

Understanding and Improving Human Data Relations

Alex Bowyer

Contents

1 Methodology	2
1.1 Forming a Research Paradigm: Ontology & Epistemology	2
1.2 Research Approach: Participatory Action Research & Experience-centred Design	4
1.3 Research Objectives	9
1.3.1 Research Question 1 (RQ1): What is the human experience of personal data, and what do people want from their data?	9
1.3.2 Research Question 2 (RQ2): What role does data play in people's service relationships and how could relationships involving data be improved?	10
1.3.3 Research Question 3 (RQ3): What challenges and opportunities are relevant when attempting to establish these ideals for human data relations?	11
1.4 Overview of Research Contexts and Activities	12
1.4.1 Context One: Civic Data Use and Access to Data in the Early Help Context	13
1.4.2 Context Two: Accessing the Personal Data from your Digital Life using GDPR	15
1.4.3 Context Three: The Practical Pursuit of Better Human Data Relations	17
1.5 Methodologies Employed in Case Studies	19
1.5.1 Sensitisation of Researcher and Participants	20
1.5.2 Discussion and Exploration with Stimuli	24
1.5.3 Participatory Co-Design of Possible Solutions	26
1.5.4 Practical Data Experiments, Interface Design and Prototyping	29
1.5.5 Analysis, Modelling and Learning	32
1.5.6 Recruitment	34
1.5.7 Ethics	39
1.5.8 Thesis structure approach	40
Bibliography	40

1 Methodology

“Research is defining the invisible.” — Steven Magee (author and researcher)

In the previous chapter, I described three research areas this thesis seeks to explore: how people think about data and what they want from it, how data fits into people’s relationships with organisations and how they want it to be used, and how could people’s desires for the role data plays in their lives be brought closer to reality. In this chapter I will explain my approach to conducting research in this area, detail the types of methods used, and explain how the different research activities I carried out contribute to those three research aims.

1.1 Forming a Research Paradigm: Ontology & Epistemology

[TODO consolidate this, remove objectivism, add more about human centric focus] To develop a research paradigm it is important to begin with reflecting upon your outlook on the nature of reality (ontology) and your beliefs on how knowledge of that reality is formed (epistemology) (Guba, 1990). It will already be evident from the literature review and the framing of this thesis so far that individual human perspectives are at the centre of my research questions. This is a reflection of my ontological stance which is that everyone experiences their own reality, informed by their own concepts and mental models of the world. This is known as *constructivism* (Guba, 1990), where new knowledge is formed by developing one’s own mental models in order to explain new experiences, as distinct from the positivist view that there is a single universal reality that needs to be uncovered. However, in parallel to this individual learning through experience, people’s realities are constantly shifting and changing, especially when it comes to the rapidly changing technological landscape we live in today reality – consider that today our reality now includes concepts that did not exist in our youth, from ‘feeds’ and ‘posts’ to ‘link sharing’, ‘syncing’ and ‘blocking’. As new technologies and practices emerge, we develop new mental models to help us make sense of and find value in new capabilities. This idea of reality as something constantly renegotiated by the individual is known as *pragmatism* (Campbell, 2011). To me this is an overriding truth about reality and this focus on understanding change, as perceived by individuals, is a key research motivation. Where constructivists may focus more upon deeply understanding an individual’s reality at a moment in time, I am more interested in understanding the ways in which people’s understanding of the world, and of themselves, changes as a result of their lived experience. At this point we must consider the individual’s motivation for constructing and pragmatically changing their concepts of the world, and to understand this we can look to *objectivism* (Peikoff, 1993), the philosophy put forward by Ayn Rand, which is a belief that the mind, informed by the senses, is the means by which we discover truths about the world, and it does so by forming concepts and using *inductive reasoning* (Smith, 2011) (in

essence, “if these things are true then what else must be true?”) to acquire knowledge. In essence, people’s conceptions of reality are constantly tested and re-evaluated by their experiences of the world. Objectivism also states that individual’s motivation in life is the pursuit of one’s own happiness and wellbeing, and that this self-interest is what drives his pursuit of deeper knowledge and understanding about the world; in essence, everyone wants to improve their own life, and they need knowledge to do it, and for me this view of understanding the nature of reality, so that one might be able to change it for the better is also a key driver behind my research. As a final philosophical element to incorporate, I also look to *Deweyan pragmatism*, which states that our knowledge and thinking are tested by actions, not just reason, and that this is how we learn - and that communication and interaction with others is a key part of that learning. Dewey recognises that every individual is not solitary, he exists within a society; he “is a social being, a citizen, growing and thinking in a vast complex of interactions and relationships.” (Dewey and Archambault, 1964) People create systems and meanings through those interpersonal interactions – which they can then use to understand everyday life; this is particularly important in the social world, as unlike the physical, natural world, many concepts are abstract and subject to individual interpretation.

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

Based upon this, we can now look to epistemology: how can knowledge be acquired? Having a constructivist rather than a positivist stance means that this is best done not through direct observation of the world and empirical testing of hypothesis, but though interacting and communicating with individuals so that we can interpret how they view reality; this is known as an *interpretivist* epistemology. Most of the techniques used will therefore be *qualitative* (understanding perspectives and collecting non-numerical data) rather than *quantitative* (measuring behaviours and collecting numerical data). The focus of my research is to acquire understanding of people’s views and mental models around data and digital living, so that I can further these concepts in order to develop theories - powerful explanations that can be understood and benefitted from by ordinary people - to fill the knowledge gaps in existing research that I have identified. Given my strong focus on pragmatism and interpreting people’s constructed social realities in terms of practical usefulness to them, I will not be deeply analysing their words through language analysis techniques like discourse analysis, but will instead focus on the social, interpersonal level - understanding how people navigate the world of data and data-based relationships and change their understandings as they seek to achieve their goals in practice; and how they are affected by the systems, relationships and society they exist within. It is this practical focus, recognising that within a society there are objective truths that will affect all individuals that means the methods used will not be

solely qualitative, but rather a *mixed methods approach* where I will adopt the most appropriate methods, usually qualitative but sometimes quantitative, as appropriate to the particular research context and question being explored.

1.2 Research Approach: Participatory Action Research & Experience-centred Design

[TODO shrink this section]

As we move away from general research approach to the specifics of this study, it is important to be clear about what it seeks to achieve. The purpose of the research is to formulate theories that can facilitate change - to map out a research and development agenda that might help the world to move from a data-centric (see section 2.1) to being human-centric (see section 2.3) operating paradigm. By learning about people's understandings of their reality, this will inform my own thinking, and using by an inductive research approach we can identify patterns common to multiple people and form theories that might explain these patterns. As a student of *Digital Civics* (Vlachokyriakos *et al.*, 2016) I believe that research can surface the ways in which current service provisions fail to meet people's needs, and through research we can show how the world might better empower citizens if it were configured differently with services closer to what they desire. The role of the researcher is to understand the world and to figure out how to change it. It is an accepted view that research cannot be value-free, but in fact we can go further, the researcher can be an activist, seeking to correct an imbalance in the world through their research. As such, the design elements of this research can be considered as political, this is *adversarial design* (DiSalvo, 2012) and I view this as necessary to counterbalance the strong forces outlined in Chapter 2 that are acting against individual interests; by creating space to reveal and confront power relations and influence, we can identify new trajectories for action (DiSalvo, 2010). Therefore the purpose of the research is to inform myself as adversarial designer, with the acquired insights from the experiences of research participants helping me to develop my own understanding, models and designs.

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

However such perspectives may ultimately be ignored or diluted in favour of expert designs or organisational motives.

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

As design approaches, I see merit in both UCD and PD. The participant should play a role as an informant - one who can provide critical insights into their own perspective on a design space and help us understand how the world is to them - but also as a designer - one who can imagine how they would like the world to be. As we involve the participant, our role as the researcher is to elicit the richest possible responses from the participant, by using questions to bring them to consider new questions and by giving them stimulating materials to trigger their thinking. The researcher also often needs to *sensitise* the participant to a design space, so that they may properly engage with the questions being posed, but equally the researcher cannot arrive at a model or theory unless he has developed *empathy* for the participant’s perspective. One of pragmatism’s founding philosophers, Peirce, put forward the *pragmatic maxim*, which states that the meaning of anything we experience in the world is understood through the conception of its practical effect, and that theories that are more successful at controlling and predicting our world can be considered closer to the truth (Campbell, 2011). Applying this philosophy in to the challenge of design, I find merit in the different, less political, take on involving users as participants in design exhibited in McCarthy and Wright’s *experience-centred design* (McCarthy and Wright, 2004) framework, which emphasises the importance of understanding the user’s experience to inform technology design. It identifies six sensemaking processes users go through. These can be considered to help acquire user empathy:

1. *anticipating*: We never come to technology unprejudiced.
2. *connecting*: We make a judgement in an instant, without much thought.
3. *interpreting*: We work out what’s going on and how we feel about it.
4. *reflecting*: We examine and evaluate what is happening in an interaction.

5. *appropriating*: We work out how a new experience fits with other experiences we have had and with our sense of self.
6. *recounting*: We enjoy storytelling and make sense of experience in stories.

Through my research I will at times be more participatory, to understand these aspects of user experience or to co-design solutions with participants, but I will at other times act more like an expert designer. Taken to the extreme, the PD view is that designs made without the direct involvement of users are invalid, because they inherently no longer represent the desires of those people the designs claim to serve. I oppose this view, because I believe that new ideas will not always arise from participants themselves, especially for this research area where a more expert-led experience-centred design approach is the most pragmatic way to proceed, because by its nature this research involves thinking about data, information, organisational relations and interaction (topics that are not often theorised about as part of everyday life) at a level which the layman is not accustomed or well-equipped to do; therefore while I strive to always include participant viewpoints, I give ultimate precedence in design to my own position of learning that I will acquire through the research I undertake with participants and which I will develop through theoretical & design work that I will undertake by myself. In doing so, I will also be a participant in my own research, incorporating my own experiences of living in a data-centric world (and my attempts to challenge it) into my learnings.

It is important to be clear about what constitutes good research in this context; if the outcome of the research is to be my own interpretations and theories, how will we know these are sound? Firstly it is important to say that this is not about measuring the effectiveness of proposed changes upon the world. There will be no deployment of systems to test the ideas I put forward. This is not because such an activity would not be worthwhile—it would—but simply because by its nature, to develop, build and deploy new data interaction paradigms that function in real life with real personal data at the sociotechnical level would be too large an endeavour for a single researcher (or even a single research group) to undertake. Therefore what I seek in this thesis is not to change the world, but to articulate with the greatest possible clarity discrete theories on how the world should, and could, be changed. Good evidence for the proposed changes will be achieved by ensuring that findings themes and discussion contributions are backed up by participant quotes, and where an idea is suggested or agreed upon by many participants or where it resonates with my own embedded experience, that can be seen as adding weight or validation to that idea. However, each person's experience is unique and needs to be put into context; not every insight will be shared by many participants and individual unique insights remain important.

The mixed methods approach I will be adopting closely follows the discipline of *participatory action research (PAR)*, which is an approach to research that encompasses both the involvement of participants' perspectives while also retaining a role for the reflection and learning of the researcher themselves. PAR's creator Kurt Lewin observed that “there is nothing so practical as a good theory”

(Lewin, 1951) which shows the pragmatic nature of this approach. PAR combines self-experimentation, fact-finding, reasoning and learning, and makes sense of the world through collaborative efforts to transform the world rather than just observing and studying it (Chevalier and Buckles, 2008). Central to this is the idea that research and action must be done with, not on or for, people; participants are not subjects but co-researchers, evolving and addressing questions together (Reason and Bradbury, 2001). To embody the three ingredients of PAR (Chevalier and Buckles, 2019) – participation, action, and research – my research will include three types of activity:

1. *participatory co-design activities* – where I will discuss and explore experiences, challenges and possible solutions with participants through conversations and design activities
2. *self-experimentation activities* – where I will carry out experiments, ranging from thought exercises to practical tests of what is possible, to develop ideas and explore the problem space myself, and
3. *embedded research activities* – where I will participate as an involved team member, in external organisations' projects that are trying to change the world in this space, so that I may learn about the challenges faced on the basis of the grounded experience of myself and others (Cheetham *et al.*, 2018).

