

Understanding and Designing Human Data Relations

Alex Bowyer

- [Frontmatter](#)
- [1 Introduction](#)
- [2 Literature Review](#)
 - [2.1 Data-centrism and the Need for Access](#)
 - [2.2 Personal Data Interaction](#)
 - [2.3 Practical Human-centred Design](#)
- [3 Methodology](#)
 - [3.1 Forming a Research Paradigm: Ontology & Epistemology](#)
 - [3.2 Research Approach: Participatory Action Research & Experience-centred Design](#)
 - [3.3 Research Objectives](#)
 - [3.4 Overview of Research Contexts and Activities](#)
 - [3.5 Methodologies Employed in Case Studies](#)
- [4 Case Study One: Accessing and Using Civic Data in Early Help](#)
 - [4.1 Context: Data Use in Early Help](#)
 - [4.2 Preliminary Explorations of Family Civic Data: Families' and Support Workers' Perspectives](#)
 - [4.3 Thematic Findings](#)
 - [4.4 Discussion](#)
 - [4.5 Summation](#)
- [5 Case Study Two: The Human Experience of GDPR](#)
- [6 Bridge: An Understanding of Human Data Relations](#)
- [7 Case Study Three: Personal Data Interface Design & Development](#)
- [8 Discussion: Designing Better Human Data Relations](#)
- [Bibliography](#)

Frontmatter

This is the example frontmatter file. Use it for your abstract, dedications, acknowledgements etc.

1 Introduction

2 Literature Review

2.1 Data-centrism and the Need for Access

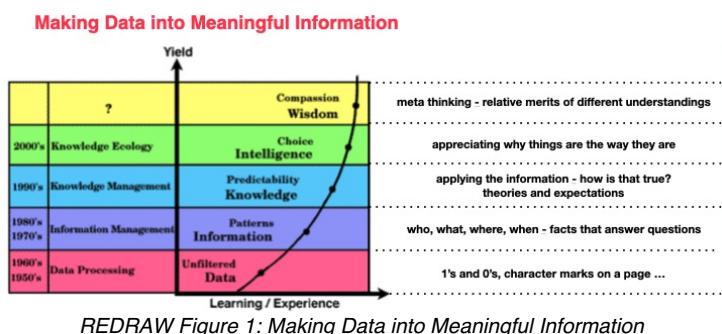
2.1.1 What is Data?

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

Wellisch expressed this as ‘the representation of concepts or other entities, fixed in or on a medium in a form suitable for communication, interpretation, or processing by human beings or by automated systems’ (Wellisch, [1996](#)), which is a useful definition as it includes the fact that both humans and algorithms can use data, and that data is something that needs interpretation.

From a strict grammatical stance, ‘data’ is a plural of the singular ‘datum’ thus it is more correct to write ‘the data are correct’ - but this usage is rapidly declining from use (‘Data’, [no date](#)) and throughout this thesis I use the more widely adopted usage of treating data as a singular mass noun, as in ‘the data is correct’.

The concepts of ‘data’ and ‘information’ are closely related, so much so that they are often used interchangeably. Ackoff presented a model for distinguishing data, information, knowledge, understanding/intelligence and wisdom, in which he describes data as the physical symbols, effectively the 1’s and 0’s stored in a computer or the ink marks on a page, which becomes useful when humans or algorithms are able to deduce facts from those symbols to answer simple questions - at this point it becomes ‘information’. Being able to interpret deeper how and why questions allow information to become knowledge and understanding, towards the ultimate goal of wisdom (Ackoff, [1989](#)). This is often represented as the DIKW pyramid (DIKW being shorthand for the data-information-knowledge-wisdom transformation that occurs as you move up through the layers), the origin of which is unknown (Wallace, [2007](#)). Figure 1 builds upon a representation by George Pór (Pór, [1997](#)) of the pyramid as a ‘wisdom curve’, showing how increasing meaning and value can be obtained from data as deeper questions can be asked of it. This theme of obtaining meaning and value from data is an important aspect of my research that I will refer back to.



