the main aim of CI (Gurstein, 2007), has been around for over a decade now, and there have been different theoretical bases proposed for CI over the years. For example, Stillman and Linger (2009) proposed that CI should be adopted as a part of a *'fragmented adhocracy'* of IS, which enables the theoretical basis of IS to the field of CI and also helps to expand IS design to include more social and community agendas. This aligns with the views of Carroll and Rosson (2007) on CI research, which looks at the design and management of IS and infrastructures for hyper-local level entities. However, they put a much larger emphasis on

[11] https://urbanobservatory.ac.uk/explore
[12] https://www.ukcric.com
[13] https://feed-finder.co.uk
[14] https://data-in.place/open-source
[15] https://github.com/300000kms/rawmaps

participatory design activities for undertaking this type of research within communities. Another attempt to develop a theoretical basis for CI was proposed by Gurstein (2012) by looking at the work and case studies conducted as part of the Alliance for Community Innovation and Networking (CRACIN) in a Canadian context. His realisation, however, was that because work in CI is applied and case study specific, it is difficult to develop a definitive theoretical framing for the fields. Instead, we should use theory as a pragmatic practice or process in order to help conceptualise and understand the factors and processes influencing the adoption and use of technologies by communities. Similar to Gurstein (2012), Moor (2009) has stated that CI research is a wide-ranging field that is predominantly *'practice built'*, with connections to several methodological pillars. For Moor (2009), the connecting elements between these pillars are '*collaboration patterns'* that can be applied in CI practice. Furthermore, Stillman and Denison (2014) have argued that because of the activist and social justice nature of CI research, we should adopt more critical social theory and approaches such as the *Capability Approach*, which would broaden the research from having a narrow technical orientation to having a more comprehensive critical and social lens, thus enabling research agendas to be expanded upon to examine the long-term effects and adoption of the technologies by communities.

It seems that there is no 'one' theoretical framing that would work with all the research agendas surrounding CI. However, most scholars in CI seem to agree that the way to extend the theoretical basis of CI is through analysing practical approaches. For this reason, the research in this thesis took an investigative approach through two case studies (Chapters 2, 3 and 4), looking at different aspects and elements of the effective use of data by communities, and provided practical examples of the design and implementations of technologies. Furthermore, reflecting on and analysing the findings from the first two case studies enabled the author to expand upon the theoretical basis of CI by outlining a model of CDI (Chapter 5), which provided a set of guiding principles for thinking about and designing technologies and processes for the effective use of data by communities, which were then put into practice in Case Study III (Chapter 6). Taking a more holistic view to CI research in practice, and using a dual approach for the development of CI tools and methods, enabled the thesis to illustrate the use of technologies not only for access and the use of data but also for the training, education and organisational support aspects relating to the effective use of data by communities (Gurstein, 2003, 2012), resulting in a better understanding of the ways that people interact with data within complex sociotechnical systems and helping to identify and underline the existing challenges. This aids further research in CI around the effective use of data not only to understand the practical forms that the key elements of CDI may take, but also to give indications on how to apply new strategies for supporting communities to make effective use of data. Furthermore, analysing not only technical aspects but also the social and organisational impacts on CDI will enable researchers to make better design choices for tools and methods for the purposeful and effective use of data in civic participation.

Thus, the research carried out in this thesis has contributed to and documented the practical cases of participatory design processes for the effective use of data by communities (Case Study I and II); helped further extend the design and development of ISs for community use (e.g. the

Data:In Place and SenseMyStreet platforms); provided a wider conceptualisation around the effective use of data by communities through a CDI model (Chapter 5) that extends on the theoretical basis of IS, CI and community commissioning; and showed how one would design technologies and methods in the frame of the CAF and CDI model and for the purposes of community action (Case Study III in Chapter 6). The reminder of this chapter is dedicated to looking at new opportunities for designing CDI – new ways of thinking around the use of ICT for data discovery and use and ways that we could truly democratise data science practices.

7.3. The Future Research in CDI

Although we all live in the 'information society', the way that each individual (or community) can benefit from it differs marginally. There are rising tensions between those who have access and can effectively use new technologies and information sources for economic and social benefits and those who lack such means (Gurstein, 2012). The research in this thesis looked at ways we can uncover and start identifying this divide in society by understanding ways to enable better uses of ICT and data for those who might benefit from it the most. However, in order to start addressing these issues on a large scale across communities, there needs to be an ability to scale up this research and the practical uses it has produced. This could also mean adopting new ways of thinking about data and data use and reuse.

One approach for future developments of CDI systems could be to design features that promote exploration and curiosity, accompanied by using embedded context, to improve the retrieval of useful information from different data sources. This criterion is important because people do not acquire information in the manner that machines do (e.g. through semantic links); to satisfy their curiosity and the need for information on the web, people posit questions that are often open-ended and do not map to any specific structure. With this in mind, Paritosh (2018) has theorised that instead of a 'web of data' or Semantic Web, *'what if we rethink the web with the primary goal of fostering and satisfying human curiosity?'*. Furthermore, how would that affect the way we retrieve data from the web?

Currently, people need the skills and tools of data scientists who know where to look and can plug into the APIs of data publishers or use automated interfaces that could contextualise that data for people (e.g. the Data:In Place platform). In order for non-experts to start repurposing and using abstract datasets from the web, there is a need for a system that links questions with the data that might bring about the answers. Using this criterion and drawing on previous work and the experiences of the author, in addition to looking at the available infrastructures on the web, it is possible to speculate about how we might reconfigure the way people interact with open data sources on the web and start compiling and making use of this rich dataset of questions with linked open data sources.

7.3.1. Adjacent Possible: State of The Art of Available Infrastructures

It seems that data needs anchors for people to start exploring some of the questions that these datasets answer. Common anchors for open data publishing platforms are things like domains or topics (e.g. transport, health, education and local services), publishing entities (e.g. NHS Digital, local governments, Ordnance Survey and ONS) and data access formats (e.g. machine readable files, APIs and PDF files). Related work has also shown that place can be a valuable anchor for accessing and exploring datasets (Puussaar et al., 2018; Taylor et al., 2015). Case Study II (Chapter 4) illustrated that place and situated activities that people engage in there provide valuable context for starting to explore data that might answer questions about those places. It seems that in order to make data *useful*, it needs to be somehow contextualised – made relevant to people and their situations. Anchoring in this sense is just another form of grounding or contextualising the data. Furthermore, Chapter 5 presented further practical approaches for contextualising by linking data to *issues* and *inquires*.

Despite the fact that data can be anchored in place, linked to issues and inquires, because of how it has been generated and archived for access, it still falls short of having a direct route for answering questions people come up with every day. In searching for information on the web, people often rely on the experiences of others finding out answers to similar questions. The use of online forums like Quora[16] and Reddit[17] and problem-solving platforms like StackOverflow[18] are just a few examples of places people go to look for answers. Often, the information or datasets people are seeking are already out there, but someone (often a professional data analyst) needs to make the link between the questions and the abstract datasets, which often requires an understanding of how the data was compiled or collected and what hidden values it represents. Furthermore, the link between the actual data that answered the question might get lost in a noisy forum thread.

From the software development processes, people might be familiar with issue tagging when they discover a bug or come up with an idea for improving the software. With the increasing popularity of the open-source software movement and people having access to open code repositories such as GitHub[19], issue tagging functionalities are becoming more and more similar to forum- or question & answer (QA)-based interactions with the systems. We can already see projects on GitHub that leverage this infrastructure to compile and archive links to open datasets, e.g. the Awesome Public Datasets[20] project, which compiles a topic-centric list of open datasets in a public domain. It uses Github issues and pulls requests to receive new proposals for dataset links. The project has tens of thousands of followers, and the majority of people who contribute to the datasets already seem to know what data they are looking for and how to find it. However, a brief observation reveals that some of the issues are actually more rephrased as questions without any link or routes to data, e.g. *'How can I apply for a driver's job in Canada/USA'*. This

[16] https://www.quora.com
[17] https://www.reddit.com
[18] https://stackoverflow.com
[19] https://github.com
[20] https://github.com/awesomedata/awesome-public-datasets

exact request posted on the *Issues* already has some of the anchors mentioned previously: (1) a place - Canada or USA and (2) an issue or a topic - working as a driver. There have been a number of other projects that make use of the open-source community and the capabilities of platforms such as GitHub. Similar ideas were applied in Case Study III (Chapter 6), where people used the Data:In Place platform's *Data Challenges* to posit questions for data. However, the platform automatically collected some of the contextual data – the place of the challenge issuer, the drawn boundary, and the interactions with existing datasets on the platform. Furthermore, the challenge description was automatically translated into Markdown language, which is often what these issue-tagging and QA systems (e.g. GitHub) use. With these opportunities, we could start leveraging the ideas and infrastructure of the open-source ecosystem to organise, archive and make accessible this contextual information, which helps people link their questions to datasets that may hold the answers.

7.3.2. Human Computation Approach to Organising Data on the Web

The problem with using already available structured data systems that organise information on the web is that the questions people come up with do not obey any laws of structured data sources (e.g. Linked Data and JSON-LD). Questions that people come up with are not polished queries for exact pieces of information but are more driven by people's needs and curiosity (Paritosh, 2018). Often, these questions need to be understood by someone else with the knowledge about the possible information to address them, i.e. a data scientist. It has also been reported that the majority, about 79%, of data scientist work is actually 'janitorial' – collecting, untangling, matching and cleaning data[21]. Furthermore, having somebody else make decisions about what is important can also reveal some further problems around the multiplicity of views and biases of people thinking they have the correct link to information answering the proposed questions. There is no single truth – there is only a matter of perspective corresponding to the values of the person or group looking for answers. Nevertheless, the first step for organising this information and making it accessible for re-enquiry is to start documenting these 'knowledge enquiries' and routes to data where the answer was found. Having this laid out for others to learn from might make their discovery of information a little easier.

Challenges

The first challenge is how to make these links between data and questions useful. The current analysis of previous work suggests that there is a need for new mechanisms for people to posit questions and have a number of previously asked questions generated. That would help link similar questions together and also show the route of discovery and data sources for those that have already answered. For that tagging, merging and hierarchies, with a combination of HITS or NLP algorithms (Section 5.2.3 in Chapter 5), could be used to automatically make these associations. We can already see this first part happening on the Google Search Engine through

[21] 2016 CrowdFlower Data Science Report. Available at: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf

the *People Also Ask* section. However, this needs to be extended to not only linking questions to each other and links to websites but also links to actual datasets on the web where one might find the answer. Building on this, Google has recently come out with a new dataset Search[22]. However in order for the dataset search to discover a particular dataset on the web, it needs to be described through adding metadata in schema.org. From there, it could be possible to start linking questions to datasets. Furthermore, there is a need for a validation process or feedback loop, which likely involves human intelligence, to assess the usefulness of retrieved datasets for each question. This could take the form of a QA or forum site that connects questions to datasets on the web. A second challenge is what the scaling strategies of this solution are. Through the use of Data:In Place's *Data Challenges*, Case Study III (Chapter 6) illustrated how we might start collecting this corpus of data of questions and linked datasets on a local community level with isolated questions. However, these links often had to be made by a professional data scientist (i.e. the author) who knew where to look for data and how to put it into a form that could be used by communities. Moving forward, there needs to be some type of process that enables and incentivises data professionals to start making these links and answering these questions with datasets. One option could be to connect these challenges to a learning and skills training platform such as mooqita.org. This would enable data scientists in training to take up the challenges and provide solutions and datasets; in addition, if companies who are hiring could steward the challenges, it would offer monetary value for challenge-takers in addition to the social credit.