[TODO diminish or place number 3 outside of PhD]

Action research also carries with it the idea that research is done in cycles: you learn something, carry out some action in the world based on your learning, learn from what happened, and repeat. This has become an established approach in HCI research (Hayes, 2011) and the importance of collecting stakeholder feedback at regular intervals is also seen in the software industry though *agile development* (Fowler and Highsmith, 2001) which can be seen as a practical implementation of action research. In startups, terms like ‘fail fast’ (Brown, 2015) and ‘pivot’ (Ries, 2011) illustrate the idea that it’s crucial to test ideas on real people then adapt quickly based on how that goes. To me, action research does not mean that you must test every single idea with an audience for it to be considered valid, but rather that user engagement is not a one-off, but a repeated component that affects the research path. Each new research activity will draw from your past learnings and theories and your acquired understanding so far, which will be further developed through its exposure to ‘real life’ in the process of participatory and embedded research activities.

Figure 3 shows the cycle of action research, as I will apply it in this study. In each area of life or context that I identify as a setting for a research activity, I will first carry out initial background reading, experimentation or exploration to familiarise myself with the area, then I will design a research activity that helps to explore my research question in that area. After carrying out the planned activity (be it participatory, self-experimentation or embedded research) I will analyse any data from that activity (or just reflect upon my experience), and then use these findings to update my overall understanding of the answer to

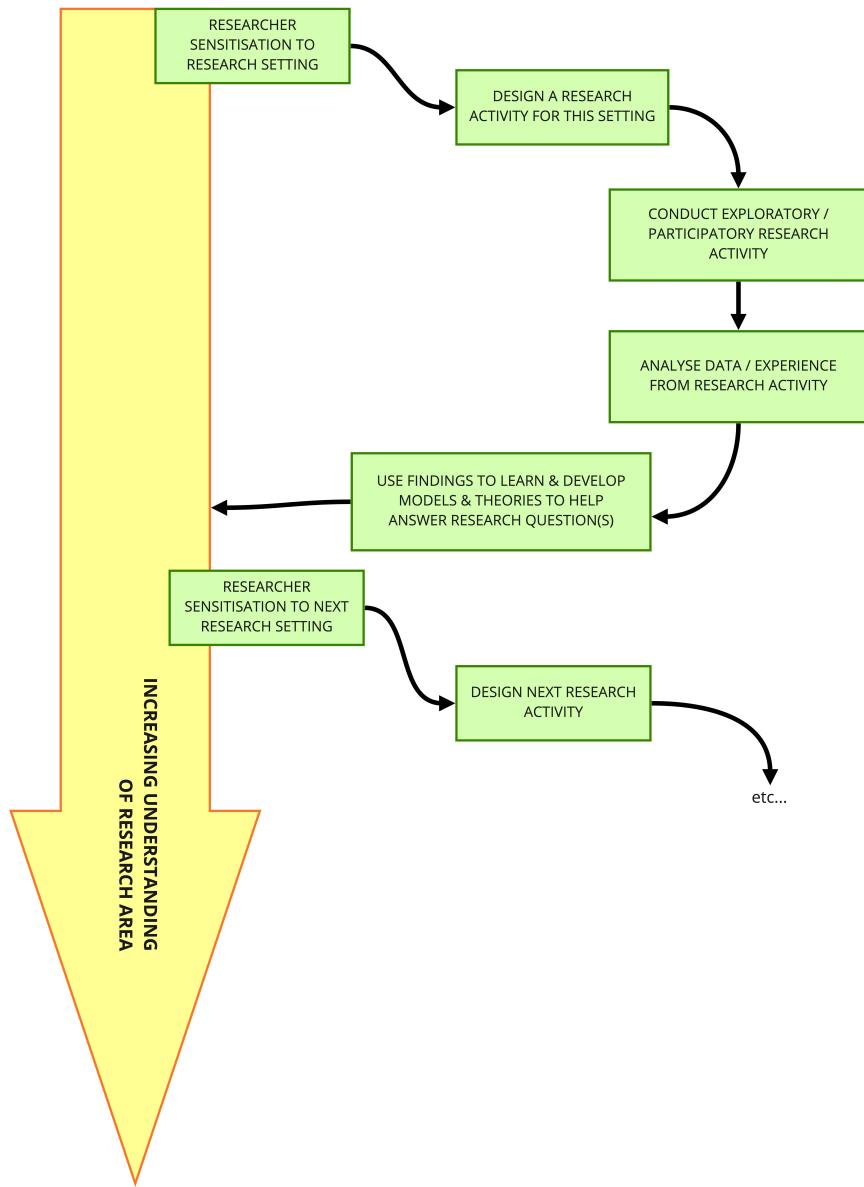


Figure 1: Figure 3: My action research approach

my research questions. I will then go on to repeat this cycle, with the next study, but beginning with more developed theories or understandings than the previous. In the case of embedded research activities these are likely to go on for several months alongside other activities, so analysis and learning will happen throughout, resulting in a continually updating current understanding that will form the baseline for later research activities. In the next section I will describe the three specific research objectives that will be targeted through the research activities.

1.3 Research Objectives

At the end of chapter 2, I introduced my research question, which is:

“What relationship do people need with their personal data, and how could that be achieved?”

[TODO restructure this to deformatise RQ3] Corresponding to the three research gaps I am focusing on as identified in 2.1.5, 2.2.5 and 2.3.5 respectively, there are three distinct subquestions I will explore using the approach detailed above. Each of my research activities will be designed to advance my understanding and theories towards at least one, sometimes more than one, of these three research objectives:

1.3.1 Research Question 1 (RQ1): What is the human experience of personal data, and what do people want from their data?

As established in section 2.1, personal data, and its collection and use by commercial and civic organisations, is an established and inevitable part of modern life, yet the concept of data is abstract and poorly understood. The first strand of research I will be advancing through this thesis is to establish a solid understanding of what mental models people have constructed about data. We need to understand what makes data meaningful to people, and given HDI’s belief that everyone needs a relationship with their data, we need to understand what relationship people currently have with their data. What *is* data to people? If we are to design new human data relations, we must begin by understanding people’s current relationship to their data, the ways in which that relationship affects them, and their unmet desires for improving their relationship to their personal data. We need to find out what aspects of data cause positive emotions, what problems do people experience with their data, and what people want from their data.

In order to approach this objective, we must take a participatory approach; gathering individual perspectives on data, and looking for patterns or trends in those perspectives, will be the primary means to advance this research objective. The first challenge here will be to find ways to sensitise participants to be able to conduct an informed and productive conversation about the topic of data, which to the layman may seem a dry, boring topic. This challenge will be addressed by leading participants into the subject of data using meaningful representations

of data as stimulus for conversation, or starting with the individual's own life experience to discover the data in their life, which they are more likely to have opinions and emotions about, rather than talking about the subject in the abstract.

1.3.2 Research Question 2 (RQ2): What role does data play in people's service relationships and how could relationships involving data be improved?

In section 2.2 and 2.3, I established that as of yet, designers of PIM and personal data interfaces have not yet risen to the sociotechnical challenge of looking at the reality of personal data today: that it is scattered, inaccessible and largely unusable. There is no way for people to view their data holistically, nor any tools to help people manage the many relationships that individuals have with companies, employers, councils, governments and other organisations that rely heavily upon the collection and processing of their personal data. Almost every civic or commercial service we use today handles our data. We know that the world is data-centric, and that data controllers use data as an asset to inform their decision-making, creating a serious imbalance of power (Hoffman, 2010, 2011, 2013, 2014a, 2014b). But what is like to conduct a relationship with an organisation that holds your data? What emotions do people experience? How does it affect their daily life, and what sort of problems do people face as a result of this data-centricity? If your data is used in ways you do not understand or consent to, how does this affect your outlook on the world? This is the second strand of research I will be exploring: to gain an understanding of the data world beyond the individual, so that we can design not just better individual relationships to one's data, but improve people's relationships with organisations that hold and use data. (Note: for the purposes of this study, we only pay attention to service relationships, not social or interpersonal relationships). In this thesis and its title I use the term "human data relations" to encompass both of these aspects - human-data relations (the individual's relationship to their data, as imagined by HDI), but also human data relations, i.e. human relationships that involve data.

To tackle RQ2, participatory research approaches are appropriate here, as our questions relate to the individual mental constructs that people have about their wider digital lives and relationships. But there is another aspect here, and that is that a relationship involves two parties. Consistent with Dewey's belief in the importance of interaction in creating meaning, the structuralist philosopher Michel Foucault said that "meaning comes from discourse" (Adams, 2017), in other words people do not construct their reality in isolation, but in fact it is shaped by the social constructs and systems they operate within. Deweyan pragmatism also takes the view that research must seek solutions to real world problems that are generalisable to use in society at large (Dewey and Archambault, 1964; Friedman, 2006). This implies that any such solutions arising from my research must work for all parties. For both these reasons, I

will conduct participatory research to understand both perspectives: that of the data controller and that of the data subject, and where possible I will engage both parties together in discourse so that the two parties' worldviews can be brought together to design solutions that could work in practice for all involved.

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

1.3.3 Research Question 3 (RQ3): What challenges and opportunities are relevant when attempting to establish these ideals for human data relations?

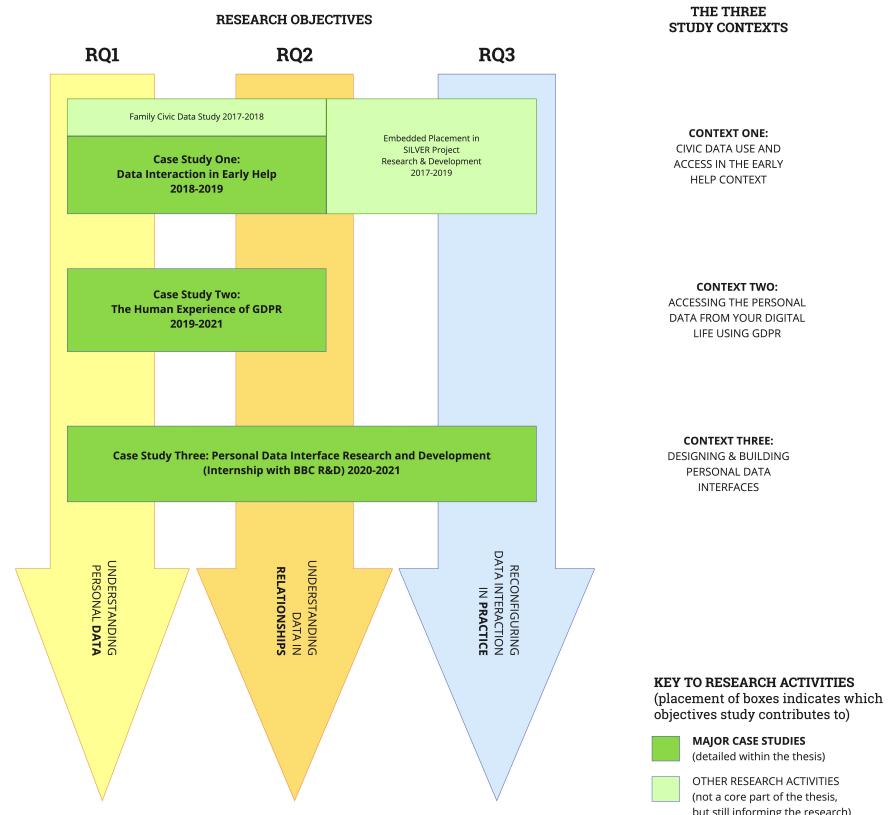
[TODO remove/deformalise this - instead talk about how this will be explored in the discussion, looking at external embedded research and the work of others]

As a software industry professional, and as a pragmatic digital civics researcher, I believe it is important that the outcome of my research is not purely theoretical. While the goal of this PhD is not to build a new data interaction system, it is important that we pay attention to *how* the problems outlined in section 2 might be addressed, and how the individual desires and needs we uncover in RQ1 and RQ2, might be met in practice. This involves gaining understanding of the technical, economic, political and legal landscape that personal data interaction occurs within; gaining clarity on the motivations that service organisations have for being data-centric; and understanding the current systems and organisational practices that influence current system and process designs. Just as Li showed that users of SI systems experience a barriers cascade as they try and achieve more human-centric data goals (Li, Forlizzi and Dey, 2010), it follows that there are also likely to be a series of obstacles that service organisations would have to overcome if they were to approach these goals. We need to uncover these obstacles so that we can design approaches to overcome them. The third strand of my research is to outline practical steps and guidance, both for researchers and personal data interaction system developers, as well as policymakers and civic activists to make it clearer how they can pursue the goals we identify for improved human data relations.