This model that turning data into information can be thought of as using that data to answer questions is consistent with the idea that “information can be thought of as the resolution of uncertainty” (‘Information’, [no date](#)). The exact origin of this definition is unknown but it is often attributed to mathematician Claude Shannon (Shannon, [1948](#)). Indeed from an etymological stance, one who is informed is one who has received knowledge or concepts as a result of what has been communicated to them. Thus we can consider that data is the material from which information can be received. It follows also that data contains uncertainty that must be resolved in order for it to become meaningful information.

2.1.2 The Rise of Data-centrism

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

The personal computer revolution ('The personal computer revolution', [no date](#)) of the late 1970s and 1980s put computers in every office and eventually every home too, and it soon became commonplace that each individual would have data stored about them in companies' databases. In the subsequent years three factors have combined to accelerate this trend of storing data about people: i) labour costs have remained high and companies have sought ways to automate their businesses and to implement online services and call centres in place of in-person staff interaction, ii) computer processing and storage has become ever cheaper thanks to the advent of cloud computing, meaning that many business processes could be reduced to data processing tasks or entire businesses be moved online, and iii) the rise of smartphones and web-enabled devices have meant that the public are now ready and willing to conduct much of their daily business online through the web and apps. These factors have encouraged both commercial and civic providers to centralise their services and to 'go digital' to the greatest degree possible. In doing so they collect ever more data about people (now 'service users' or just 'users'). Data is now seen as a resource which can be mined for value, and harnessed for profit and business efficiency - 'the new oil' (Toonders, [2014](#)). Zuboff, in her 2019 book on 'surveillance capitalism', characterises this new digital world as the collection of human behaviour data so that it can be used as free raw material and converted into profit through hyper-personalised advertising and targeting by software platforms (Zuboff, [2019](#)). This philosophy is also known as 'data-ism' (Brooks, [2013](#)) and the analysis and exploitation of such data at scale is known as 'big data' (Neef, [2015](#)).

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

Unfortunately, the large-scale systems which collect data about us now function as 'data traps' (Abiteboul, André and Kaplan, [2015](#)) - where data about us is easily gathered but very hard to remove or even to access. This creates a lack of agency for the individuals living in this data-centric world. The World Economic Forum's "Rethinking Personal Data" project recognised the critical role that data, (specifically *personal data* - data created by and about people) now holds, and identified that "an asymmetry of power exists today [...] created by an imbalance in the amount of information about individuals held by industry and governments, and the lack of knowledge and ability of the same individuals to control the use of that information" (Hoffman, [2011](#), [2013](#), [2014b](#), [2014a](#)).

2.1.3 Data Protection & GDPR

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

“the right of individuals to access and challenge personal data is [...] the most important privacy protection safeguard” and issued recommendations that individuals should be given basic privacy rights, including the right to be informed whether data is stored about them, and the right to an intelligible copy of that data (Organisation for Economic Co-operation and Development, [1980](#)).

Over the subsequent decades, lawmakers began to enact laws to deliver these rights to individuals, notably the UK’s Data Protection Act 1984 (which set up an independent body, the Data Protection Registrar (now the Information Commissioner’s Office) with which organisations were required to register their usage of personal data), Ireland’s Data Protection Act 1988 (which introduced the concept of a ‘duty of care’ for data collectors - that they are expected to avoid causing damage or distress to data subjects), the EU’s Data Protection Directive in 1995 and the UK’s Data Protection Act in 1998. However, such laws were generally found to be ineffective - in 2002 Simon Davies, director of Privacy International said that the UK’s DPA was “almost useless in limiting the growth of surveillance” (Millar, [2002](#)).