7.3.3. Engineering Approach to Organising Data on the Web

On the other hand, perhaps using big data and building better algorithms and contextually aware AIs could solve this problem of human-data interaction. The affordances of IoT and sensing technologies that constantly record large quantities of data (e.g. big data) could emulate some of those processes at the level of syntax and associations. Similar to using the standards of the Semantic Web to describe and organise data on the web for better access and use of data, computer algorithms and AI could possibly emulate this and organise it for people.

In the early days, AI development was built on the assumption that in order for computers to understand natural language, they needed to understand grammar and semantics. This meant that people were developing semantic, syntactic and pragmatic theories of language and implementing them into computer systems so that the systems could use those rules of grammar to parse the syntax and the semantics of language and work out what the pragmatics were. For example, people implemented online lexicons such as WordNet[23] to advance their AI systems in understanding context and pragmatics of language. However, the constant flow of single sentences from people's personal devices, such as Google Assistance, Amazon Alexa and Apple Siri, accompanied by contextual metadata (e.g. people, places and situations) has accumulated a massive dataset for training and improving the responses of AI. Furthermore, the way that these

[22] https://datasetsearch.research.google.com
[23] https://wordnet.princeton.edu/citing-wordnet

technologies have become more personalised (e.g. understanding unique voices and people) and more aware of historical interactions between each individual enables them to give information and answers unique to a person. This seamless feedback loop between individuals and devices could potentially emulate the processes of inquiry and replace the work of data scientists (or the Semantic Web) making these connections. However, a deeper discussion is needed around these technologies and whether they are actually helping people achieve control or giving it away to private hands.

7.4. Summary

This chapter reflected on the aims and objectives of the thesis and discussed the findings of the research and its contributions and implications to the wider fields of research in IS engineering, HCI and CI. Furthermore, the outcomes and sustainability of the built technologies were discussed, offering solutions for further use and adoption by communities. Finally, future research in CI around CDI was discussed, focusing on the adjacent possibility of current digital infrastructures and potential solutions to scale up research efforts to better address the effective use of data by communities. The next chapter will present a summary of the contributions of this thesis and concludes by answering its research questions.

Chapter 8. Conclusions

I had nothing to offer anybody except my own confusion.

— Jack Kerouac

This chapter summarises the contributions of the research to the multiple fields of HCI, IS engineering and CI and concludes by offering short answers to the research questions introduced in the Introduction.

8.1. Contributions

8.1.1. HCI

Despite a growing body of research in HCI focusing on the design of inclusive technologies for civic participation, there is a lack of research with regard to understanding what enables citizens to become active users of data and technologies for participating in decision-making. This thesis has established an understanding of what makes data useful (i.e. effective use) for non-expert citizens and communities, contributing new knowledge about the relationships between data, communities, civic participation, advocacy and action. Through PADRE, which is a common approach for studies in HCI, and applying engineering sprints to develop and deploy technologies, this study actively responded to community problem-solving activities. Furthermore, by analysing the way people made use of community data sources, new understandings of civic advocacy and action were uncovered, which were condensed into a CAF that helped to discuss advocacy efforts and plan for future activities relating to local action by citizens (Chapter 3). However, the analysis and evaluation of these technologies also revealed the temporality of such solutions responding to ever-changing communities and their activities. Building on community commissioning research (Garbett, 2017), this thesis extended the digital civics approach (Asad and Le Dantec, 2017; Olivier and Wright, 2015; Vlachokyriakos et al., 2016) with the inclusion of designing for building community capacity as a step towards the successful transfer of ownership and the adoption of community information resources (Chapters 3 and 5), in addition to presenting a deeper understanding of the role of expert professionals and instructional designers or researchers in all of this.

8.1.2. IS

Taking a human-centred design approach to designing and developing community IS, this research documented the design processes and also provided two open-source platforms for IS:

(1) the SMS toolkit (Chapter 2), which is a set of tools and platforms that enable communities to use scientific-grade environmental monitors to produce and commission data about their neighbourhoods and (2) Data:In Place (Chapter 4), an open-source web platform built to support citizens in accessing, interpreting and using open data for the purposes of civic advocacy and action. Furthermore, through the community problem-solving activities and the user-centred design processes, the thesis contributed new ways to accommodate more social agendas into the design and development of future IS.

8.1.3. CI

This thesis provided a new model of CDI for the field of CI that helps discuss and design community resources for knowledge making and civic action through the effective use of data by citizens (Chapter 5). CDI extends the view relating to the design of CI tools from focusing only on the access and instrumentation of control (Stillman and Linger, 2009) to looking at how technologies could be used also to communicate ideas, share knowledge, create community links and increase capacities for the effective use of data by communities. The thesis also contributed findings from the practical use of CDI as a technical solution and social facilitation method that support communities making effective use of data for civic advocacy and action (Chapter 6). This provides practical examples for the key elements of CDI to take, enabling researchers to better adopt the model for further research in CI and CDI.

8.2. Answering the Research Questions

What is involved in citizens creating actionable knowledge using data?

The assumption about a lack of skills, access or tools to use data being the only reason for communities not leveraging that data is not completely true. However, what was learned from the research was that building better systems for the access and use of data does not necessarily increase data's actionability by communities. In the same way, providing new ways for citizens to actively produce or commission data will not guarantee the effective use of that data by citizens. The usefulness of data for citizens is dictated by the activities they plan to carry out and what they want to achieve. Citizens need a purpose for using the data, and only then can it become useful to them and be put in action. Furthermore, such purposes need to be integrated into the investigation and issue enquiry process at an early stage.

How do we build and configure tools and processes to support knowledge making and enable the effective use of data by citizens in civic advocacy and action?

In order to enable communities to effectively use data for the local benefit, we need to move away from building tools solely for accessing and using data and start looking at ways to build tools for social facilitation around the use of data. This includes providing training, methods and digital tools to help people articulate their concerns, communicate and discuss ideas, understand what resources are needed and plan steps for action. Furthermore, some of these tools are already

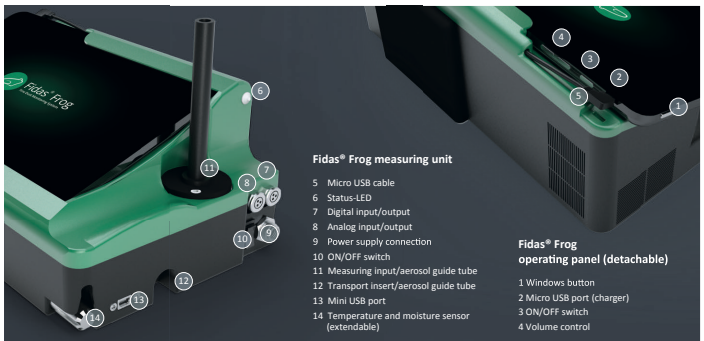
used by communities for other purposes, which could be reconfigured to serve communities' problem-solving agendas. Indeed, there is a need for increased capacities for communities and either the transfer of or support of data science skills. However, this does not mean coming in and doing the work for the communities. In order to maintain the engagement and use of community information resources and solutions, the community needs to be able to take ownership and, through increasing community links, adapt to a constantly changing social economy.

Appendix A. Guide for Using Hand-held Monitor

SENSEMYSTREET
User Manual
V2.0

SENSEMYSTREET
DATA FOR EVERYONE

SenseMyStreet is a toolkit developed in collaboration between Open Lab and Urban Observatory at Newcastle University. Toolkit enables citizens and local communities commission sensors from Urban Observatory and locate them on their local streets to collect evidence for informing or even changing the community. For the first time, the public has access to precision equipment to gather data relevant to their local issues.



Fidas® Frog measuring unit

5 Micro USB cable
6 Status-LED
7 Digital input/output
8 Analog input/output
9 Power supply connection
10 ON/OFF switch
11 Measuring input/aerosol guide tube
12 Transport insert/aerosol guide tube
13 Mini USB port
14 Temperature and moisture sensor (extendable)

Fidas® Frog operating panel (detachable)

1 Windows button
2 Micro USB port (charger)
3 ON/OFF switch
4 Volume control

Charge the batteries

- Plug the power supply of the Fidas® Frog measuring unit into the power supply connection (9).
- Plug the Micro USB cable (5) into the Micro USB port (2) on the Fidas® Frog operating panel.
- Plug the power supply into the mains outlet.
  - ✔ The battery in the Fidas® Frog measuring unit is recharged.
  - ✔ The battery in the Fidas® Frog operating panel is recharged.

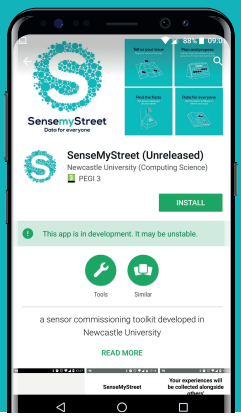
What's Included?

- Fidas® Frog measuring unit, Fidas® Frog operating tablet, aerosol guide tube
- Power supply for the Fidas® Frog measuring unit
- Charger for the operating panel
- Transport bag
- SenseMyStreet High-Vis
- User Manual

Download the app

***if your device has a Bluetooth button**

1. Go to sensemystreet.uk/app and sign up to be a beta tester.
2. After that you can download the application of "Google Play Store" by using the your Android phone. Just search for "sensemystreet".
3. After the installation, the app will guide you how to register and use the application with the Bluetooth button attached to the measuring unit.