This strand will be addressed in parallel to RQ1 and RQ2, so that practical discoveries may inform those research questions too. This also means that as new needs and desires emerge from RQ1 and RQ2, they can become "requirements" for the more technical design work of RQ3 [TODO deformalise RQ3]. As an approach, this will be action research in its purest sense - I will embed myself in projects working in the personal data space, as a developer and a researcher, so that I can gain deep field experience of the constraints and opportunities that affect the design of data interaction systems and processes while also feeding research learnings back into those projects in pursuit of practical impact. Unlike

RQ1 and RQ2, this strand of research will be explored not through strictly configured PhD case studies, but rather through a process of acculturation to the world of building data systems and practical project operations in the personal data economy. This will developing my knowledge through design, technical prototyping and pushing the boundaries of the systems that do exist so that they may be better understood. The objective is to gain insights that can afford me greater expertise, backed by the empirical findings from RQ1 & RQ2, to allow me to draw with confidence conclusions about how the discipline of human-centred data relations should proceed in its future research and development to best serve individual and societal interests.

1.4 Overview of Research Contexts and Activities



[TODO: UPDATE Figure 4 to diminish BBC and add Hestia and remove RQ3/Case Study Three.]

As explained in the last section, the three sub-research questions RQ1, RQ2 and RQ3 [TODO deformatise RQ3] have been addressed in parallel throughout this research. They can be considered as three parallel trajectories of research and learning, each informed by some or all of my research activities as they progress,

in cycles of action research as described in section 3.2 above. Figure 4 shows these three parallel research objectives as downward arrows. Considered as three areas of understanding, RQ1 can be seen as understanding personal *data*, RQ2 as understanding data in *relationships*, [TODO de-formalize RQ3] and RQ3 as understanding how to reconfigure data interaction in *practice*. Figure 4 also illustrates how the three contexts of study, addressed through two major case studies and a number of other activities, which I will explain below, contribute to advancing my understanding of each area - with the positioning of the box over an arrow indicating that it contributes to that area of understanding.

1.4.1 Context One: Civic Data Use and Access to Data in the Early Help Context

The first research context I explored in this PhD was “Early Help”. This is explained in detail in Chapter 4, but in brief: Early Help is a particular type of social support offered by UK local authorities as voluntary help to families who are considered to be at risk of falling into poverty, crime, truancy, addiction or other issues which are both problematic for the individuals and costly to the state. Families enrolled in the scheme meet a social worker (called a ‘support worker’ in this context) regularly who can provide advice and connect the family with appropriate health, lifestyle and social services to their needs. As part of this, the support worker has access to a variety of data from civic sources: school records, employment and benefits data, social housing data, criminal records, and more, so that they might be better informed about the family’s situation. However the families do not have any access to this data, and thus despite this being a scheme that is on the face of it intended to empower families to help themselves, it runs the risk of disempowering the families through the same data-centric power imbalance described in section 2.1.2. Therefore, this setting provides a very interesting context in which to examine both RQ1 (finding out how these supported families feel about their data) and RQ2 (examining the impacts of data use within a service relationship) as well as to explore how the families and support workers could imagine their data relations being improved.

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

1.4.1.1 Embedded Research Placement in CHC SILVER Project

From March 2017 to March 2019, I joined Connected Health Cities’ “SILVER” project (Connected Health Cities, 2017) as a part-time research engineer alongside my PhD. This research project was funded by the UK’s Department for Health (now the Department of Health and Social Care) and brought together local authorities, health authorities, University researchers and technology partners in the North East of England, in exactly the Early Help context described above. Its goal was to explore how to unify civic data about a supported family, with their consent, to allow support workers to provide better care to those families. This made it an ideal place to explore my research objectives: Because it was aiming

to build a real-world technical solution, this would provide practical insights that would serve RQ3 [TODO remove], and as it was also using direct research with families and support workers to inform the system requirements, this would also provide an opportunity for deeper understanding of the use of data within the Early Help support relationship (RQ2), and both parties attitudes to this highly personal and real civic data (RQ1). My role was two-fold: as a software engineer, to design and develop user interfaces that would be used to view this unified data, and as a participatory researcher, to assist with the design and execution of focus groups and workshops with staff and supported families that could inform the proof-of-concept data system being built. This embedded placement is not considered a major case study of this thesis, however it has contributed to the research objectives and the developing understandings of this context so will be referenced in the subsequent chapters, especially Chapter 4 and Chapter 7. Chapter 7 incorporates observations from participating in the project. The final report from the project was not published but a public summary is available (Northern Health Science Alliance, 2020).

1.4.1.2 Pilot: Understanding Family Civic Data study [TODO make sure this is properly positioned as prior work - give the 2018 paper a ‘special status’ and put it in Appendix]

In the summer of 2017, in the MRes year of this doctoral training programme, I carried out an initial participatory field study in order to deepen my understanding of data use and attitudes within this context (RQ1) and develop appropriate research methods. This study consisted of home visits to four different families in the North East who had interacted in the past with social care & support services. During the course of these two hour visits I carried out participatory co-design activities and interviewed the families (both adults and children) about their civic data, and in particular their views on how risky different types of data were and how that data should be handled. While this fieldwork took place prior to the start of this PhD, the data analysis and publication of the findings took place within the scope of this PhD. Again, this is not considered a primary study for this PhD, but will be referenced within this thesis. The paper which published the study is (Bowyer *et al.*, 2018), which is included in [ADD APPENDIX REFERENCE TO CHI2018 PAPER HERE].

1.4.1.3 CASE STUDY ONE: Data Interaction in Early Help study

In the summer of 2018, informed by the SILVER project and the Understanding Family Civic Data study, I designed and conducted my first major case study of this thesis: a series of three participatory co-design workshops with people directly involved in Early Help relationships in North East England. The workshops were funded by CHC and conducted by myself and were designed with a dual purpose: to inform the design of the SILVER system but also to serve RQ1 and RQ2 of this thesis. These workshops built upon the Understanding Family Civic Data study, in order to validate the earlier findings – but aimed to develop a deeper understanding of what supported families (workshop 1) and support workers

(workshop 2) perceive as problems with data use in the Early Help context and to explore perceived solutions to these problems. The third workshop was specifically designed to focus on the use of data within the support relationship, and was a joint workshop involving staff and parents working together. This case study is described in detail as Chapter 4, and contributes to the general findings about RQ1 and RQ2 presented in Chapter 6.

1.4.2 Context Two: Accessing the Personal Data from your Digital Life using GDPR

From the start, a core motivation for my interest in this research has been to look at the power imbalance around personal data from the “everyday life” perspective - to explore our relationship with and through the data that we hold, use or live with as we go about our lives, online and in person. It seems that this power imbalance is something that touches everyone, and therefore for my second research context I chose not to focus on a particular community or group but to look at these problems at the level of our day-to-day digital lives. I designed research activities where I would talk to people about their everyday experiences of data in their lives (RQ1) and their views on the usage of data within their relationships with commercial or civic service providers (RQ2). In 2018, during this PhD, the European Union’s GDPR regulations came into force, enabling people to obtain copies of their own data. This enabled me to take the research deeper than a simple conversation and to guide my participants through the GDPR process to obtain their data from providers, and then to use this retrieved data as a stimulus for discussion; this I hoped would result in a far more grounded and less theoretical perspective. In parallel to this, I began to conduct my own experiments using GDPR to see and explore my own data. This allowed me to sensitise myself to the research space, and to enhance my understanding of RQ3 (finding out more about what is and is not possible in practice when it comes to everyday personal data access) [TODO deformalise RQ3] but also crucially it enabled me to become a participant in my own research, enabling a deeper understanding of this research context.

[TODO unify first three into “Pilot studies” and make clear which were outside the PhD] Within this context, I carried out four research activities between 2016 and 2020:

1.4.2.1 Smartphone Usefulness study This early study was carried out in late 2016. Its goal was to deepen my understanding of people’s perceived values around everyday technology use and to validate some of my own perspectives. Using participatory interviewing techniques I explored attitudes to smartphone use, with particular attention to perceived usefulness or barriers. This was designed to provide background on what motivates people as users of technology, an important consideration when looking at disempowerment. The thematic findings from this study are detailed in a report in [INSERT APPENDIX REFERENCE HERE].

1.4.2.2 Digital Life Mapping study In order to further acclimatise myself to people's attitudes to data and to provide balance to my own attitudes and opinions, I conducted 5 two-hour interviews with individuals about their digital lives, looking at how they mentally segment their life, and the roles and functions of different technologies, and especially of data, across those different parts of their lives. As part of this I also explored the participants' perceptions of their relationships with service providers, in order to identify the ways in which individuals might feel disempowered by the ways their data was handled or to identify what they would like to change about their data relationships. The interviews were conducted using the Sketching Dialogue (Hwang, 2021) technique, which uses collaborative sketches as a basis for a semi-structured interview. A light summary of observations and findings are presented in [INSERT APPENDIX REFERENCE HERE].

1.4.2.3 Self GDPR Experiments As preparation for Case Study Two, and in order to increase my own empathy and participation in the research, I have throughout the last three years from 2018 made numerous efforts to obtain my own data from companies and organisations in my own life. This has entailed over 70 GDPR requests to a variety of organisations including retailers, device manufacturers, online service providers, local and health authorities, banks and leisure services. Additionally I have experimented with self-service download dashboards and third party 'get my data' tools. In some cases I have engaged providers in communication to try and get better data or ask questions about my data. These activities have provided multiple benefits: they have enabled me develop a detailed understanding of what actual stored personal data looks like (which informs RQ1), they have given me an awareness of the evolving response to GDPR from data controlling organisations (which informs RQ2), and has allowed me to test the limits of what is and is not possible with GDPR (which informs RQ3) [TODO deformalise RQ3]. A summary of observations and findings are presented in [INSERT APPENDIX REFERENCE HERE - OR DON'T!].

1.4.2.4 CASE STUDY TWO: The Human Experience of GDPR As described above, the major study for this context was to guide participants through the process of GDPR and retrieving their own personal data, to enable a conversation that included not only attitudes to personal data and the use of data within service relationships, but discussion of how those attitudes were changed by the experience as it happened and how well expectations and hopes were met by the process. 11 participants were engaged 1-on-1 in a 4 to 5 hour process over a series of months which involved five stages:

1. Sensitisation, using a set of wall posters about data holding organisations, types of personal data, GDPR rights and possible uses for your retrieved data
2. A life mapping exercise, similar to that in 3.4.2.2, using the Sketching Dialogue (Hwang, 2021) technique, at the end of which 3-5 target companies

were selected for GDPR.

3. A discussion and guided walkthrough of the target organisations' privacy policies, in particular their stated data collection practices.
4. Guidance and support in making and seeing to conclusion a GDPR request from each individual to each of their target organisations
5. A 2-hour interview in which participants were guided through the reviewing their data and were asked about their experiences and reactions to the data and the GDPR process.

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

In the final data exploration interviews, which were conducted online over Zoom due to COVID-19 restrictions, a spreadsheet-based approach was used, where participants were walked through a series of Yes/No questions about different categories of their data, and then asked to expand verbally on their reasoning. This produced both qualitative and quantitative data for later analysis.

This case study is described in detail as Chapter 5, and contributes to the general findings about RQ1 and RQ2 presented in Chapter 6.