It was only in 2018, when the EU’s General Data Protection Regulation (GDPR) came into force, carrying with it significant designed-to-hurt fines for non-compliance (Kelly, [2020](#); Leprince-Ringuet, [2021](#)), that individuals have been able to practically exercise their data rights to any meaningful degree (‘The GDPR: Does it Benefit Consumers in Any Practical Way?’, [2020](#)). The GDPR – which gives individuals key rights including rights to timely data access, explanation, erasure and correction (Information Commissioner’s Office, [2018](#)) – can be seen as the first serious attempt to rebalance the aforementioned power imbalance over data between citizens and organisations and is generally regarded as a landmark piece of legislation and a strong template for individual data protection. Around the world, companies have overhauled their privacy policies and updated their business practices to comply with the GDPR and other similar legislation, such as Japan’s 2017 Act on the Protection of Personal Information, India’s 2019 Personal Data Protection Bill and the 2020 California Consumer Protection Act. In the USA, there has been no national privacy law yet, but the GDPR’s influence is being felt in court rulings (Hoofnagle, Sloot and Borgesius, [2019](#)).

Also in 2018, the Cambridge Analytica scandal (‘Facebook–Cambridge Analytica Data Scandal’, [2014](#)) broke; the personal data of 87 million people, acquired from Facebook, was exploited with the apparent intent of influencing voting outcomes including the UK’s 2016 Brexit referendum and the USA’s 2017 election of Donald Trump. This combined with widespread public information campaigns about GDPR have led to a heightened awareness of personal data rights (European Union Agency for Fundamental Rights, [2020](#)) and at the time of writing in 2021, personal data protection laws and individual digital rights remain a rapidly evolving area.

From the GDPR and its antecedents, a number of key terms have been established which I will adopt in this thesis, specifically (Information Commissioner’s Office, [2014](#); The European Parliament and the Council of the European Union, [2016](#)):

- *Personal data* means any information relating to an identifiable natural person - one who can be identified directly or indirectly by reference to an identifier such as a name, identification number or location or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person.
- The *data subject* is the identified individual, living or deceased, who the personal data relates to.
- A *data controller* is the legal entity (company, public authority, agency, individual or other body) which collects or stores personal data about an individual and determines the means and purposes for which it is processed. Liability for data protection compliance rests with the data controller.
- A *subject access request* is the right to a copy of your personal data.
- *Data portability* is the right to receive a copy of all stored data about you, not just that which you provided, in an accessible and machine-readable format such as a CSV file, so that you can transport it to another service or make use of it.

2.1.4 The Need for Practical and Effective Data Access

The World Economic Forum called in 2011 for a balanced ecosystem around personal data, and identified

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

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

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

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

Beyond these operational concerns over effective access, there are practical limitations affecting people’s ability to make use of their data. Where people are given interfaces their data, access is typically via a list or feed combined with a search box. Studies have shown that people prefer to find information by *orienteering* rather than search - associatively traversing related datapoints (Teevan *et al.*, 2004; Karger and Jones, 2006). Having our documents distributed across multiple platforms, applications and devices makes interrogation and orienteering hard (Krishnan and Jones, 2005). Abowd and Mynatt highlight that in presenting information about people and their activities, everyday computing needs to address the facts that users activities rarely have a clear beginning or end, are often interrupted, are often concurrent with other activities; that *time* is an important factor in finding and interpreting information; and that associative modelling of information is more useful than hierarchical models, because future usage goals cannot always be anticipated (Abowd and Mynatt, 2000). Recognising these needs, Krishnan and Jones identify that an effective information access system should support giving historical context, *finding trends and patterns*, time-based contextual retrieval, automatic structuring and multiple perspectives of the information (Krishnan and Jones, 2005). Shneiderman, in the context of considering the effectiveness of interactive information visualisations, identified the need to support seven types of information interaction: *overview*, *pan & zoom*, *focus* (context & distortion), *detail on demand*, *filter*, *relate*, *history* and *extract* (Shneiderman, 1996). While any one of the capabilities mentioned in this paragraph does exist in at least some data interfaces today, it is clear that no such general-purpose personal information access system exists with all or even most of those capabilities exists today. The development and state of the art in the field of Personal

Information Management Systems is explored in section 2.2 below.