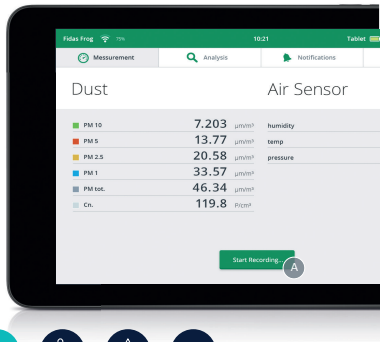
Preparing to use the monitor

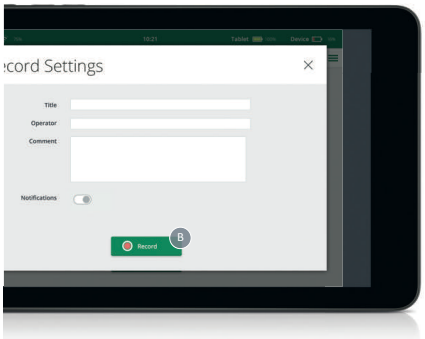
- Insert the aerosol guide tube into the measurement input (11).
- Switch on the Fidas® Frog measuring unit at the ON/OFF switch (10).
- Switch on the Fidas® Frog operating panel at the ON/OFF switch (3).
  - ✔ The Fidas® Frog operating panel boots up.
  - ✔ After the booting process the starting screen is displayed and the fine dust measurement starts automatically (measurement data is not recorded).

Starting screen

Press [Start Recording] (A) to start recording.

- ✔ A query window appears in which you can enter details about the measurement.





Query window

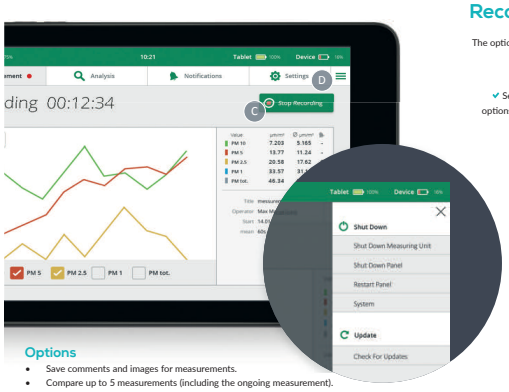
Enter [Title] - date and time on the "Keyboard Input".
Leave the [Operator] as default.
Press [Record] (B) to start the recording.

- ✔ The query window will close and the measurement is displayed.

Recording screen

The option [Stop Recording] (C) is used to stop the recording.

✔ Select [Menu] (D) to display the options for switching off the system.



Options

- Save comments and images for measurements.
- Compare up to 5 measurements (including the ongoing measurement).
- Save limit values.
- Export measurement data.

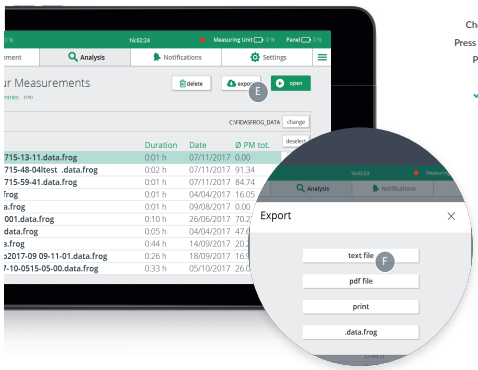
Export screen

Choose the file you would like to export.

Press [Export] (E) to start exporting the file.

Press [text file] (F) from export pop-up.

✔ An export window appears in which you can specify the export details.

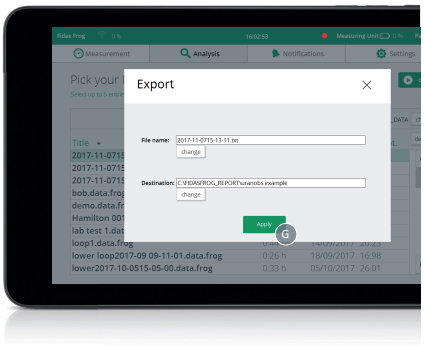


Export window

You can leave everything as default.

Press [Apply] (G) to export the reading.

✔ Your recording will be exported to folder and synced to cloud storage.



UK and EU Air Quality Limits

National air quality objectives and European Directive limit and target values for the protection of human health

Pollutant	Applies	Objective	Concentration measured as	Date to be achieved by (and maintained thereafter)	European Obligations	Date to be achieved (by and maintained thereafter)
Particles (PM_{10})	UK	50 µg/m³ not to be exceeded more than 35 times a year	24 hour mean	31 December 2004	50 µg/m³ not to be exceeded more than 35 times a year	1 January 2005
	UK	40 µg/m³	annual mean	31 December 2004	40 µg/m³	1 January 2005
	Indicative 2010 objectives for PM_{10} (from the 2000 strategy and Addendum) have been replaced by an exposure reduction approach for $PM_{2.5}$ (except in Scotland – see below)					
	Scotland	50 µg/m³ not to be exceeded more than 7 times a year	24 hour mean	31 December 2010	50 µg/m³ not to be exceeded more than 35 times a year	1 January 2005
	Scotland	18 µg/m³	annual mean	31 December 2010	40 µg/m³	1 January 2005
Particles ($PM_{2.5}$) Exposure Reduction	UK (except Scotland)	25 µg/m³	annual mean	2020	Target value - 25 µg/m³	2010
	Scotland	10 µg/m³		31 December 2020	Limit value - 25 µg/m³	1 January 2015
	UK urban areas	Target of 15% reduction in concentrations at urban background		Between 2010 and 2020	Target of 20% reduction in concentrations at urban background.	Between 2010 and 2020

Air Quality Limits. Source: Department for Environment Food & Rural Affairs uk-air.defra.gov.uk/air-pollution/uk-eu-limits

Working on behalf of people

Aare Puussaar - Open Lab
a.puussaar2@newcastle.ac.uk

Philip James - Urban Observatory
urbanobservatory@newcastle.ac.uk

This project is developed in collaboration between Open Lab and the Urban Observatory at Newcastle University; it is funded by Engineering and Physical Sciences Research Council (EPSRC)

Open Lab
Newcastle University
urban observatory
EPSRC Engineering and Physical Sciences Research Council
UKCRIC

Appendix B. Community Action Day Worksheets

DATA:IN PLACE RQI

Democratising Community Action

Citizens taking power of data

Question Formulation Technique (QFT) - PART 1

Rules for Producing Questions

- Ask as many questions as you can
- Do not stop to discuss, judge or answer the questions
- Write down every question exactly as it is stated
- Change any statement into a question

What is Your Focus?

Open Lab Newcastle University

DATA:IN PLACE RQI

Democratising Community Action

Citizens taking power of data

ASKING QUESTIONS

1. Follow the Rules for Producing Questions.
2. Number your questions.

Open Lab Newcastle University

DATA:IN PLACE RQI

Democratising Community Action

Citizens taking power of data

IMPROVE YOUR QUESTIONS

Categorize questions as Closed or Open-ended:

- Closed-ended questions can be answered with "yes" or "no" or with one word.
- Open-ended questions require an explanation and cannot be answered with "yes" or "no" or with one word.

Find closed-ended questions.
Mark them with a "C."
The other questions must be open-ended.
Mark them with an "O."

Discuss the value of each type of question:

Advantages & disadvantages of closed-ended questions
Advantages & disadvantages of open-ended questions

Open Lab Newcastle University

DATA:IN PLACE RQI

Democratising Community Action

Citizens taking power of data

Change questions from one type to another:

Change one closed-ended question to open-ended.

Change one open-ended question to closed-ended.

Open Lab Newcastle University

DATA:IN PLACE RQI

Democratising Community Action

Citizens taking power of data

PRIORITISE YOUR QUESTIONS

Choose your three most important questions:

Q1

Q2

Q3

Why did you choose these three as the most important?
What are the numbers of your priority questions?

DATA:IN PLACE RQI

Democratising Community Action

Citizens taking power of data

Creating a Challenge - PART 2

What data and other resources (technology, people, skills) are needed?

What do we already know?
What data do we need to answer these questions:

- What is out there/what exists?
- How to get access to it?
- Do we need to collect something ourselves?

DATA:IN PLACE RQI

Democratising Community Action

Citizens taking power of data

How do turn the data into something useful?

Using pen and paper please try to sketch the visual representation of the output – what would you expect the data to look like?

DATA:IN PLACE RQI

Democratising Community Action

Citizens taking power of data

Making data actionable!

How can we use the sensemaking and the visualisations to take action?

Appendix C. Data Consultation Worksheets

Data:In Place - Data Consultation

Citizens taking power of data

PART 1

What do we want to investigate?

H0 - Null hypothesis (statement we want to disprove)

H1 - Alternative hypothesis (statement we want to show is true)

H2 - Alternative hypothesis (statement we want show is true)

Data:In Place - Data Consultation

Citizens taking power of data

What data and other resources (technology, people, skills) are needed?

What do we already know?

What data do we need to answer these questions?

1) What is out there/what exists?
2) How to get access to it?
3) Do we need to collect something ourselves?

Data:In Place - Data Consultation

Citizens taking power of data

How do turn the data into something useful (Visualising)?

Using pen and paper please try to sketch the visual understanding of the output – what would you expect the data to look like?

Data:In Place - Data Consultation

Citizens taking power of data

Making data actionable!

How can we use the sensemaking and the visualisations to take action? How can we produce some outputs that could be potentially use for funding bids or reports?

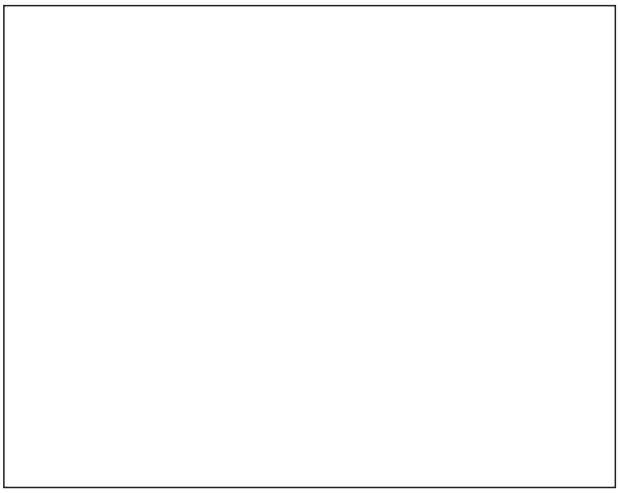
Data:In Place - Data Consultation

Citizens taking power of data

PART 2

Using the tool to investigate hypotheses.

This part is primarily run by me where I would use the tool and try to visualise the data to prove the alternative hypotheses and disprove the null hypothesis.



References

Aanensen, D. M., Huntley, D. M., Feil, E. J., al Own, F., and Spratt, B. G. (2009). EpiCollect: Linking Smartphones to Web Applications for Epidemiology, Ecology and Community Data Collection. *PLoS ONE*, 4(9).

Abowd, G. D. (2012). What next, ubicomp?: celebrating an intellectual disappearing act. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*, page 31, Pittsburgh, Pennsylvania. ACM Press.

Ackoff, R. L. (1989). From Data to Wisdom. *Journal of Applied Systems Analysis*, 16(1):3–9.