1.4.3 Context Three: The Practical Pursuit of Better Human Data Relations

[TODO move / reposition this to be external to the thesis, maybe omit entirely]

The third context for this PhD, which has remained a parallel focus throughout, is a more practical one; to go beyond just understanding people's perspectives but to look, in the context of what we learn about people's desires for their data and their relationships, at what is currently possible in practice. The goal is to find out what factors shape the design and implementation of real world data interaction systems and processes, to understand what legal, social, economic, technical or political factors come into play and importantly, to explore what technologies or techniques might be able to pursue human-centric design goals in a data-centric world. In scope, this context is a broad one, encompassing all forms of personal data interaction; as such it is able to draw on the findings of RQ1 and RQ2 from the first two contexts, viewing those as "needs" or "requirements" that would ideally be met through the design and pursuit of new interfaces, policies and practices.

In total four separate research activities beginning in 2017 and continuing through to 2022 and beyond have taken place within this practical research context, outside of the Case Studies but very much drawing upon and feeding insights back into the developing findings, and thus part of the action research

cycle of this thesis (3.2).

These activities relate to this thesis not as case studies but as exemplars to illustrate possible avenues for bringing about better Human Data Relations in practice (RQ3) [TODO deformatilse RQ3]. Chapter 7 examines the practicalities of bringing about the desired changes.

1.4.3.1 Health Data Interface Development: CHC's SILVER project

The embedded role I took in the SILVER project described in section 3.4.1.1 contributes also to this context, as part of my role was as a front-end software developer for a personal data health interface intended for use by support workers in the Early Help context. Learnings from that experience also helped to serve RQ3 [TODO deformatilse / remove RQ3]. This aspect of the SILVER project is considered out of scope for this thesis, though reference is made to it in Chapter 7.

1.4.3.2 Obtaining Data and Reconfiguring Data Interfaces through Web Augmentation [TODO update this section with new paper, possibly shorten a bit too]

As a software developer I have been aware for a long time that one of the biggest challenges in building new data interfaces is to gain programmatic access to the necessary data. As part of the trend towards cloud-based services and data-centric business practices, it has become increasingly difficult to access all of the data held about users by service providers. Application Programming Interfaces (APIs) are a technical means for programmers to access a user's data so that third party applications may be built using that data. Unfortunately, as a result of commercial incentives to lock users in and keep data trapped (Abiteboul, André and Kaplan, 2015; Bowyer, 2018), much of users' data can no longer be accessed via APIs. While GDPR data portability requests do open up a new option for the use of one's provider-collected data in third party applications, this is an awkward and time-consuming route for both users and developers. *Web augmentation* provides a third possible technical avenue for obtaining data from online service providers. It relies on the fact that a user's data is loaded to the user's local machine and displayed within their web browser everytime a website is used, and therefore it is possible to extract that data from the browser using a browser extension. Similarly, once loaded into the browser, a provider's webpage can be modified to display additional data or useful human-centric functionality that the provider failed to provide.

In order to better understand what is and is not possible using this technique, I participated from 2018 to 2020 as a part time web developer in a project which was using the web augmentation technique to improve the information given to users of Just Eat, a takeaway food ordering platform in the UK. While this particular use case does not concern personal data, the technology being used by the project were considered highly relevant, and the goals of the research project were also human-centric, and consistent with our own research goals - tackling

power imbalance of service providers in order to better serve individual needs. This research project is not detailed within this thesis, and is not considered a primary study for this PhD, but is referenced within Chapter 7. The paper which published the study is [ADD REF goffe *ET AL*], which is included in [ADD APPENDIX REFERENCE TO GOFFE ET AL PAPER HERE].

1.4.3.3 Personal Data Store Research: The BBC R&D Cornmarket Project [TODO: Diminish this so it is no longer a case study just a contributing piece of work]

Within the personal data interface design context, I undertook my second embedded research activity within the PhD. For an eight month period (three months full time and five months part time) beginning in early summer of 2020, I was a research intern in the British Broadcasting Corporation's Research and Development department. The BBC has a public remit to carry out research and development in the broadcast, media and information space, including HDI (BBC R&D, 2017), and has over 200 researchers. I was assigned to a project codenamed Cornmarket, a collaboration between user experience designers, researchers and developers which aimed to explore a new role for the BBC in extending its public service role beyond broadcasting into personal data stewardship. The main task was to develop a prototype personal data locker into which people could store everyday data including TV and music media streaming data, health data, and financial data. This provided an excellent opportunity to put all of my learnings acquired thus far for all three RQs into practice, and further deepen my understanding of RQ3 - the barriers and opportunities to actually building new human-centric data interfaces in the real world. [TODO deformalise RQ3] Throughout the internship I was able to explore the problem space from many different angles - sharing my own research expertise, doing competitor analysis and background research, information architecture, data modelling, user experience and user-centred design, technology prototyping and supporting participatory research activities. This embedded research provided numerous new insights and an opportunity to iterate and develop my theories and models with BBC colleagues.

1.4.3.4 Understanding the Power of Data: Sitra's #digipower Project with Hestia.ai [TODO add details here]

1.5 Methodologies Employed in Case Studies

In the previous sections I introduced my research approaches and the three research contexts and the different case studies and research activities I carried out. In this section I will explain which methods were used across those studies and why they were chosen.

The methods used in my research can be loosely grouped into five stages, though not every activity involved all stages:

1. Sensitisation of Researcher and Participants
2. Discussion and Exploration with Stimuli
3. Participatory Co-Design of Possible Solutions
4. Practical Data Experiments, Interface Design and Prototyping
5. Analysis, Modelling and Learning

[TODO: consider if Hestia methods need to come into here at all]

I will now explain each of these stages, with examples from the different studies, as well as providing information about recruitment, ethics and thesis structuring at the end of this section.

1.5.1 Sensitisation of Researcher and Participants

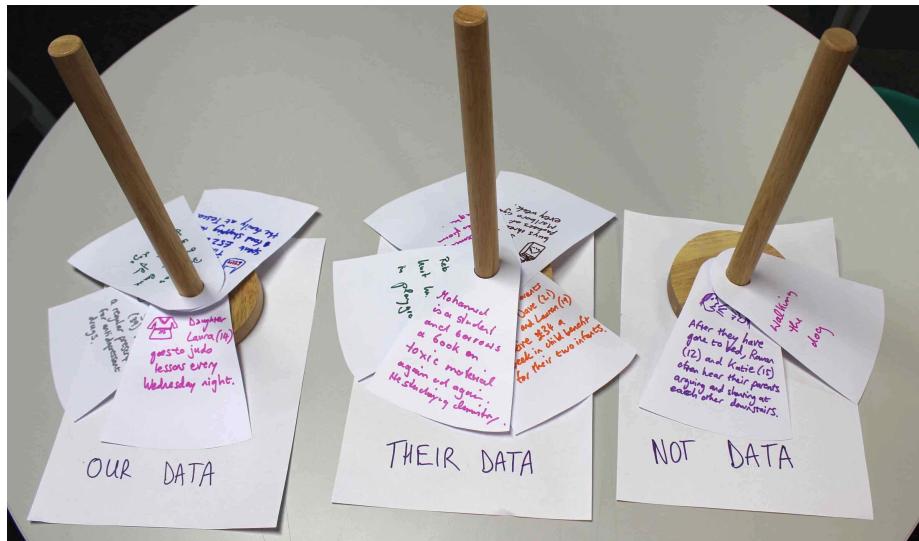


Figure 2: Figure 5: “Family Facts” – What is Data?

As I described in section 3.2, an important first step before any research activity is to *sensitise* myself as researcher to the research context, which means to become familiar with relevant issues, systems and practices and increase one's empathy for the participants. In the Understanding Family Civic Data study, this entailed a review of grey literature to identify the different types of civic data that councils stored, and conversations with colleagues and partner organisations within the SILVER project to deepen my understanding of Early Help. This same study served as researcher sensitisation for Case Study One, as through that study which introduced me to families that had had some contact with the care system, I was able to gain empathy for supported families and acquire some initial understandings of likely perspectives, before working with supported families directly; and through participation in fieldwork with support workers through the SILVER project I was able to gain empathy for the data needs

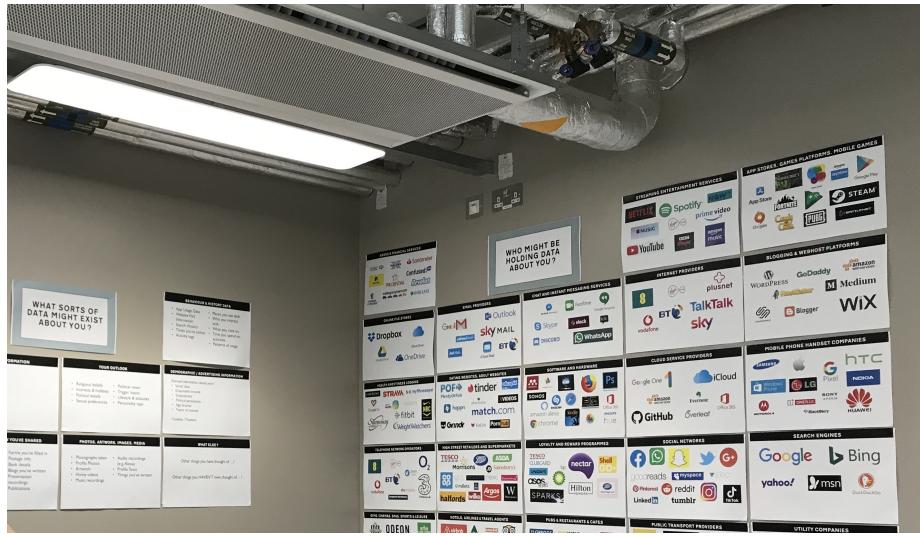


Figure 3: Figure 6: Walls of Data – Sensitising participants to the world of commercially-held data and GDPR

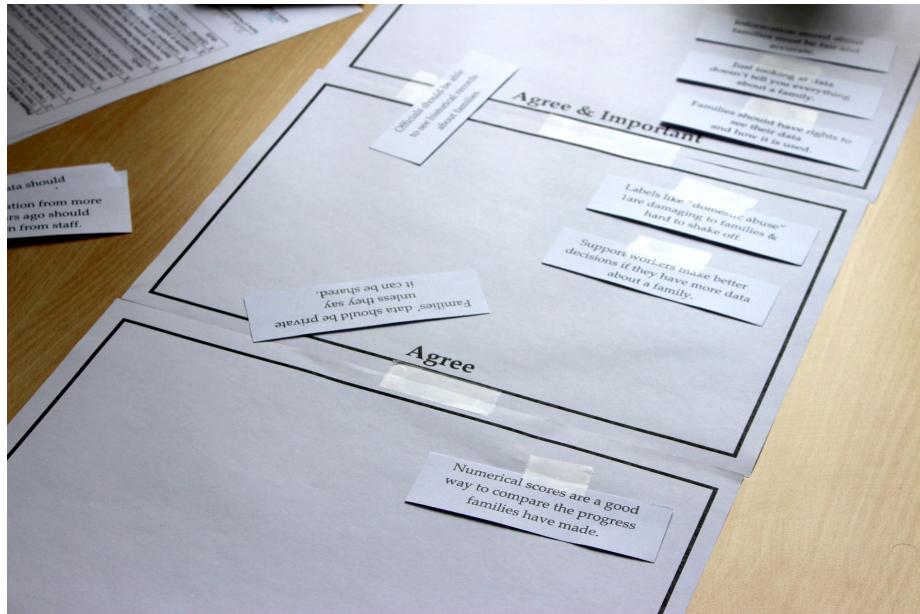


Figure 4: Figure 7: Sentence Ranking – Bringing staff and families to a shared problem space

of staff within the care service. In Case Study Two, my self-experiments with GDPR as well as researching privacy policies and GDPR rights provided me with similar sensitisation before engaging participants.