Ackoff, R. L. (1999). From Data to Wisdom. In *Ackoff's Best*, pages 170 – 172. John Wiley & Sons, New York, NY, USA.

Aclima (2018). Aclima & Google Scale Air Quality Mapping to More Places Around the World.

Adams, M., Cox, T., Moore, G., Croxford, B., Refaee, M., and Sharples, S. (2006). Sustainable Soundscapes: Noise Policy and the Urban Experience. *Urban Studies*, 43(13):2385–2398.

Afzalan, N. and Evans-Cowley, J. (2015). Planning and Social Media: Facebook for Planning at the Neighbourhood Scale. *Planning Practice & Research*, 30(3):270–285.

Ahas, R., Silm, S., Järv, O., Saluveer, E., and Tiru, M. (2010). Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones. *Journal of Urban Technology*, 17(1):3–27.

Al-Hader, M., Rodzi, A., Sharif, A. R., and Ahmad, N. (2009a). Smart City Components Architicture. In *Modelling and Simulation 2009 International Conference on Computational Intelligence*, pages 93–97.

Al-Hader, M., Rodzi, A., Sharif, A. R., and Ahmad, N. (2009b). SOA of Smart City Geospatial Management. In *2009 Third UKSim European Symposium on Computer Modeling and Simulation*, pages 6–10.

Aldrich, D. P. and Meyer, M. A. (2015). Social Capital and Community Resilience. *American Behavioral Scientist*, 59(2):254–269.

Arnstein, S. R. (1969). A Ladder Of Citizen Participation. *Journal of the American Institute of Planners*, 35(4):216–224.

Asad, M. and Le Dantec, C. A. (2017). Tap the "Make This Public" Button: A Design-Based Inquiry into Issue Advocacy and Digital Civics. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pages 6304–6316, Denver, Colorado, USA. ACM Press.

Asad, M., Le Dantec, C. A., Nielsen, B., and Diedrick, K. (2017). Creating a Sociotechnical API: Designing City-Scale Community Engagement. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pages 2295–2306, Denver, Colorado, USA. ACM Press.

Ausserhofer, J., Gutounig, R., Oppermann, M., Matiasek, S., and Goldgruber, E. (2017). The datafication of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive newswork. *Journalism*, page 1464884917700667.

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, 2(2):205395171559463.

Balaam, M., Comber, R., Jenkins, E., Sutton, S., and Garbett, A. (2015). FeedFinder: A Location-Mapping Mobile Application for Breastfeeding Women. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1709–1718, New York, NY, USA. ACM. event-place: Seoul, Republic of Korea.

Balestrini, M., Diez, T., Marshall, P., Gluhak, A., and Rogers, Y. (2015). IoT Community Technologies: Leaving Users to Their Own Devices or Orchestration of Engagement? *EAI Endorsed Transactions on Internet of Things*, 1(1):150601.

Balestrini, M., Marshall, P., and Diez, T. (2014). Beyond boundaries: the home as city infrastructure for smart citizens. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp '14 Adjunct*, pages 987–990, Seattle, Washington. ACM Press.

Balestrini, M., Rogers, Y., Hassan, C., Creus, J., King, M., and Marshall, P. (2017). A City in Common: A Framework to Orchestrate Large-scale Citizen Engagement around Urban Issues. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pages 2282–2294, Denver, Colorado, USA. ACM Press.

Bandura, A. (1977). *Social learning theory*. Prentice Hall. Google-Books-ID: IXvuAAAAMAAJ.

Bartindale, T., Varghese, D., Schofield, G., and Tsukamoto, M. (2019). Our Story: Addressing Challenges in Development Contexts for Sustainable Participatory Video. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 437:1–437:12, New York, NY, USA. ACM. event-place: Glasgow, Scotland Uk.

Batty, M. (2010). The Pulse of the City. *Environment and Planning B: Planning and Design*, 37(4):575–577.

Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography*, 3(3):274–279.

Bauer, U. E., Briss, P. A., Goodman, R. A., and Bowman, B. A. (2014). Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. *Lancet (London, England)*, 384(9937):45–52.

Behrens, M., Valkanova, N., gen. Schieck, A. F., and Brumby, D. P. (2014). Smart Citizen Sentiment Dashboard: A Case Study Into Media Architectural Interfaces. In *Proceedings of The International Symposium on Pervasive Displays - PerDis '14*, pages 19–24, Copenhagen, Denmark. ACM Press.

Benkler, Y. (2007). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, New Haven London.

Bernarda, J., Ferreira, A. M., Silva, C., and Queiroz, I. (2017). Design as a process tool of collaborative and multidisciplinary learning in society. *The Design Journal*, 20(sup1):S900–S914.

Berners-Lee, T. (2010). The year open data went worldwide.

Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The Semantic Web: A New Form of Web Content That is Meaningful to Computers Will Unleash a Revolution of New Possibilities. *ScientificAmerican.com*.

Berntzen, L. and Karamagioli, E. (2010). Regulatory Measures to Support eDemocracy. In *2010 Fourth International Conference on Digital Society*, pages 311–316. ISSN: null.

Berntzen, L., Rohde Johannesen, M., Böhm, S., Weber, C., and Morales, R. (2018). Citizens as Sensors - Human Sensors as a Smart City Data Source. pages 11–18.

Beyer, H. and Holtzblatt, K. (1997). *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann, San Francisco, Calif, 1 edition edition.

Bilandzic, M. and Venable, J. (2011). Towards Participatory Action Design Research: Adapting Action Research and Design Science Research Methods for Urban Informatics. *Journal of Community Informatics*, 7(3):24.

Binder, T., Michelis, G. D., Ehn, P., Jacucci, G., Linde, P., and Wagner, I. (2011). *Design Things*. MIT Press. Google-Books-ID: XTvKPmnEsk8C.

Bird, J. and Rogers, Y. (2010). The Pulse of Tidy Street: Measuring and Publicly Displaying Domestic Electricity Consumption. In *Workshop on Energy Awareness and Conservation through Pervasive Applications*, page 6.

Björgvinsson, E., Ehn, P., and Hillgren, P.-A. (2010). Participatory design and "democratizing innovation". In *Proceedings of the 11th Biennial Participatory Design Conference on - PDC '10*, page 41, Sydney, Australia. ACM Press.

Boehner, K. and DiSalvo, C. (2016). Data, Design and Civics: An Exploratory Study of Civic Tech. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 2970–2981, Santa Clara, California, USA. ACM Press.

Bolton, A., Butler, L., Dabson, I., Enzer, M., Evans, M., Fenemore, T., Harradence, F., Keaney, E., Kemp, A., Luck, A., Pawsey, N., Saville, S., Schooling, J., Sharp, M., Smith, T., Tennison, J., Whyte, J., Wilson, A., and Makri, C. (2018). Gemini Principles. Technical report, Apollo - University of Cambridge Repository.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., and Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298.

Bontcheva, K. and Rout, D. P. (2014). Making sense of social media streams through semantics: A survey. *Semantic Web*.

Boyte, H. C. (2004). *Everyday Politics: Reconnecting Citizens and Public Life*. University of Pennsylvania Press.

Boyte, H. C. and Fretz, E. (2010). Boyte, Fretz - 2010 - Civic professionalism.pdf. *Journal of Higher Education Outreach and Engagement*, 14(2):67–90.

Brabham, D. C. (2009). Crowdsourcing the Public Participation Process for Planning Projects. *Planning Theory*, 8(3):242–262.

Bradley, G. (2006). *Social and Community Informatics: Humans on the Net*. Routledge, London ; New York, 1 edition edition.

Brady, S. R., Young, J. A., and McLeod, D. A. (2015). Utilizing Digital Advocacy in Community Organizing: Lessons Learned from Organizing in Virtual Spaces to Promote Worker Rights and Economic Justice. *Journal of Community Practice*, 23(2):255–273.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101.

Brown, J. S. and Newman, S. E. (1985). Issues in Cognitive and Social Ergonomics: From Our House to Bauhaus. *Human-Computer Interaction*, 1:359–391.

Calvillo, C. F., Sánchez-Miralles, A., and Villar, J. (2016). Energy management and planning in smart cities. *Renewable and Sustainable Energy Reviews*, 55:273–287.

Cameron William Bruce (1963). *Informal Sociology: A Casual Introduction to Sociological Thinking*. Random House, New York, NY, USA.

Carroll, J. M., Beck, J., Dhanorkar, S., Binda, J., Gupta, S., and Zhu, H. (2018). Strengthening Community Data: Towards Pervasive Participation. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, dg.o '18, pages 65:1–65:9, New York, NY, USA. ACM. event-place: Delft, The Netherlands.

Carroll, J. M. and Rosson, M. B. (2003). A Trajectory for Community Networks Special Issue: ICTs and Community Networking. *The Information Society*, 19(5):381–393.

Carroll, J. M. and Rosson, M. B. (2007). Participatory design in community informatics. *Design Studies*, 28(3):243–261.

Carson, J. (2017). Fake news: What exactly is it – and how can you spot it? *The Telegraph*.

Chandler, D. (2015). A World without Causation: Big Data and the Coming of Age of Posthumanism. *Millennium*, 43(3):833–851.

Chi, E. H. (2000). A taxonomy of visualization techniques using the data state reference model. In *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*, pages 69–75.

Chi, E. H.-h. and Riedl, J. T. (1998). An operator interaction framework for visualization systems. In *Proceedings IEEE Symposium on Information Visualization (Cat. No.98TB100258)*, pages 63–70.

Choi, J. and Tausczik, Y. (2017). Characteristics of Collaboration in the Emerging Practice of Open Data Analysis. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*, pages 835–846, Portland, Oregon, USA. ACM Press.

Chou, S., Li, W., and Sridharan, R. (2014). Democratizing Data Science. In *Proceedings of the KDD 2014 20th ACM SIGKDD*, pages 24–27, New York, NY, USA.

Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K., Pardo, T. A., and Scholl, H. J. (2012). Understanding Smart Cities: An Integrative Framework. In *2012 45th Hawaii International Conference on System Sciences*, pages 2289–2297, Maui, HI, USA. IEEE.

Christoffel, K. K. (2000). Public health advocacy: process and product. *American Journal of Public Health*, 90(5):722–726.

Claes, S. and Vande Moere, A. (2013). Street Infographics: Raising Awareness of Local Issues through a Situated Urban Visualization. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays - PerDis '13*, page 133, Mountain View, California. ACM Press.

Clear, A. K., Finnigan, S. M., Olivier, P., and Comber, R. (2017). "I'd Want to Burn the Data or at Least Nobble the Numbers": Towards Data-mediated Building Management for Comfort and Energy Use. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*, pages 2448–2461, Portland, Oregon, USA. ACM Press.

Clement, A., Gurstein, M., Longford, G., Moll, M., and Shade, L. R., editors (2012). *Connecting Canadians: Investigations in Community Informatics*. AU Press.

Clement, A. and Shade, L. (1998). The Access Rainbow 1 The Access Rainbow: Conceptualizing Universal Access to the Information/ Communications Infrastructure.

Cohn, J. P. (2008). Citizen Science: Can Volunteers Do Real Research? *BioScience*, 58(3):192–197.

Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, 94:S95–S120.

Colomb, C. (2017). Participation and conflict in the formation of neighbourhood areas and forums in super-diverse cities. In Sue Brownill and Quintin Bradley, editors, *Localism and Neighbourhood Planning: Power to the People?*, pages 127–144. Bristol, UK.

Crabtree, A., Lodge, T., Colley, J., Greenghalgh, C., and Mortier, R. (2017). Accountable Internet of Things? Outline of the IoT databox model. In *2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pages 1–6, Macau, China. IEEE.

Cranshaw, J., Schwartz, R., Hong, J. I., and Sadeh, N. (2012). The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. ICWSM'12. *ICWSM 2012 - Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*.

Crivellaro, C., Comber, R., Bowers, J., Wright, P. C., and Olivier, P. (2014). A Pool of Dreams: Facebook, Politics and the Emergence of a Social Movement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 3573–3582, New York, NY, USA. ACM. event-place: Toronto, Ontario, Canada.

Crivellaro, C., Comber, R., Dade-Robertson, M., Bowen, S. J., Wright, P. C., and Olivier, P. (2015). Contesting the City: Enacting the Political Through Digitally Supported Urban Walks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2853–2862, New York, NY, USA. ACM. event-place: Seoul, Republic of Korea.

Crivellaro, C., Taylor, A., Vlachokyriakos, V., Comber, R., Nissen, B., and Wright, P. (2016). Re-Making Places: HCI, 'Community Building' and Change. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 2958–2969, Santa Clara, California, USA. ACM Press.

Dahlbom, B. and Mathiassen, L. (1993). *Computers in Context: The Philosophy and Practice of System Design*. Wiley. Google-Books-ID: DPNjABe2RzQC.

Dameri, R. P. and Cocchia, A. (2013). Smart City and Digital City: Twenty Years of Terminology Evolution. In *Proceedings Itais*, page 9.

Dantec, C. A. L. (2014). Data-Based Civic Participation. In *Second AAAI Conference on Human Computation and Crowdsourcing*.

Dantec, C. A. L., Appleton, C., Asad, M., Rosenberger, R., and Watkins, K. (2016). Advocating through data : Community visibilities in crowdsourced cycling data. In Golub, A., Hoffmann, M. L., Lugo, A. E., and Sandoval, G. F., editors, *Bicycle Justice and Urban Transformation*.

Dantec, C. A. L. and DiSalvo, C. (2013). Infrastructuring and the formation of publics in participatory design. *Social Studies of Science*, 43(2):241–264.

Davies, T. (2010). *Open data, democracy and public sector reform*. PhD thesis, University of Oxford, Oxford.

Davies, T. and Frank, M. (2013). 'There's No Such Thing As Raw Data': Exploring the Socio-technical Life of a Government Dataset. In *Proceedings of the 5th Annual ACM Web Science Conference*, WebSci '13, pages 75–78, New York, NY, USA. ACM. event-place: Paris, France.

Davies, T., Walker, S. B., and Rubinstein, M., editors (2019). *The State of Open Data: Histories and Horizons*. African Minds.

Department for Environment, F. a. R. A. D. w. g. g. u. (2018). Air Pollution in the UK report-Defra, UK. Technical report, London.

Devlin, K. (2001). *Infosense: Turning Information Into Knowledge*. W. H. Freeman, New York, NY.

Dewey, J. (1954). *The Public and its Problems*. Swallow Press, Athens.

Dewey, J. (2016). *The Public and Its Problems: An Essay in Political Inquiry*. Ohio University Press. Google-Books-ID: f3pODQAAQBAJ.

DiSalvo, C., Louw, M., Coupland, J., and Steiner, M. (2009). Local issues, local uses: tools for robotics and sensing in community contexts. In *Proceeding of the seventh ACM conference on Creativity and cognition - C&C '09*, page 245, Berkeley, California, USA. ACM Press.

DiSalvo, C. and Lukens, J. (2009). Towards a Critical Technological Fluency: The Confluence of Speculative Design and Community Technology Programs.

DiSalvo, C., Nourbakhsh, I., Holstius, D., Akin, A., and Louw, M. (2008). The Neighborhood Networks Project: A Case Study of Critical Engagement and Creative Expression Through Participatory Design. In *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008*, PDC '08, pages 41–50, Indianapolis, IN, USA. Indiana University. event-place: Bloomington, Indiana.

Dix, A. (2007). Designing for appropriation. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it - Volume 2*, BCS-HCI '07, pages 27–30, University of Lancaster, United Kingdom. BCS Learning & Development Ltd.

Dodds, L. (2015). What is a data portal?

Dourish, P. (2004). *Where the Action Is*. MIT Press, Cambridge, Mass., new ed edition edition.

Ehn, P. (2008). Participation in Design Things. In *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008*, PDC '08, pages 92–101, Indianapolis, IN, USA. Indiana University. event-place: Bloomington, Indiana.

Eitzel, M. V., Cappadonna, J. L., Santos-Lang, C., Duerr, R. E., Virapongse, A., West, S. E., Kyba, C. C. M., Bowser, A., Cooper, C. B., Sforzi, A., Metcalfe, A. N., Harris, E. S., Thiel, M., Haklay, M., Ponciano, L., Roche, J., Ceccaroni, L., Shilling, F. M., Dörler, D., Heigl, F., Kiessling, T., Davis, B. Y., and Jiang, Q. (2017). Citizen Science Terminology Matters: Exploring Key Terms. *Citizen Science: Theory and Practice*, 2(1):1.

Eltantawy, N. and Wiest, J. B. (2011). Social Media in the Egyptian Revolution: Reconsidering Resource Mobilization Theory. *International Journal of Communication*, 5(0):18.

Epstein, D. A. (2015). Personal informatics in everyday life. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers - UbiComp '15*, pages 429–434, Osaka, Japan. ACM Press.

Epstein, D. A., Ping, A., Fogarty, J., and Munson, S. A. (2015). A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, pages 731–742, Osaka, Japan. ACM Press.

Erete, S. L., Ryou, E., Smith, G., Fassett, K. M., and Duda, S. (2016). Storytelling with Data: Examining the Use of Data by Non-Profit Organizations. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, pages 1271–1281, San Francisco, California, USA. ACM Press.

Evans, C., Abrams, E., Reitsma, R., Roux, K., Salmonsen, L., and Marra, P. P. (2005). The Neighborhood Nestwatch Program: Participant Outcomes of a Citizen-Science Ecological Research Project. *Conservation Biology*, 19(3):589–594.

Ferreira, A. M., Souleles, N., and Savva, S. (2020). Social Design, Innovation and Ergonomics: Reflections on Education, Transdisciplinarity and New Blurred Models for Sustainable Social Change. In Goossens, R. H. and Murata, A., editors, *Advances in Social and Occupational Ergonomics*, Advances in Intelligent Systems and Computing, pages 37–48, Cham. Springer International Publishing.

Finger, F., Genolet, T., Mari, L., de Magny, G. C., Manga, N. M., Rinaldo, A., and Bertuzzo, E. (2016). Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks. *Proceedings of the National Academy of Sciences*, 113(23):6421–6426.

Foth, M. (2009). *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City*, volume 75.

Foucault, M. (1982a). *The Archaeology of Knowledge: And the Discourse on Language*. Vintage, New York, NY, 3988th edition edition.

Foucault, M. (1982b). The Subject and Power. *Critical Inquiry*, 8(4):777–795.

Franklin, R. (2019). Spatial inequality and the smart city.

Furnas, G. W. and Russell, D. M. (2005). Making sense of sensemaking. In *CHI '05 extended abstracts on Human factors in computing systems - CHI '05*, page 2115, Portland, OR, USA. ACM Press.

Gabrys, J. and Pritchard, H. (2018). Just Good Enough Data and Environmental Sensing: Moving beyond Regulatory Benchmarks toward Citizen Action. *International Journal of Spatial Data Infrastructures Research*, page 11.

Gallacher, S., Christian, J., Kalnikaite, V., McCann, J., Prendergast, D., and Bird, J. (2014). Investigating the Challenges of Crowd Sensing Lessons from Zurich. In *Workshop "Structures of Knowledge Co-creation Between Organisations and the Public"*, Baltimore.

Gallent, N. and Robinson, S. (2012). *Neighbourhood Planning: Communities, Networks and Governance*. Policy Press, Bristol.

Garbett, A., Comber, R., Jenkins, E., and Olivier, P. (2016). App Movement: A Platform for Community Commissioning of Mobile Applications. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 26–37, Santa Clara, California, USA. ACM Press.

Garbett, A. T. (2017). *Designing community driven participatory platforms : reconfiguring roles, resources, infrastructure, and constraints for community commissioning*. Thesis, Newcastle University. Accepted: 2018-06-25T13:21:08Z.

García-Peñalvo, F. J., García de Figuerola, C., and Merlo, J. A. (2010). Open knowledge: challenges and facts. *Online Information Review*, 34(4):520–539.

Gigler, B.-S., Custer, S., and Rahemtulla, H. (2011). Realizing the Vision of Open Government Data: Opportunities, Challenges and Pitfalls. In *ICT for Development*, volume 1, Washington DC.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438:900–901.

Gitelman, L., editor (2013). *"Raw data" is an oxymoron*. Infrastructures series. The MIT Press, Cambridge, Massachusetts ; London, England.

Goecks, J., Voida, A., Voida, S., and Mynatt, E. D. (2008). Charitable technologies: opportunities for collaborative computing in nonprofit fundraising. In *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08*, page 689, San Diego, CA, USA. ACM Press.

Golder, S. A. and Macy, M. W. (2011). Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures. *Science*, 333(6051):1878–1881.

Golsteijn, C., Gallacher, S., Capra, L., and Rogers, Y. (2016). Sens-Us: Designing Innovative Civic Technology for the Public Good. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS '16*, pages 39–49, Brisbane, QLD, Australia. ACM Press.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221.

Graham, G. (2014). Too-smart cities? Why these visions of utopia need an urgent reality check.

Graham, G. N., Ostrowski, M., and Sabina, A. B. (2016). Population health-based approaches to utilizing digital technology: a strategy for equity. *Journal of Public Health Policy*, 37(2):154–166.

Granovetter, M. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, 1:201–233. Publisher: [American Sociological Association, Wiley, Sage Publications, Inc.].

Gray, J., Lämmerhirt, D., and Bounegru, L. (2016). Changing What Counts: How Can Citizen-generated And Civil Society Data Be Used As An Advocacy Tool To Change Official Data Collection? Technical report.