Participants need to be sensitised too; when planning participatory research activities such as interviews or workshops, it is important to begin the session with an activity that will acclimatise participants both to the specific area of discussion, but also to the mindset of problem solving required for a constructive conversation. This goes beyond ice-breaking to thinking about what the participants bring and lack at the start of the engagement. For example, in the Understanding Family Civic Data study, I felt that data would be a hard topic for families to engage with, so I designed the “Family Facts” activity shown in Figure 5. This required family members to consider simple facts about their lives (some provided, and some created by the family members) and discuss whether or not such a fact would be considered data, and additionally whether such a fact should be in the family’s control or that of the authorities. This served a double purpose of teaching families that data is simply “stored information about you”, while also getting them used to thinking critically about data ownership. The technique is discussed further in (Bowyer *et al.*, 2018).

For Case Study Two, I wanted to get participants (and potential participants) to think about the data involved in their everyday lives, especially that stored by commercial service providers. So I put up a series of posters in the common room of my research lab which showed logos of companies that might store data, types of data that might be stored, information about GDPR rights, and possible uses that an individual might have for data they obtain from a GDPR request. Some of these posters are shown in Figure 6. These posters served both as a recruitment tool for the project and were also visited with participants at the start of each interview as a series of talking points to sensitise the participants.

Sometimes sensitisation activities can also serve an additional purpose of bringing disparate participants to be “on the same page”, this is known in participatory research as *co-experience* (Battarbee and Koskinen, 2005). An example of this is the “sentence ranking” exercise used at the start of all workshops in Case Study Two and shown in Figure 7. Here, a series of sentences were prepared containing opinions about civic data that had been observed from staff and families in earlier research, and participants were asked to rank these according to agreement and importance. This allowed me to validate whether previous findings held with these new participants, but also sensitised the participants to considering and discussing the civic data context and the problems experienced by families and staff. Since the sentences included both staff and family viewpoints, and the activity was carried out in all workshops regardless of whether staff, families or both were present, it served to establish a common set of “requirements” that would be in participants’ minds as they began the subsequent co-design activity within each workshop.



Figure 5: Figure 8: Family Civic Data Cards – things to think with, that can also be used in card sorting tasks

1.5.2 Discussion and Exploration with Stimuli

[TODO ADD THE PERSONAL DATA TYPES SLIDE HERE]



Figure 6: Figure 9: Interviewing Families in the Home – Card sorting with a family in their living room

As discussed in 3.2, my research seeks to uncover individual perspectives and worldviews. The primary method that I used in both Case Study One and Two to do this is traditional qualitative interviewing - talking to people about the topic being explored. In Case Study Two, this was largely done on 1-on-1 basis (largely because of the sensitivity of dealing with one's own personal data, and because it allowed me as researcher to get closer to the participant's individual experience). In Case Study One, group discussions and activities were mainly used. This brought the advantage of being able to 'prime' a discussion between participants and then sit back into more of an observational role, which proved particularly insightful when observing intergenerational conversations between family members in the Understanding Family Civic Data study (Bowyer *et al.*, 2018), and in Case Study One it allowed me to observe the negotiation of a 'middle ground' between support workers and supported families. In some cases,

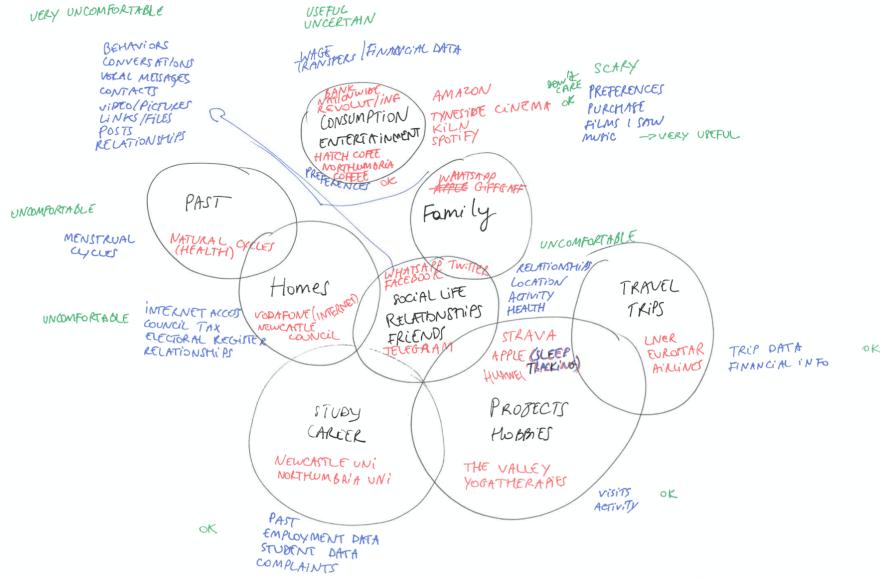


Figure 7: Figure 10: Sketching Dialogue – An example life sketch created by a participant and annotated during discussion

such as the home visits in the Understanding Family Civic Data study and some visits to council workers as part of my embedding in the SILVER project, I was able to conduct interviews-in-place (Pink *et al.*, 2013) in participants' own environments, which allowed for additional ethnographic observations to be made as "life happens around" (Mannay and Morgan, 2015) the participants, as discussed in (Bowyer *et al.*, 2018).

I wanted to go beyond 'just talking' to achieve a deeper and more detail-oriented conversation, and so in all of my interviews and group engagements I also ensured that suitable stimuli were created to seed and progress the discussion. Given the abstract nature of the topic of data, it does not always carry a clear meaning in people's everyday lives, so I needed to find a way to make the topic more vivid and real. Having sensitised myself to civic data as mentioned in the previous section, I constructed a taxonomy and lexicon for Family Civic Data, and created "Family Civic Data Cards" (shown in Figure 8) for use in activities and discussions. These serve as *boundary objects* (Star, 1989, 2010; Bowker *et al.*, 2015) - representational artifacts that are understandable by people who come from different perspectives, providing a common vocabulary for discussion (as well as serving to enable co-experience, detailed above). Each card represents a different category of data, including a summary and meaningful examples to make them be easy to digest, yet still containing sufficient detail to stimulate thinking. The cards were designed to be bright, child-friendly and appealing to engage with. The tangibility of these artifacts was important too,

they became *things to think with* (Papert, 1980; Brandt and Messeter, 2004) that could be used in discussions and in activities. Researchers have had success with the use of tangible objects to embody discussion concepts in order to stimulate and structure discussion, for example Coughlan's use of a dolls' house to explore attitudes to home energy use (Coughlan, Leder Mackley, *et al.*, 2013) or more recently Xie's Data City which used AR-enhanced cardboard models to represent data-processing functions (Xie, Ho and Wang, 2021). Many of these approaches have their roots in Dourish's concept of *embodied interaction* (Dourish, 2001). These cards were used throughout the Civic Data research in both sensitisation and *card sorting* (Spencer and Warfel, 2004) tasks, for example asking participants to position the cards on a pinboard according to perceptions about risk and ownership (see Figure 9), or sorting them into trays according to relative personal importance. The cards proved very effective at enabling a personal and detail oriented discussion: participants voluntarily opened up about sensitive topics (e.g. domestic violence or criminal records) raised by the cards because of their detached-but-relatable nature. The sketching dialogue technique (Hwang, 2021) used in the digital life context can also been as another application of this technique; by putting both participant and researcher's focus upon the page, rather than on each other, it can feel less invasive, more collaborative and makes it easier to focus on details (see Figure 10). Of course the ultimate stimulus for discussion about data is to view the actual data itself. Exploring data together with participants to elicit opinions and insights is a well-established technique (Coughlan, Brown, *et al.*, 2013; Chung *et al.*, 2016; Puussaar, Clear and Wright, 2017). This is the technique used within Case Study Two, asking participants about the data they retrieved from GDPR requests. The spreadsheet-based approach mentioned above was another example of a stimulus for discussion, and it allowed the Zoom-based interviews to retain a "gathered around the table looking at things together" ambiance despite the remoteness necessitated by COVID-19 restrictions.

1.5.3 Participatory Co-Design of Possible Solutions

In 3.2 I also introduced the concepts of participatory co-design (PD) as an additional research approach. This becomes particularly important when exploring solutions and ideals rather than understanding what participants perceive as problems. It involves bringing participants into a new mental space where they can imagine the realm of the possible, rather than just their current lived experience. Within Case Study One, PD was an important part of the research with both family and staff groups. In the early stages of a PD activity, it is important that participants are able to generate a wide range of ideas, even fantastical ones, without constraints, self-censoring or judgements. This is known as the 'discovery' phase in the UK Design Council's *double diamond* framework. (Design Council UK, 2004). Golembewski's *ideation decks* technique (Golembewski and Selby, 2010) was chosen for this purpose, as it allows participants to both select 'ingredients' of a design based on their own experience but also to combine them in a variety of different ways to generate novel ideas, guiding



Figure 8: Figure 11: Ideation Grids – Combining random design ingredients to generate new ideas

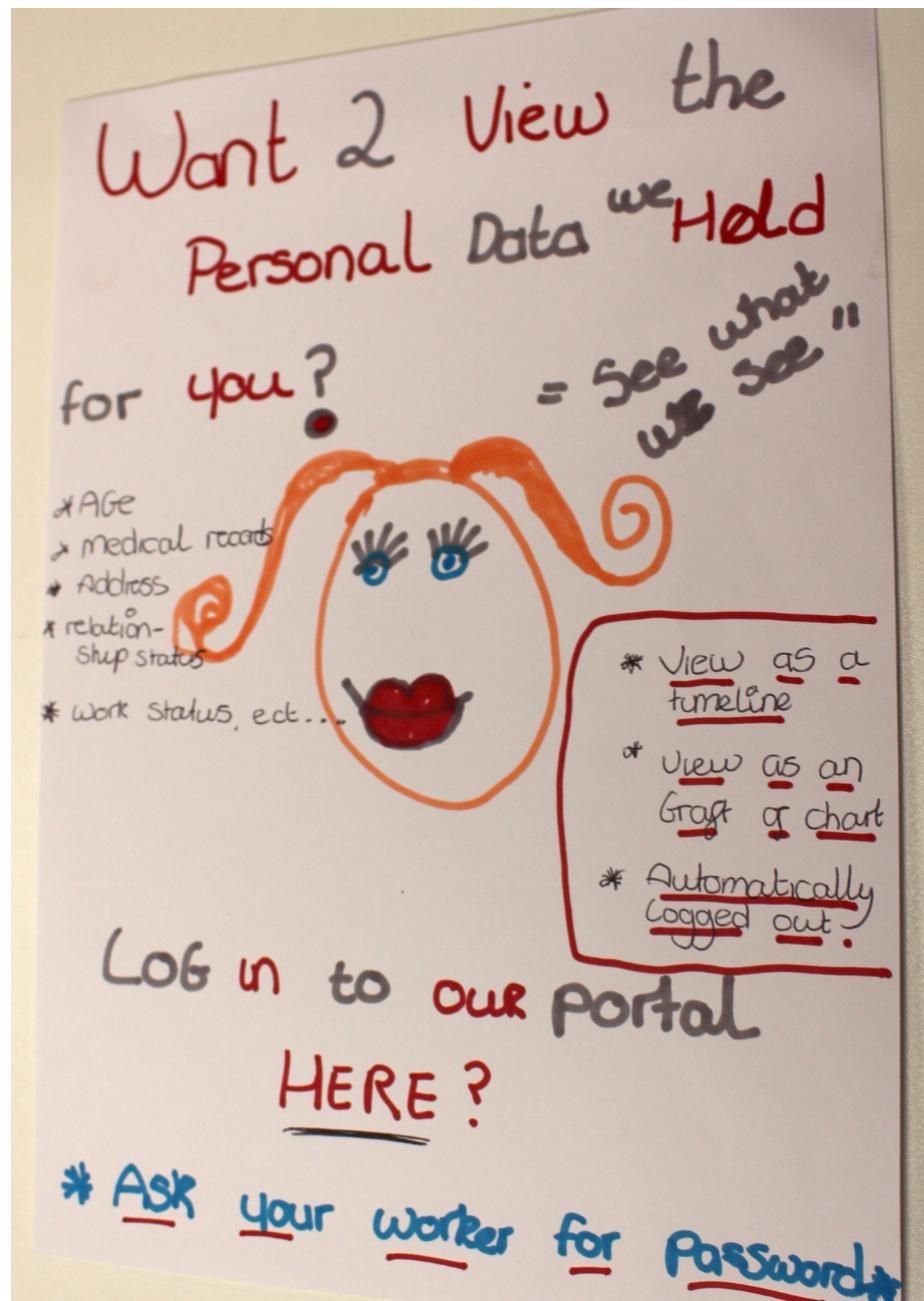


Figure 9: Figure 12: Group poster design – A participant-designed poster to advertise features of imagined data interface products

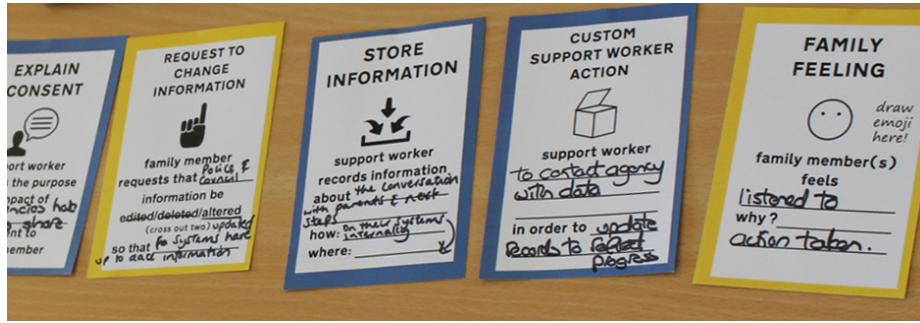


Figure 10: Figure 13: Storyboarding cards – A mutually constructed narrative created through discussion from a palette of possible parent and staff actions

them into a previously unconsidered solution space.