Greenfield, A. (2013). *Against the Smart City*. Do projects, New York, NY, USA.

Gurin, J., Bonina, C., and Verhulst, S. (2019). Open Data & the Private Sector. In Davies, T., Walker, S., Rubinstein, M., and Perini, F., editors, *The State of Open Data*. Public draft edition.

Gurstein, M. (2000). Introduction Community Informatics : Enabling Community Uses of Information and Communications Technology.

Gurstein, M. (2003). Effective use: A community informatics strategy beyond the Digital Divide. *First Monday*, 8(12).

Gurstein, M. (2007). *What is Community Informatics (and Why Does It Matter)?* Polimetrica. arXiv: 0712.3220.

Gurstein, M. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday*, 16(2).

Gurstein, M. (2012). Toward A Conseptual Framework For A Community Informatics. In Clement, A., Gurstein, M., Longford, G., and Shade, L. R., editors, *Connecting Canadians: Investigations in Community Informatics*, pages 35–60. AU Press.

Gyongyi, Z., Berkhin, P., Garcia-Molina, H., and Pedersen, J. (2006). Link spam detection based on mass estimation. In *Proceedings of the 32nd international conference on Very large data bases*, VLDB '06, pages 439–450, Seoul, Korea. VLDB Endowment.

Hackler, D. and Saxton, G. D. (2007). The Strategic Use of Information Technology by Nonprofit Organizations: Increasing Capacity and Untapped Potential. *Public Administration Review*, 67(3):474–487.

Haj-Bolouri, A., Bernhardsson, L., and Rossi, M. (2015). Introducing PADRE : Participatory Action Design Research.

Haj-Bolouri, A., Bernhardsson, L., and Rossi, M. (2016). PADRE: A Method for Participatory Action Design Research. pages 19–36.

Hakken, D. (2003). A Social Informatics Approach to Knowledge in Cyberspace. In *The Knowledge Landscapes of CyberSpace*, pages 145–177. Routledge, New York, New York, USA.

Haklay, M. (2013). Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation. In Elwood, S., Goodchild, M., and Sui, D., editors, *Crowdsourcing Geographic Knowledge*, pages 105–122. Springer Netherlands, Dordrecht.

Hansen, D. L., Koepfler, J. A., Jaeger, P. T., Bertot, J. C., and Viselli, T. (2014). Civic Action Brokering Platforms: Facilitating Local Engagement with ACTion Alexandria. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, pages 1308–1322, New York, NY, USA. ACM. event-place: Baltimore, Maryland, USA.

Hartung, C., Lerer, A., Anokwa, Y., Tseng, C., Brunette, W., and Borriello, G. (2010). Open Data Kit: Tools to Build Information Services for Developing Regions. In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*, ICTD '10, pages 18:1–18:12, New York, NY, USA. ACM. event-place: London, United Kingdom.

Hasenfratz, D., Saukh, O., Sturzenegger, S., and Thiele, L. (2012). Participatory Air Pollution Monitoring Using Smartphones. In *2nd International Workshop on Mobile Sensing*, page 5.

Heidegger, M. (1976). *What is Called Thinking?* HarperPerennial, New York etc., reprint edition edition.

Herbert Paul Grice (1975). Logic and conversation. In Cole, P. and Morgan, J. J., editors, *Syntax and Semantics*, volume 3, pages 41–58. Academic Press, New York, NY.

Hespanhol, L., Tomitsch, M., McArthur, I., Fredericks, J., Schroeter, R., and Foth, M. (2015). Vote as you go: blending interfaces for community engagement into the urban space. In *Proceedings of the 7th International Conference on Communities and Technologies - C&T '15*, pages 29–37, Limerick, Ireland. ACM Press.

Hirschheim, R., Klein, H. K., and Lyytinen, K. (1996). Exploring the intellectual structures of information systems development: A social action theoretic analysis. *Accounting, Management and Information Technologies*, 6(1):1–64.

Hoffman, L. H. (2012). Participation or Communication? An Explication of Political Activity in the Internet Age. *Journal of Information Technology & Politics*, 9(3):217–233.

Hossain, M. A., Dwivedi, Y. K., and Rana, N. P. (2016). State-of-the-art in open data research: Insights from existing literature and a research agenda. *Journal of Organizational Computing and Electronic Commerce*, 26(1-2):14–40.

Howard, P. N., Duffy, A., Freelon, D., Hussain, M. M., Mari, W., and Maziad, M. (2011). Opening Closed Regimes: What Was the Role of Social Media During the Arab Spring? SSRN Scholarly Paper ID 2595096, Social Science Research Network, Rochester, NY.

Huck, J. J., Whyatt, J. D., Coulton, P., and Gradinar, A. (2014). Mapping Traffic Pollution Exposure: the Quantified Self. *Proceedings of the GIS Research UK 22nd Annual Conference*, page 11.

Jankowski, P. (2009). Towards participatory geographic information systems for community-based environmental decision making. *Journal of Environmental Management*, 90(6):1966–1971.

Jenkins, T., Le Dantec, C. A., DiSalvo, C., Lodato, T., and Asad, M. (2016). Object-Oriented Publics. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 827–839, Santa Clara, California, USA. ACM Press.

Jennex, M. E. (2017). Big Data, the Internet of Things, and the Revised Knowledge Pyramid. *SIGMIS Database*, 48(4):69–79.

Jennex, M. E. and Bartczak, S. E. (2013). A Revised Knowledge Pyramid. *International Journal of Knowledge Management*, 9(3):19–30.

Jetzek, T., Avital, M., and Bjorn-Andersen, N. (2014). Data-Driven Innovation through Open Government Data. *Journal of theoretical and applied electronic commerce research*, 9(2):15–16.

Johnson, I., Al-Shahrabi, D., and Vines, J. (2020). From Creating Spaces for Civic Discourse to Creating Resources for Action. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, Honolulu, Hawaii, USA. ACM.

Johnson, I. G., MacDonald, A., Briggs, J., Manuel, J., Salt, K., Flynn, E., and Vines, J. (2017). Community Conversational: Supporting and Capturing Deliberative Talk in Local Consultation Processes. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pages 2320–2333, Denver, Colorado, USA. ACM Press.

Johnson, I. G., Puussaar, A., Manuel, J., and Wright, P. (2018). Neighbourhood Data: Exploring the Role of Open Data in Locally Devolved Policymaking Processes. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):83:1–83:20.

Johnson, I. G., Vines, J., Taylor, N., Jenkins, E., and Marshall, J. (2016). Reflections on Deploying Distributed Consultation Technologies with Community Organisations. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 2945–2957, Santa Clara, California, USA. ACM Press.

Kamel Boulos, M. N., Resch, B., Crowley, D. N., Breslin, J. G., Sohn, G., Burtner, R., Pike, W. A., Jezierski, E., and Chuang, K.-Y. S. (2011). Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *International Journal of Health Geographics*, 10(1):67.

Karpf, D. (2010). Online Political Mobilization from the Advocacy Group's Perspective: Looking Beyond Clicktivism. *Policy & Internet*, 2(4):7–41.

Kim, S., Mankoff, J., and Paulos, E. (2013). Sensr: evaluating a flexible framework for authoring mobile data-collection tools for citizen science. In *Proceedings of the 2013 conference on Computer supported cooperative work - CSCW '13*, page 1453, San Antonio, Texas, USA. ACM Press.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1):1–14.

Kitchin, R., Lauriault, T. P., and McArdle, G. (2015). Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. *Regional Studies, Regional Science*, 2(1):6–28.

Klein, C. and Kaefer, G. (2008). From Smart Homes to Smart Cities: Opportunities and Challenges from an Industrial Perspective. In *Next Generation Teletraffic and Wired/Wireless Advanced Networking*, pages 260–260. Springer, Berlin, Heidelberg.

Kleinberg, J. M. (1999). Hubs, Authorities, and Communities.

Kling, R. (2000). Learning About Information Technologies and Social Change: The Contribution of Social Informatics. *The Information Society*, 16(3):217–232.

Koeman, L., Kalnikaite, V., Rogers, Y., and Bird, J. (2014). What Chalk and Tape Can Tell Us: Lessons Learnt for Next Generation Urban Displays. In *Proceedings of The International Symposium on Pervasive Displays - PerDis '14*, pages 130–135, Copenhagen, Denmark. ACM Press.

Kontokosta, C. E. (2017). Urban informatics for social good: definitions, tensions, and challenges. In *Proceedings of the 2nd International Workshop on Science of Smart City Operations and Platforms Engineering - SCOPE '17*, pages 52–56, Pittsburgh, Pennsylvania. ACM Press.

Kuznetsov, S. and Paulos, E. (2010). Participatory Sensing in Public Spaces: Activating Urban Surfaces with Sensor Probes. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, DIS '10, pages 21–30, New York, NY, USA. ACM. event-place: Aarhus, Denmark.

Lakhani, K. R. and Wolf, R. G. (2003). Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. SSRN Scholarly Paper ID 443040, Social Science Research Network, Rochester, NY.

Lamb, R. and Kling, R. (2003). Reconceptualizing users as social actors in information systems research. *MIS Quarterly*, 27(2):197–235.

Lambton-Howard, D., Anderson, R., Montague, K., Garbett, A., Hazeldine, S., Alvarez, C., Sweeney, J. A., Olivier, P., Kharrufa, A., and Nappey, T. (2019). WhatFutures: Designing Large-Scale Engagements on WhatsApp. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–14, Glasgow, Scotland Uk. Association for Computing Machinery.

Lanier, J. (2014). *Who Owns The Future?* Penguin, London.

Larkin, A. and Hystad, P. (2017). Towards Personal Exposures: How Technology Is Changing Air Pollution and Health Research. *Current environmental health reports*, 4(4):463–471.

Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, Cambridge England ; New York, 1st edition edition edition.

Le Dantec, C. A. (2016). *Designing Publics*. The MIT Press, 1st edition.

Le Dantec, C. A., Asad, M., Misra, A., and Watkins, K. E. (2015). Planning with Crowdsourced Data: Rhetoric and Representation in Transportation Planning. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*, pages 1717–1727, Vancouver, BC, Canada. ACM Press.

Leonardi, P. M. and Barley, S. R. (2008). Materiality and change: Challenges to building better theory about technology and organizing. *Information and Organization*, 18(3):159–176.

Leroux, K. (2007). Nonprofits as Civic Intermediaries: The Role of Community-Based Organizations in Promoting Political Participation. *Urban Affairs Review*, 42(3):410–422.

Li, H.-C. (2015). Optimal delivery strategies considering carbon emissions, time-dependent demands and demand–supply interactions. *European Journal of Operational Research*, 241(3):739–748.