After generating a wide range of ideas using the ideation decks, participants were then invited to pick just one or two ideas to develop into posters, each with three ‘features’ highlighted. An example is shown in Figure 12. This activity corresponds to the ‘define’ phase of the double diamond, where participants narrow down the options.

For the final workshop of Case Study One, where both parents and staff were brought together to explore possibilities of shared data interaction within the support relationship, I used a Storyboarding activity. Drawing from the world of film production, storyboarding is a well-established technique in participatory design (Spinuzzi, 2005; Moraveji *et al.*, 2007). Usually it involves the participants drawing out a series of sketches in the form of a comic strip ‘telling the story’ of an interaction, encounter or activity. However as I wanted to focus on the interpersonal relations and process rather than the visual aspects of storytelling or interface design, I used a card-based approach to storyboarding, where participants selected actions from a palette of action cards representing different possible human or data interaction possibilities and annotated these with specific details. These cards are shown in Figure 13 and described in more detail in Chapter 4. The cards were designed with colour-coded borders to distinguish staff member actions (blue), parent actions (yellow) and shared actions (green), and participants demonstrated that they were confident to make their own decisions on their own action types, but to reach collaborative decisions on the shared actions.

1.5.4 Practical Data Experiments, Interface Design and Prototyping

In the BBC placement in particular, and also in the Self GDPR Experiments of Context Two and the development aspects of the embedded SILVER placement in Context One, the focus was not on uncovering individual perspectives, but on direct experimentation in the world to discover constraints and possibilities – in

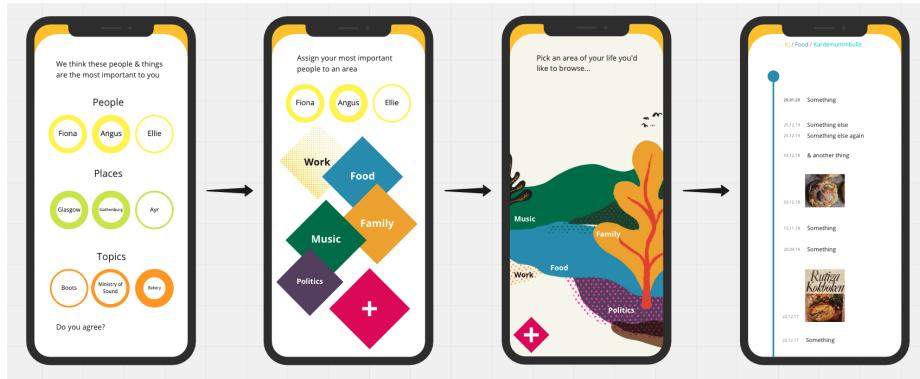


Figure 11: Figure 14: Visual Design Mockup for Life Partitioning in a PDS – A visual design mockup collaboratively created with BBC Research colleague Jasmine Cox

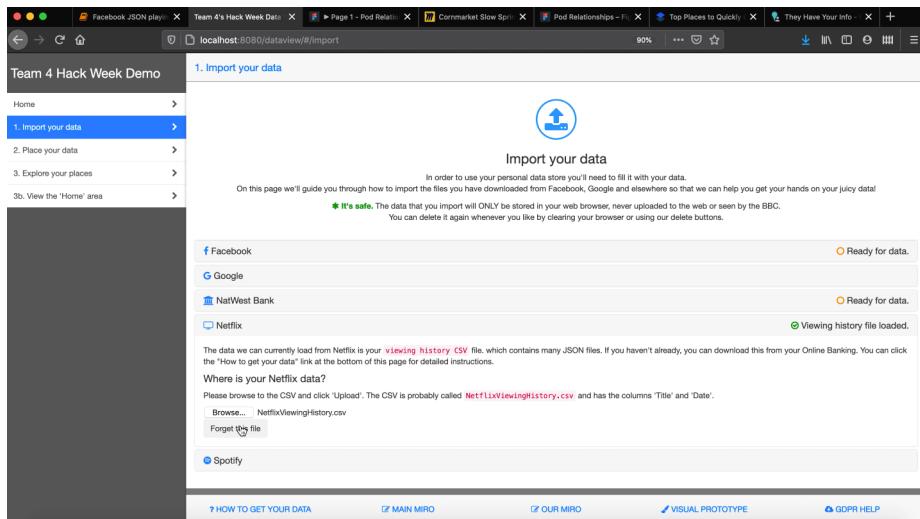


Figure 12: Figure 15: Prototype interface for GDPR Data Viewing – A working prototype that I developed during a hack week at BBC R&D

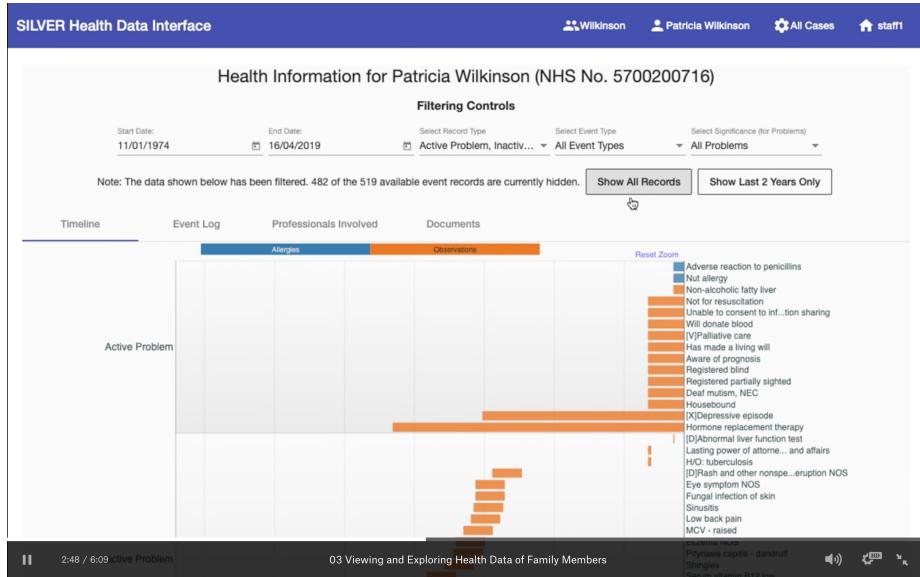


Figure 13: Figure 16: SILVER Health Data Viewing Interface – A working health data viewing interface for Early Help support workers that I developed as part of the SILVER project

line with the philosophy of Deweyan pragmatism referenced in 3.1. To design a better future, we must understand the world at it is, not just as people perceive it. Another justification is that as a designer or software developer, we need not only user requirements but knowledge of actual constraints and possibilities for implementation if we are to create something that is realistic and feasible for use in the real world. With this in mind, I conducted many practical explorations of data interaction throughout this thesis. Loosely these could be divided into design activities, prototyping, and interface development.

As part of my placement at BBC R&D, I co-designed a conceptual personal data locker interface for unifying a user's data from different sources and then partitioning it into different 'areas of life'. Our design was mocked up visually by BBC colleague Jasmine Cox and is shown in Figure 14. Imagining and iterating on possible interface designs and user flows is an important part of the process of prototyping possibilities - some ideas seem viable until you actually try to detail them.

As mentioned in 3.4.2.3, I had been gathering my own data from GDPR requests since 2018. This 'testing what is possible' of GDPR processes provided valuable insights to inform both RQ2 and RQ3 [TODO deformatilse RQ3], but also provided me with copies of my own personal data. At BBC R&D, I participated in 'hack week' as part of which explored possibilities for personal data locker interface designs. I used the data I had retrieved via GDPR and built a prototype

user interface in JavaScript, shown in Figure 15, that would import data files from different parts of life and extract information that could then be used to categorise and display my own data. Doing this activity heightened my understanding of what is possible with real GDPR-retrieved data, and the complexities of dealing with it and analysing it in practice.

As a front-end developer embedded within the SILVER project, I was responsible to build a functional user interface for support workers to explore health data, illustrated in Figure 16. This provided an opportunity to put the ideas of timelines and Temporal PIM (see section 2.2.2) into practice and explore which features are most useful; the SILVER project ran an evaluation workshop of this software with support workers at a local council which provided further insights into which features are most valuable when interacting with personal data.

1.5.5 Analysis, Modelling and Learning

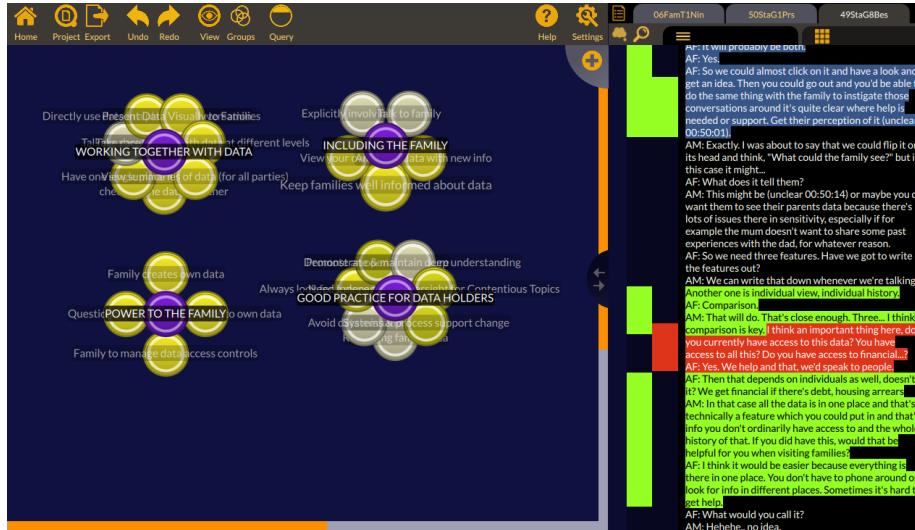


Figure 14: Figure 17: Thematic Analysis – A screenshot of thematic coding of qualitative data using Quirkos for Case Study One

In order to find common viewpoints and extract insights from the many participatory activities I conducted in Case Study One and Two, I needed to analyse the qualitative data. The general approach taken was to audio record (and occasionally video record) all interviews and workshops, and to produce a written transcript of the words spoken. Digital photos were taken to capture card arrangements, rankings and other transitory choices, as well as designs, life sketches and other participant creations. While it is possible to analyse participant designs in more detail, I chose to give them the sole purpose of adding contextual understanding to conversation transcripts and did not examine them further. Field notes were captured during or soon after each engagement. Then

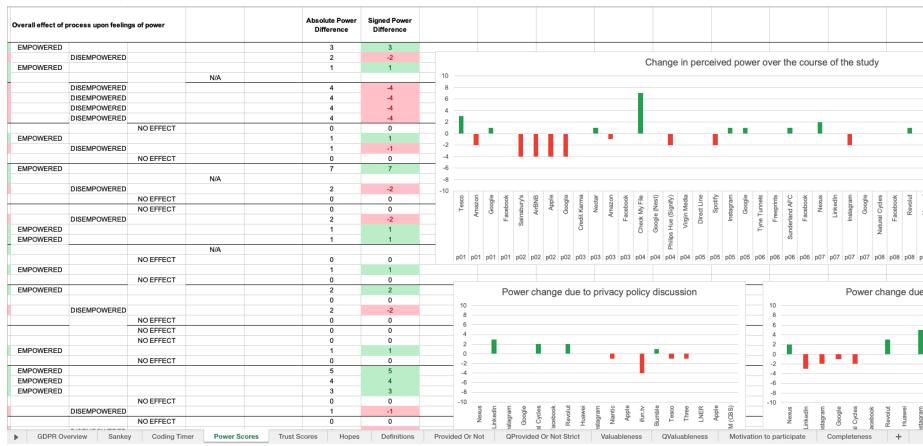


Figure 15: Figure 18: Quantitative Analysis – A screenshot of spreadsheet-based quantitative analysis of interview data from Case Study Two

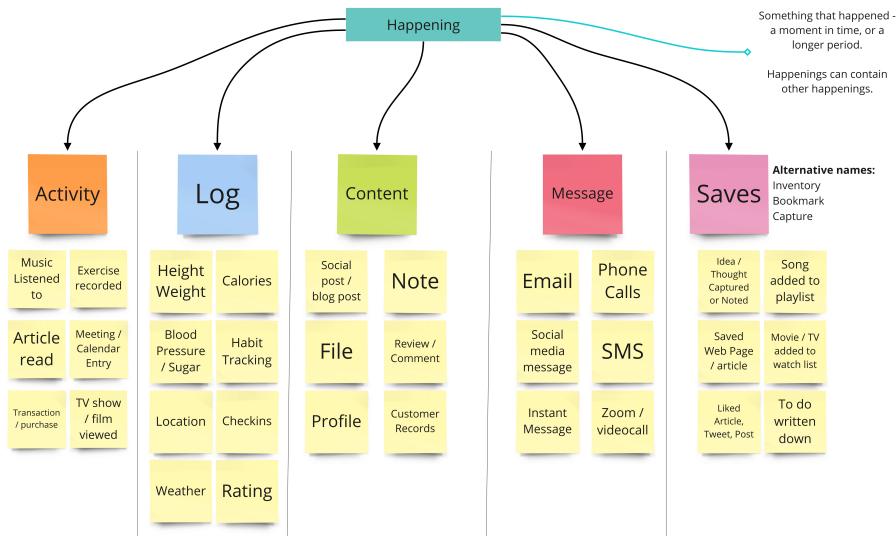


Figure 16: Figure 19: A Model for Personal Data – Developing a common model for personal data imported into a PDS a part of BBC Cornmarket R&D work

a process of thematic analysis was undertaken. This involved examining the text of the transcripts (with reference to all relevant digital artifacts to add context), and identifying the underlying ideas, themes and opinions of the participants. Thematic coding is a well established technique in qualitative research (Braun and Clarke, 2006). I selected the Quirkos software for this purpose, as shown in Figure 17, due to it having a more visual organisation and simpler approach than the more commonly used nVivo. After initial coding of transcripts, a process of reductive data display cycles (Huberman and Miles, 2002) was used to group codes into themes which became the key findings of the data chapters 4 and 5. In chapter 7, a similar approach was used, although in this case as this was not a participatory engagement, the source text was my own captured field notes informed by design materials and other digital files created as part of the research placement.

While the participant data in Case Study One and Two was largely free-flowing and very loosely structured conversation, the structure of some activities allowed some data to be captured numerically, notably the sentence rankings and data card placements in the Understanding Family Civic Data study and the trust/power ratings and GDPR spreadsheets produced in Case Study Two. These data points were captured into Excel spreadsheets, and where appropriate analysed using formulae to produce weighted mean averages and standard deviations to help contextualise the findings. An example is shown in Figure 18. Due to the qualitative focus of my research, participant numbers were too low to seek statistically significant findings, so all quantitative findings are not intended to be representative of any population at large.

As well as analysing participant data, an important aspect of pursuing answers to the three research questions was to develop theories, models and ideas and then to iteratively develop those models over time. This was particularly important in my BBC placement [TODO and Hestia, need to update this paragraph], which was the place where theoretical knowledge acquired from the first two case studies collided with practical reality. As part of this process, I produced many different models of personal data and of personal data interaction. In some cases I was able to test these by discussing them with expert colleagues at the BBC; in other cases by disseminating ideas through blogs, tweets, workshop papers and lectures, a process which helps to refine and clarify ideas but also stimulate valuable discussions with interested people to gain feedback that helps develop the models further. Figure 19 shows an example of a model I was developing for unifying personal data in the PDS context while embedded at BBC R&D [TODO and Hestia].

1.5.6 Recruitment

Table 1: Table 3.1 - Context One (Civic Data & Early Help):
Participants involved in research activities leading into Case Study
One.{#table-3.1}

Research Activity	Engagement	Stage or Phase	Duration	Number of Participants	Recruitment Method
Pilot (prior to PhD): Under-standing Family Civic Data study	4 x Home-based Interview	preliminary	4 x 2 hours	7 adults and 6 children from 4 families	Posters and Visits to Local Community Centre
Main study (Data Interaction in Early Help)	1 x Group Design Workshop for Families	1A	1 x 2 hours	8 adults and 9 children from 5 supported families	Selected by Local Authority Care Services
Main study (Data Interaction in Early Help)	2 x Group Design Workshop for Staff	1B	2 x 2 hours	36 supporters & related staff	Selected by Local Authority Care Services

Research		Stage or Phase	Duration	Number of Participants	
Activ-ity	Engagement			Recruitment Method	
Main study (Data Inter-action in Early Help)	1 x Combined Staff and Parents Group Design Workshop	2	1 x 2 hours	3 supporters and 4 parents from supported families	Selected by Local Authority Care Services

[TODO update above table so that naming is consistent with C4/C5]

Tables 1 and 2 summarises the participants involved in this research. In Case Study One, recruitment was initially attempted using posters placed in local libraries, as shown in Figure 20 below. When this approach was unsuccessful, participants were successfully recruited with the assistance of a local community centre [SHOULD I NAME IT?] which allowed me to visit a community social meeting and talk to residents about my study. This community was located in a low income area that was known to include a number of support families; in this way we were able to access for this informative study a population very similar to that which would reach through the local care authorities for the main study, avoiding some bureaucratic obstacles which were delaying recruitment through official channels. For the main engagement of Case Study One, I was able to work with two local authorities, Newcastle City Council and North Tyneside Council, who were partners on the SILVER project, and provided suitable participants who were actively involved in their Early Help programmes. In the preliminary study and in the first families workshop of the main study (stage 1A), activities were designed to include children as active participants in the research, as it was felt they would bring valuable contributions to the somewhat abstract creative co-design work and because it would be valuable to be able to observe intra-family conversations. The final combined workshop with staff (stage 2) however was designed to only include adult participants. This is because the focus on processes and on the care relationship itself was thought to be too boring and potentially sensitive for the children to participate.

Have you ever wondered what happens to
YOUR FAMILY'S INFORMATION
when you speak to councils, doctors and officials?

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



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

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

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

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

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



Open Lab
at Newcastle University

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

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

Figure 17: Figure 20: Recruitment Poster – Poster used to recruit participants for Understanding Family Civic Data study

Research Activity		Engagement	Stage or Phase	Duration	Number of Participants	Recruitment Method
Pilot:	5 x 1-on-1 interview		preliminary	5 x 2 hours	5 adults	Convenience sample
Digital Life Mapping study	Main study interview (Life Sketching) (Guided GDPR)		1	11 x 1 hour	11 adults	Convenience sample
Main study interview (Privacy Policy Reviewing) (Guided GDPR)	10 x 1-on-1 interview		2	10 x 1 hour	10 adults	Continuation from previous stage ¹
Main study interview (Viewing GDPR returned data)	(Guided GDPR)		3	10 x 2 hours	10 adults	Continuation from previous stage

Table: Table 2 - Context Two (Digital Life): Participants Involved in Digital Life Research Activities Leading into Case Study Two.

In Case Study Two, the digital life study, it was felt that no special population was needed, as the issues of living in a data-centric world would be likely to affect everyone. Therefore, a convenience sample (largely 20-40 year old postgraduate students from Newcastle University) was used. Care was taken to find an even split of male and female participants, but other than that no selection criteria was applied. The participants used for this study were thought likely to have a larger awareness of societal issues around personal data use, and greater familiarity with participatory co-design, than the average layperson, but this was considered an advantage as it would reduce the amount of sensitisation required.

In all cases² for both case studies, participants were compensated for their time with vouchers – either online/offline shopping vouchers or in the case of the families workshop, vouchers for a family day out of the family's choice.

¹One participant withdrew from the study after the first interview of the Guided GDPR study due to COVID-19. The other 10 participants took part in all three stages.

²(with one exception - the staff workshops within Case Study Two. Because the participants were attending the workshops through their employers (the local authorities), we were not allowed to provide vouchers for participation.)

1.5.7 Ethics

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

1. *Visiting private homes*: In order to protect myself and other researchers from any physical risks or any accusations of impropriety, all home visits took place with two researchers present, and contact was made with a colleague before and immediately after the interviews to confirm everything was ok.
2. *Working with children*: Activities were designed to be child-friendly (not just safe, but engaging). The families workshop took place at a park with a nearby cafe and playgrounds for children, and catering was provided. Within the room, an activity area was provided for smaller children who were not directly participating to play while their parents and older siblings engaged. There was always more than one researcher present and the research team was never alone with children.
3. *Protecting personal data privacy*: In Case Study Two, particular care was taken to design ways for researchers to talk to people about their personal data without violating participants' right to privacy. The research was positioned that the data retrieved from companies was participants' own data, that would never be directly collected or handled by the research team, it was made clear that as researchers we were only interested in what was said, not the data itself. Initially a privacy monitor was developed which could only be seen with viewing glasses that were in the participant's control. This would allow a researcher to sit next to a participant who was viewing his/her personal data, without the researcher being able to see it. Additional measures to protect users' data included clear instructions on how to keep data safe before, during and after the study. A complaints procedure was also written at the request of the Ethics board.
4. *Adapting to COVID-19*: As COVID-19 changed working and living conditions in early 2020, Case Study Two was adapted to no longer rely on face-to-face engagement. The in-person privacy monitor approach was abandoned and replaced with an online Zoom-based approach. In this model participants would share parts of their data using screen sharing instead, and could move windows off screen to protect their privacy. The full study plan for Case Study Two was rewritten for online-based participation

and was re-approved by the Ethics Board.

1.5.8 Thesis structure approach

[TODO CUT THIS SECTION AND COVER IT IN INTRO]

In writing up this thesis, I made a choice to foreground my two most major research activities as Case Studies, and not to detail the other activities carried out beyond the high level summaries included in this chapter. Case Study One and Two both span research questions RQ1 and RQ2 (see Figure 4 in section 3.4) as they explore both people's direct relationship with data *and* the relationships people have that indirectly involve data. The learnings that allow me to explore and draw conclusions that serve RQ3 [TODO deformalise RQ3] come from a variety of practical activities, described in 3.4.3 above, all focused on designing and bringing about better human data relations in practice.