Li, I., Dey, A., and Forlizzi, J. (2010). A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 557–566, Atlanta, Georgia, USA. Association for Computing Machinery.

Lin, N. (2004). *Social Capital: A Theory of Social Structure and Action*. Cambridge University Press, Cambridge, United Kingdom.

Lindley, S. E., Thieme, A., Taylor, A. S., Vlachokyriakos, V., Regan, T., and Sweeney, D. (2017). Surfacing Small Worlds through Data-In-Place. *Computer Supported Cooperative Work (CSCW)*, 26(1-2):135–163.

Liu, S. B. (2014). Crisis Crowdsourcing Framework: Designing Strategic Configurations of Crowdsourcing for the Emergency Management Domain. *Computer Supported Cooperative Work (CSCW)*, 23(4):389–443.

Loosen, W., Reimer, J., and De Silva-Schmidt, F. (2017). Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *Journalism*, page 1464884917735691.

Loreto, V., Haklay, M., Hotho, A., Servedio, V. D., Stumme, G., Theunis, J., and Tria, F., editors (2017). *Participatory Sensing, Opinions and Collective Awareness*. Understanding Complex Systems. Springer International Publishing, Cham.

Lowndes, V. and Pratchett, L. (2012). Local Governance under the Coalition Government: Austerity, Localism and the 'Big Society'. *Local Government Studies*, 38(1):21–40.

Madni, A. M. (2007). Transdisciplinarity: Reaching Beyond Disciplines To Find Connections. *Journal of Integrated Design & Process Science*, 11(1):1–11.

Manuel, J., Vigar, G., Bartindale, T., and Comber, R. (2017). Participatory Media: Creating Spaces for Storytelling in Neighbourhood Planning. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pages 1688–1701, Denver, Colorado, USA. ACM Press.

Mark, G. and Semaan, B. (2008). Resilience in Collaboration: Technology As a Resource for New Patterns of Action. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, CSCW '08, pages 137–146, New York, NY, USA. ACM. event-place: San Diego, CA, USA.

Marres, N. (2007). The Issues Deserve More Credit: Pragmatist Contributions to the Study of Public Involvement in Controversy. *Social Studies of Science*, 37(5):759–780.

Marshall, M., Kirk, D. S., and Vines, J. (2016). Accountable: Exploring the Inadequacies of Transparent Financial Practice in the Non-Profit Sector. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 1620–1631, Santa Clara, California, USA. ACM Press.

Maskell, T., Crivellaro, C., Anderson, R., Nappey, T., Araújo-Soares, V., and Montague, K. (2018). Spokespeople: Exploring Routes to Action through Citizen-Generated Data. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, pages 1–12, Montreal QC, Canada. ACM Press.

Mauri, M., Elli, T., Caviglia, G., Uboldi, G., and Azzi, M. (2017). RAWGraphs: A Visualisation Platform to Create Open Outputs. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter - CHItaly '17*, pages 1–5, Cagliari, Italy. ACM Press.

Mayer-Schonderger, V. and Cukier, K. (2013). *Big Data: The Essential Guide to Work, Life and Learning in the Age of Insight*. John Murray, Great Britain, 1 edition.

Mccarthy, J., Wright, P., Friedman, K., and Stolterman, E. (2015). *Taking [A]part*. MIT Press, Cambridge, Massachusetts.

McMillan, D., Engström, A., Lampinen, A., and Brown, B. (2016). Data and the City. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 2933–2944, Santa Clara, California, USA. ACM Press.

Merkel, C. B., Xiao, L., Farooq, U., Ganoe, C. H., Lee, R., Carroll, J. M., and Rosson, M. B. (2004). Participatory Design in Community Computing Contexts: Tales from the Field. In *Proceedings of the Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices - Volume 1*, PDC 04, pages 1–10, New York, NY, USA. ACM. event-place: Toronto, Ontario, Canada.

Michael, M. and Lupton, D. (2016). Toward a manifesto for the 'public understanding of big data'. *Public Understanding of Science*, 25(1):104–116.

Misra, A., Gooze, A., Watkins, K., Asad, M., and Le Dantec, C. A. (2014). Crowdsourcing and Its Application to Transportation Data Collection and Management. *Transportation Research Record: Journal of the Transportation Research Board*, 2414(1):1–8.

Moere, A. V., Tomitsch, M., Hoinkis, M., Trefz, E., Johansen, S., and Jones, A. (2011). Comparative Feedback in the Street: Exposing Residential Energy Consumption on House Façades. In Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., and Winckler, M., editors, *Human-Computer Interaction – INTERACT 2011*, Lecture Notes in Computer Science, pages 470–488. Springer Berlin Heidelberg.

Molloy, J. C. (2011). The Open Knowledge Foundation: Open Data Means Better Science. *PLOS Biology*, 9(12):e1001195.

Moor, A. (2007). Using System Dynamics to Construct Design Theory for Community Information Systems. *J. Community Informatics*, 3.

Moor, A. (2009). Collaboration Patterns as Building Blocks for Community Informatics. In *Proceedings of the 6th Community Informatics Research Network Conferenc*, pages 1–21, Prato,Italy.

Moor, A. and Cindio, F. (2007). Beyond Users to Communities - Designing Systems As Though Communities Matter - An Introduction to the Special Issue. *The Journal of Community Informatics*, 3.

Mortier, R., Haddadi, H., Henderson, T., McAuley, D., and Crowcroft, J. (2014). Human-Data Interaction: The Human Face of the Data-Driven Society. *SSRN Electronic Journal*.

Morton, M. J. and Lurie, N. (2013). Community Resilience and Public Health Practice. *American Journal of Public Health*, 103(7):1158–1160.

Nam, T. and Pardo, T. A. (2011). Smart City As Urban Innovation: Focusing on Management, Policy, and Context. In *Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance*, ICEGOV '11, pages 185–194, New York, NY, USA. ACM. event-place: Tallinn, Estonia.

National Collaborating Centre for Determinants of Health (2013). Let's Talk: Universal and targeted approaches to health equity | National Collaborating Centre for Determinants of Health. Technical report, NS: National Collaborating Centre for Determinants of Health,St. Francis Xavier University, Antigonish.

Newman, G., Graham, J., Crall, A., and Laituri, M. (2011). The art and science of multi-scale citizen science support. *Ecological Informatics*, 6(3):217–227.

Nissen, B. and Bowers, J. (2015). Data-Things: Digital Fabrication Situated within Participatory Data Translation Activities. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 2467–2476, Seoul, Republic of Korea. ACM Press.

Obar, J. A., Zube, P., and Lampe, C. (2012). Advocacy 2.0: An Analysis of How Advocacy Groups in the United States Perceive and Use Social Media as Tools for Facilitating Civic Engagement and Collective Action. *Journal of Information Policy*, 2:1–25.

Olivier, P. and Wright, P. (2015). Digital civics: taking a local turn. *interactions*, 22(4):61–63.

Panigutti, C., Tizzoni, M., Bajardi, P., Smoreda, Z., and Colizza, V. (2017). Assessing the use of mobile phone data to describe recurrent mobility patterns in spatial epidemic models. *Royal Society Open Science*, 4(5).

Parasie, S. and Dagiral, E. (2013). Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. *New Media & Society*, 15(6):853–871.

Paritosh, P. (2018). What if the Primary Goal of the Web was to Foster Curiosity? In *International Semantic Web Conference*.

Pierangeli, I., Nieuwenhuijsen, M. J., Cirach, M., and Rojas-Rueda, D. (2020). Health equity and burden of childhood asthma - related to air pollution in Barcelona. *Environmental Research*, page 109067.

Polanyi, M. (2009). *The Tacit Dimension*. University of Chicago Press, Chicago and London, revised ed. edition edition.

Prakash, A. and Gugerty, M. K. (2010). *Advocacy Organizations and Collective Action*. Cambridge University Press. Google-Books-ID: KZPrtkJibCYC.

Procter, R., Vis, F., and Voss, A. (2013). Reading the riots on Twitter: methodological innovation for the analysis of big data. *International Journal of Social Research Methodology*, 16(3):197–214.

Puiu, D., Barnaghi, P., Tönjes, R., Kümper, D., Ali, M. I., Mileo, A., Parreira, J. X., Fischer, M., Kolozali, S., Farajidavar, N., Gao, F., Iggena, T., Pham, T., Nechifor, C., Puschmann, D., and Fernandes, J. (2016). CityPulse: Large Scale Data Analytics Framework for Smart Cities. *IEEE Access*, 4:1086–1108.

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 - CHI '17*, pages 6936–6942, Denver, Colorado, USA. ACM Press.

Puussaar, A., Johnson, I. G., Montague, K., James, P., and Wright, P. (2018). Making Open Data Work for Civic Advocacy. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):143:1–143:20.

Rainey, J., Montague, K., Briggs, P., Anderson, R., Nappey, T., and Olivier, P. (2019). Gabber: Supporting Voice in Participatory Qualitative Practices. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 377:1–377:12, New York, NY, USA. ACM. event-place: Glasgow, Scotland Uk.

Ratti, C. and Claudel, M. (2016). *The City of Tomorrow: Sensors, Networks, Hackers, and the Future of Urban Life*. Yale University Press, New Haven ; London.

Raymond, E. S. (2001). *The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly Media, 1 edition edition.

Rein Ahas (2010). Mobile Positioning. In Urry, J., Witchger, K., and Monika Büscher, editors, *Mobile Methods*, pages 183–199. Routledge. Google-Books-ID: yk95AgAAQBAJ.

Renee E. Sieber and Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly*, 32(3):308–315.

Rheingold, H. (2000). *The Virtual Community: Homesteading on the Electronic Frontier*. MIT Press. Google-Books-ID: fr8bdUDisqAC.

Ritchie, H. and Roser, M. (2018). Urbanization. *Our World in Data*.

Rocha, H., Ferreira, A. M., and Azevedo, J. (2018). Paradigm Shift in Design Education: An overview on issues and possibilities for change.

Rodger, S., Vines, J., and McLaughlin, J. (2016). Technology and the Politics of Mobility: Evidence Generation in Accessible Transport Activism. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 2417–2429, New York, NY, USA. ACM. event-place: San Jose, California, USA.

Rogers, Y. (2011). Interaction design gone wild: striving for wild theory. *interactions*, 18(4):58.

Rudolf Giffinge, Christian Fertner, Hans Kramar, Robert Kalasek, Nataša Pichler-Milanović, and Evert Meijers (2007). Smart cities – Ranking of European medium-sized cities. Technical report, Centre of Regional Science, Austria, Vienna.

Salim, F. and Haque, U. (2015). Urban computing in the wild: A survey on large scale participation and citizen engagement with ubiquitous computing, cyber physical systems, and Internet of Things. *International Journal of Human-Computer Studies*, 81:31–48.

Sall, M. (2014). Using Charted.

Salovaara, A., Jacucci, G., Oulasvirta, A., Saari, T., Kanerva, P., Kurvinen, E., and Tiitta, S. (2006). Collective creation and sense-making of mobile media. In *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*, page 1211, Montréal, Québec, Canada. ACM Press.

Saunders, T. and Baeck, P. (2015). Rethinking Smart Cities from the Ground Up. Technical report, Nesta.

Schofield, G., Bartindale, T., and Wright, P. (2015). Bootlegger: Turning Fans into Film Crew. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 767–776, Seoul, Republic of Korea. ACM Press.

Schrock, A. R. (2016). Civic hacking as data activism and advocacy: A history from publicity to open government data. *New Media & Society*, 18(4):581–599.

Sheppard, S. A. (2012). wq: a modular framework for collecting, storing, and utilizing experiential VGI. pages 62–69. ACM.

Sieber, R. (2006). Public Participation Geographic Information Systems: A Literature Review and Framework. *Annals of the Association of American Geographers*, 96(3):491–507.

Sizov, S. (2010). GeoFolk: latent spatial semantics in web 2.0 social media. In *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, pages 281–290, New York, New York, USA. Association for Computing Machinery.

Sonntag, U., Schmucker, D., Potier, F., Nurmi, O., Madre, J.-L., Karus, E., Tiru, M., Esko, S., Armoogum, J., Ahas, R., Ilves, M., European Commission, and Eurostat (2014). Feasibility study on the use of mobile positioning data for tourism statistics consolidated report. Technical report, Publications Office, Luxembourg. OCLC: 1044333555.

Sorce, S., Ruggieri, S., Gentile, V., Gentile, A., and Malizia, A. (2017). Human-to-Human Interaction: The Killer Application of Ubiquitous Computing? In Kurosu, M., editor, *Human-Computer Interaction. User Interface Design, Development and Multimodality*, Lecture Notes in Computer Science, pages 86–93, Cham. Springer International Publishing.

SPACE for Heaton (2017). Air pollution in Heaton - The desire for data to help residents make decisions.

Stanton, J. (2014). The Big Society and Community Development: Neighbourhood Planning under the Localism Act. *Environmental Law Review*, 16(4):262–276. Publisher: SAGE Publications.

Star, S. L. and Gasser, L. (1997). *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*. Routledge, Mahwah, N.J, 1 edition edition.

Starbird, K. and Palen, L. (2011). "Voluntweeters": Self-organizing by Digital Volunteers in Times of Crisis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1071–1080, New York, NY, USA. ACM. event-place: Vancouver, BC, Canada.

Starbird, K., Palen, L., Hughes, A. L., and Vieweg, S. (2010). Chatter on the Red: What Hazards Threat Reveals About the Social Life of Microblogged Information. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, CSCW '10, pages 241–250, New York, NY, USA. ACM. event-place: Savannah, Georgia, USA.

Stillman, L. and Denison, T. (2014). The Capability Approach Community Informatics. *The Information Society*, 30(3):200–211.

Stillman, L. and Linger, H. (2009). Community Informatics and Information Systems: Can They Be Better Connected? *The Information Society*, 25(4):255–264.

Su, K., Li, J., and Fu, H. (2011). Smart city and the applications. In *2011 International Conference on Electronics, Communications and Control (ICECC)*, pages 1028–1031.

Suchman, L. A. (1985). *PLANS AND SITUATED ACTIONS: The Problem of human-machine communication*. Palo Alto Research Center (Xerox), Palo Alto, California.

Suchman, L. A. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge University Press, Cambridge ; New York, 2 edition edition.

Suárez, D. F. (2009). Nonprofit Advocacy and Civic Engagement on the Internet. *Administration & Society*, 41(3):267–289.

Swan, M. (2012). Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen. *Journal of Personalized Medicine*, 2(3):93–118. Number: 3 Publisher: Molecular Diversity Preservation International.

Sweeney, D., Regan, T., Taylor, A., and Lindley, S. (2019). A Pixel is Not a Pixel: Simple Data Charting Made Physical. page 3809513 Bytes.

Tan, W., Blake, M. B., Saleh, I., and Dustdar, S. (2013). Social-Network-Sourced Big Data Analytics. *IEEE Internet Computing*, 17(5):62–69.

Taylor, A., Wilson, R., and Agamanolis, S. (2009). Supporting carers in their caring role through design. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA '09*, page 3985, Boston, MA, USA. ACM Press.

Taylor, A. S., Lindley, S., Regan, T., and Sweeney, D. (2014). Data and life on the street. *Big Data & Society*, 1(2):2053951714539278.

Taylor, A. S., Lindley, S., Regan, T., Sweeney, D., Vlachokyriakos, V., Grainger, L., and Lingel, J. (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*, pages 2863–2872, Seoul, Republic of Korea. ACM Press.

Taylor, N. and Cheverst, K. (2010). Creating a Rural Community Display with Local Engagement. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, DIS '10, pages 218–227, New York, NY, USA. ACM. event-place: Aarhus, Denmark.

Taylor, N., Cheverst, K., Wright, P., and Olivier, P. (2013). Leaving the wild: lessons from community technology handovers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1549–1558, Paris, France. Association for Computing Machinery.

Taylor, N., Marshall, J., Blum-Ross, A., Mills, J., Rogers, J., Egglestone, P., Frohlich, D. M., Wright, P., and Olivier, P. (2012). Viewpoint: empowering communities with situated voting devices. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, page 1361, Austin, Texas, USA. ACM Press.

Taylor, P. A. (2005). From hackers to hacktivists: speed bumps on the global superhighway? *New Media & Society*, 7(5):625–646.

Tenney, M. and Sieber, R. (2016). Data-Driven Participation: Algorithms, Cities, Citizens, and Corporate Control. *Urban Planning*, 1(2):101.

Thakuriah, P. V., Tilahun, N. Y., and Zellner, M. (2017). Introduction to Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics. In Thakuriah, P. V., Tilahun, N., and Zellner, M., editors, *Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics*, Springer Geography, pages 1–9. Springer International Publishing, Cham.

Thompson, J. E. (2016). Crowd-sourced air quality studies: A review of the literature & portable sensors. *Trends in Environmental Analytical Chemistry*, 11:23–34.

Thrift, N. (2014). The 'sentient' city and what it may portend. *Big Data & Society*, 1(1):205395171453224.

Townsend, A. M. (2013). *Smart cities: big data, civic hackers, and the quest for a new utopia*. W.W Norton & Company, 1 edition.

Tufekci, Z. (2014a). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *arXiv:1403.7400 [physics]*. arXiv: 1403.7400.

Tufekci, Z. (2014b). Engineering the public: Big data, surveillance and computational politics. *First Monday*.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In *ICWSM*.

Tuomi, I. (1999). Data Is More than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory. *Journal of Management Information Systems*, 16(3):103–117.

Ubaldi, B. (2013). Open Government Data.

Valkanova, N., Walter, R., Vande Moere, A., and Müller, J. (2014). MyPosition: sparking civic discourse by a public interactive poll visualization. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14*, pages 1323–1332, Baltimore, Maryland, USA. ACM Press.

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2):197–208.

Van Schalkwyk, F. and Verhulst, S. G. (2017). The state of open data and open data research. In *The Social Dynamics of Open Data*. African Minds, Cape Town, South Africa.

Vanhoof, M., Hendrickx, L., Puussaar, A., Verstraeten, G., Ploetz, T., and Smoreda, Z. (2017). Exploring the use of mobile phone data for domestic tourism trip analysis. *Netcom. Réseaux, communication et territoires*, (31-3/4):335–372.

Varnelis, K. (2008). *Networked Publics*. MIT Press, Cambridge, Mass.

Vilajosana, I., Llosa, J., Martinez, B., Domingo-Prieto, M., Angles, A., and Vilajosana, X. (2013). Bootstrapping smart cities through a self-sustainable model based on big data flows. *IEEE Communications Magazine*, 51(6):128–134.

Vines, J., Clarke, R., Wright, P., McCarthy, J., and Olivier, P. (2013). Configuring participation: on how we involve people in design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, page 429, Paris, France. ACM Press.

Vlachokyriakos, V., Comber, R., Ladha, K., Taylor, N., Dunphy, P., McCorry, P., and Olivier, P. (2014). PosterVote: expanding the action repertoire for local political activism. In *Proceedings of the 2014 conference on Designing interactive systems - DIS '14*, pages 795–804, Vancouver, BC, Canada. ACM Press.

Vlachokyriakos, V., Crivellaro, C., Le Dantec, C. A., Gordon, E., Wright, P., and Olivier, P. (2016). Digital Civics: Citizen Empowerment With and Through Technology. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*, pages 1096–1099, Santa Jose, California, USA. ACM Press.

Voida, A., Harmon, E., and Al-Ani, B. (2012). Bridging Between Organizations and the Public: Volunteer Coordinators' Uneasy Relationship with Social Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1967–1976, New York, NY, USA. ACM. event-place: Austin, Texas, USA.

Voorberg, W. H., Bekkers, V. J. J. M., and Tummers, L. G. (2015). A Systematic Review of Co-Creation and Co-Production: Embarking on the social innovation journey. *Public Management Review*, 17(9):1333–1357.

Warner, M. (2005). *Publics and Counterpublics*. Zone Books - MIT, New York, NY, new ed edition edition.

Weiser, M. (1991). The computer for the 21st Century. *Scientific American*.

Weiser, M. (1993). Some computer science issues in ubiquitous computing. *Communications of the ACM*, 36(7):75–84.

West, J. and Gallagher, S. (2006). Challenges of open innovation: the paradox of firm investment in open-source software. *R&D Management*, 36(3):319–331.

West, J. and O'Mahony, S. (2005). Contrasting Community Building in Sponsored and Community Founded Open Source Projects. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, pages 196c–196c.

Whong, C. (2015). Thoughts on Open Data Part 1.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):1–9. Number: 1 Publisher: Nature Publishing Group.

Winograd, T., Flores, F., and Flores, F. F. (1986). *Understanding Computers and Cognition: A New Foundation for Design*. Intellect Books. Google-Books-ID: 2sRC8vcDYNEC.

World Wide Web Foundation (2018). Open Data Barometer - Leaders Edition. Technical report, World Wide Web Foundation, Washington DC.

Zaman, J. and Meuter, W. D. (2015). DisCoPar: Distributed components for participatory campaigning. In *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 160–165.

Zanella, A., Bui, N., Castellani, A., Vangelista, L., and Zorzi, M. (2014). Internet of Things for Smart Cities. *IEEE Internet of Things Journal*, 1(1):22–32.

Zubelzu, S. and Álvarez, R. (2015). Urban planning and industry in Spain: A novel methodology for calculating industrial carbon footprints. *Energy Policy*, 83:57–68.