Because of the overlapping of RQs 1 and 2 in Case Study One and Two, I have structured the subsequent chapters as follows:

- Chapter 4 details Case Study One in the context of both RQ1 & RQ2.
- Chapter 5 details Case Study Two in the context of both RQ1 & RQ2.
- Chapter 6 is the first discussion chapter, which separately unifies RQ1 findings and RQ2 findings, so that they can be referenced in general terms as distinct understanding of people's wants in their direct relationships to data (RQ1) and their wants in relationships with those who hold data about (i.e. their indirect relationships to data) (RQ2), drawing from insights that span both case studies.
- Chapter 7 is the concluding discussion chapter, which examines how those needs identified in Chapters 4, 5 and 6 might be achieved in practice, through software development, education and civic action. This covers all three research subquestions and draws these together to address the main research question.

Bibliography

- Abiteboul, S., André, B. and Kaplan, D. (2015) *Managing your digital life with a Personal information management system*. 5. ACM, pp. 32–35. doi: 10.1145/2670528.
- Adams, R. (2017) ‘Michel Foucault: Discourse’. Available at: <https://criticallegalthinking.com/2017/11/17/michel-foucault-discourse/> (Accessed: 7 May 2021).
- Battarbee, K. and Koskinen, I. (2005) ‘Co-experience: user experience as interaction’, *CoDesign*. Taylor & Francis, 1(1), pp. 5–18.
- BBC R&D (2017) ‘Human Data Interaction - BBC R&D’. Available at: <https://www.bbc.co.uk/rd/projects/human-data-interaction>.
- Bjerknes, Gro. *et al.* (1987) *Computers and democracy : a Scandinavian challenge*. Aldershot [Hants, England]; Brookfield [Vt.], USA: Avebury, p.

434. Available at: <http://www.worldcat.org/title/computers-and-democracy-a-scandinavian-challenge/oclc/614994092?referer=di&ht=edition>.
- Björgvinsson, E., Ehn, P. and Hillgren, P.-A. (2010) 'Participatory design and "democratizing innovation"', in *Proceedings of the 11th biennial participatory design conference*, pp. 41–50.
- Bowker, G. C. et al. (2015) *Boundary objects and beyond : working with Leigh Star*. MIT Press, p. 548. Available at: <https://books.google.co.uk/books?hl=en&lr=&id=nmSkCwAAQBAJ&oi>-Objects and Beyond%3A Working with Leigh Star&f=false.
- Bowyer, A. (2018) 'Free Data Interfaces: Taking Human- Data Interaction to the Next Level', *CHI Workshops 2018*. Available at: <https://eprints.ncl.ac.uk/273825>.
- Bowyer, A. et al. (2018) 'Understanding the Family Perspective on the Storage, Sharing and Handling of Family Civic Data', in *Conference on human factors in computing systems - proceedings*. New York, New York, USA: ACM Press, pp. 1–13. doi: 10.1145/3173574.3173710.
- Brandt, E. and Messeter, J. (2004) 'Facilitating collaboration through design games', in *Proceedings of the eighth conference on participatory design artful integration: Interweaving media, materials and practices - PDC 04*. New York, New York, USA: ACM Press, p. 121. doi: 10.1145/1011870.1011885.
- Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*. Taylor & Francis, 3(2), pp. 77–101. doi: 10.1191/1478088706qp063oa.
- Brown, D. (2015) 'Here's what 'fail fast' really means'. Available at: <https://venturebeat.com/2015/03/15/heres-what-fail-fast-really-means/>.
- Campbell, P. L. (2011) 'Peirce, pragmatism, and the right way of thinking', *Sandia National Laboratories, Albuquerque*. Citeseer.
- Cheetham, M. et al. (2018) 'Embedded research: A promising way to create evidence-informed impact in public health?', *Journal of Public Health (United Kingdom)*. Oxford University Press, 40(suppl_1), pp. i64–i70. doi: 10.1093/pubmed/fdx125.
- Chevalier, J. M. and Buckles, D. J. (2008) *SAS2: A guide to collaborative inquiry and social engagement*. SAGE Publishing India.
- Chevalier, J. M. and Buckles, D. J. (2019) *Participatory action research: Theory and methods for engaged inquiry*. Routledge.
- Chung, C. F. et al. (2016) 'Boundary negotiating artifacts in personal informatics: Patient-provider collaboration with patient-generated data', *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 27, pp. 770–786. doi: 10.1145/2818048.2819926.
- Connected Health Cities (2017) 'SILVER Project: Smart Interventions for Local Residents'. Available at: <https://web.archive.org/web/20210308040602/https://www.connectedhealthcities.org/research-projects/troubled-families/> (Accessed: 8 March 2021).
- Coughlan, T., Leder Mackley, K., et al. (2013) 'Current issues and future directions in methods for studying technology in the home', *PsychNology Journal*, 11(2), pp. 159–184.
- Coughlan, T., Brown, M., et al. (2013) 'Methods for studying technology in the

- home', in *CHI'13 extended abstracts on human factors in computing systems*, pp. 3207–3210.
- Design Council UK (2004) 'What is the framework for innovation? Design Council's evolved Double Diamond'. Available at: <https://www.designcouncil.org.uk/news-opinion/what-framework-innovation-design-councils-evolved-double-diamond> (Accessed: 20 May 2021).
- Dewey, J. and Archambault, R. D. (1964) 'John Dewey on education: Selected writings'.
- DiSalvo, C. (2010) 'Design, Democracy and Agonistic Pluralism', *Proceedings of the Design Research Society Conference 2010*, pp. 366–371.
- DiSalvo, C. (2012) *Adversarial Design*. MIT Press (Design thinking, design theory). doi: 10.7551/mitpress/8732.003.0007.
- Dourish, P. (2001) *Where the action is: the foundations of embodied interaction*. MIT press.
- Fowler, M. and Highsmith, J. (2001) 'The agile manifesto', *Software Development*. [San Francisco, CA: Miller Freeman, Inc., 1993-, 9(8), pp. 28–35.
- Friedman, R. L. (2006) 'Deweyan Pragmatism', *William James Studies*, 1. Available at: <https://williamjamesstudies.org/deweyan-pragmatism/>.
- Golembewski, M. and Selby, M. (2010) 'Ideation decks', in *Proceedings of the 8th ACM conference on designing interactive systems - DIS '10*. New York, New York, USA: ACM Press, p. 89. doi: 10.1145/1858171.1858189.
- Guba, E. G. (1990) 'The alternative paradigm dialog', *The paradigm dialog*. Sage Publications, Inc, pp. 17–30. Available at: <http://www.jstor.org/stable/3340973>.
- Hayes, G. R. (2011) 'The relationship of action research to human-computer interaction', *ACM Transactions on Computer-Human Interaction*, 18(3), pp. 1–20. doi: 10.1145/1993060.1993065.
- Hoffman, W. (2010) 'Rethinking Personal Data'. Available at: <https://web.archive.org/web/20110220013300/http://www.weforum.org/issues/rethinking-personal-data>.
- Hoffman, W. (2011) *Personal data : The emergence of a new asset class*. World Economic Forum, pp. 1–40. Available at: <http://www.weforum.org/reports/personal-data-emergence-new-asset-class>.
- Hoffman, W. (2013) *Unlocking the Value of Personal Data: From Collection to Usage Prepared in collaboration with The Boston Consulting Group Industry Agenda*. February. World Economic Forum.
- Hoffman, W. (2014a) *Rethinking Personal Data : A New Lens for Strengthening Trust*. May. World Economic Forum, p. 35. Available at: http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf.
- Hoffman, W. (2014b) *Rethinking personal data: Trust and context in user-centred data ecosystems*. May. World Economic Forum, p. 35. Available at: http://www3.weforum.org/docs/WEF_RethinkingPersonalData_TrustandContext_Report_2014.pdf.
- Hosch, W. L. (2017) 'Web 2.0'. Available at: <https://www.britannica.com/topic/Web-20> (Accessed: 26 April 2021).
- Huberman, M. and Miles, M. B. (2002) *The qualitative researcher's companion*. Sage.

- Hwang, E. (2021) ‘Sketching Dialogue : Incorporating Sketching in Emphatic Semi-structured Interviews for HCI’.
- Jenkins, H. (2006) *Convergence Culture: Where Old and New Media Collide*. New York, USA: New York University Press. doi: 10.7551/mitpress/9780262036016.003.0012.
- Kensing, F. and Blomberg, J. (1998) ‘Participatory design: Issues and concerns’, *Computer supported cooperative work (CSCW)*. Springer, 7(3), pp. 167–185.
- Lewin, K. (1951) ‘Problems of research in social psychology’, *Field theory in social science: Selected theoretical papers*, pp. 155–169.
- Li, I., Forlizzi, J. and Dey, A. (2010) ‘Know thyself: Monitoring and reflecting on facets of one’s life’, *Conference on Human Factors in Computing Systems - Proceedings*, pp. 4489–4492. doi: 10.1145/1753846.1754181.
- Mannay, D. and Morgan, M. (2015) ‘Doing ethnography or applying a qualitative technique? Reflections from the ‘waiting field’’, *Qualitative research*. Sage Publications Sage UK: London, England, 15(2), pp. 166–182.
- McCarthy, J. and Wright, P. (2004) ‘Technology as experience’, *Interactions*, 11(5), pp. 42–43. doi: 10.1145/1015530.1015549.
- Moraveji, N. et al. (2007) ‘Comicboarding: Using comics as proxies for participatory design with children’, in *Conference on human factors in computing systems - proceedings*. ACM, pp. 1371–1374. doi: 10.1145/1240624.1240832.
- Norman, D. A. and Draper, S. W. (1986) ‘User Centered System Design; New Perspectives on Human-Computer Interaction’. L. Erlbaum Associates Inc.
- Northern Health Science Alliance (2020) *Connected Health Cities Impact Report*. Newcastle upon Tyne, UK: Northern Health Science Alliance, pp. 129–130.
- Papert, S. (1980) ‘Mindstorms: children, computers, and powerful ideas’. Basic Books, Inc.
- Peikoff, L. (1993) *Objectivism: The Philosophy of Ayn Rand*. Penguin Publishing Group (Ayn rand library). Available at: <https://books.google.co.uk/books?id=G6DDIqNftGcC>.
- Pink, S. et al. (2013) ‘Applying the lens of sensory ethnography to sustainable HCI’, *ACM Transactions on Computer-Human Interaction (TOCHI)*. ACM New York, NY, USA, 20(4), pp. 1–18.
- Puussaar, A., Clear, A. K. and Wright, P. (2017) ‘Enhancing personal informatics through social sensemaking’, in *Proceedings of the 2017 CHI conference on human factors in computing systems*, pp. 6936–6942.
- Reason, P. and Bradbury, H. (2001) *Handbook of action research: Participative inquiry and practice*. Sage.
- Ries, E. (2011) *Wiki: The Lean Startup: How Today’s Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. Crown. Available at: http://en.wikipedia.org/wiki/Lean_Startup.
- Smith, N. K. (2011) *Immanuel Kant’s critique of pure reason*. Read Books Ltd.
- Smith, R. C., Bossen, C. and Kanstrup, A. M. (2017) ‘Participatory design in an era of participation’, *CoDesign*. Taylor & Francis, 13(2), pp. 65–69. doi: 10.1080/15710882.2017.1310466.
- Spencer, D. and Warfel, T. (2004) ‘Card sorting: A definitive guide’, *Boxes and arrows*, 2(2004), pp. 1–23.

- Spinuzzi, C. (2005) 'The methodology of participatory design', *Technical Communication*. Society for Technical Communication, 52(2), pp. 163–174.
- Star, S. L. (1989) 'The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving', in *Distributed artificial intelligence*. Elsevier, pp. 37–54. doi: 10.1016/b978-1-55860-092-8.50006-x.
- Star, S. L. (2010) 'This is not a boundary object: Reflections on the origin of a concept', *Science Technology and Human Values*, 35(5), pp. 601–617. doi: 10.1177/0162243910377624.
- Vlachokyriakos, V. et al. (2016) 'Digital civics: Citizen empowerment with and through technology', *Conference on Human Factors in Computing Systems - Proceedings*, 07-12-May-, pp. 1096–1099. doi: 10.1145/2851581.2886436.
- Xie, A., Ho, J. C. F. and Wang, S. J. (2021) 'Data City: Leveraging Data Embodiment Towards Building the Sense of Data Ownership', pp. 365–378. doi: 10.1007/978-3-030-73426-8_22.