2.1.5 Research Gap: The Human Experience of Data

In this section, I have described the establishment of the data-centric world in which we live today, the imbalance this creates between data subjects and data controllers, and what can be viewed as nascent attempts by governments to redress that imbalance through the creation of new laws. I have also outlined where research thinking has exceeded the practical data capabilities we have today, in identifying many factors and capabilities that should be considered when it comes to giving people a meaningful relationship with their personal data.

To date, people's relationship with their personal data and the information within it has barely been explored. What mental models do people have around data? What value does it carry to them and what meaningful place does it (or should it) hold in their life? What is it that makes data meaningful and what do people want from their data? What is it like to live in this data-centric world where your abilities over your data are limited by lack of access to data and a lack of suitable interfaces and technologies to properly manage your digital life? This is one aspect of the research gap this thesis will address - discovering the human experience of data.

2.2 Personal Data Interaction

2.2.1 Computers as General-Purpose Information Tools

In the immediate aftermath of the second World War, Dr. Vannevar Bush wrote a landmark article for *The Atlantic Monthly* in which he envisioned a new scientific agenda for America and the world - to harness new general information-processing capabilities of computers to make the stored knowledge of mankind accessible and usable to all, for the betterment of society. He proposed the 'Memex', a device in which people would store their books, communications and records digitally so that it "might be consulted with exceeding speed and flexibility" - a personal filing system to serve as "an enlarged intimate supplement to his memory". He emphasised the importance of allowing information to be stored in "associative chains of related materials" so that people would be able to retrieve information in the same way we think of it, traversing related items or ideas (Bush, [1945](#)). During the next three decades, while computer systems were moving out of science labs and being established in workplaces as a means to automate and improve business processes, researchers began to look beyond usage in business and consider how computers might be used by 'the common man' to store one's personal information in digital files (Nelson, [1965](#)), for interpersonal communication (Shannon, [1948](#)), to augment human intellect (Engelbart, [1962](#)) and to model human thought (Simon and Newell, [1958](#)).

Collectively, these constituted a recognition that computers could be considered a general-purpose tool that anyone could use for their own purposes, and in the 1970s and 1980s the home computer revolution ('The personal computer revolution', [no date](#)) seemed to place the potential power that "having reduced your affairs to software, software can take care of them for you" (Gelernter, [1994](#)) into the hands of ordinary people.

2.2.2 Personal Information Management

Through the examination of people's desk-based working practices, researchers began to understand how people handle information to inform the design of computer information systems. In 1983, Thomas Malone observed that categorisation is hard, and that any system must not only help the user to find information, but also remind the user of things to do. Computers could help through automatic classification, but should also allow both physical and logical "piles" of information to be arranged by the user (Malone, [1983](#)).

Personal Information Management (PIM) was first mentioned in 1988 by Mark Lansdale, who identified a need to design information management systems according to the psychology of the people who use them rather than by simulating office practices. By paying attention to how people categorise, recognise and recall information, and labelling information with appropriate attributes, information can be retrieved by

different properties (Lansdale, [1988](#)). PIM includes both directly interacting with digital files, webpages and e-mails as well as ‘meta-activities’ such as finding, arranging, searching, browsing, re-finding, categorising, sensemaking, keeping and discarding personal information. William Jones summarised PIM as “the art of getting things done in our lives through information” (W. Jones, [2011a](#)).

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

Spatial PIM systems are based on the idea that people remember “where” they have put things and that this allows information to be quickly returned by associating it with a place (Negroponte and Bolt, [1978](#)), much as as people keep current information ‘in reach’ on a desk (Klein *et al.*, [2004](#)). Spatial approaches recognise that *keeping* is a valuable activity in its own right, that informs sensemaking (Marshall and Jones, [2006](#)). Placed information also performs an important *reminding* function (Barreau, [1995](#); Barreau and Nardi, [1995](#)).

Building on Bush’s ideas of “associative chains of related materials”, *networked* PIM systems focus on the relationships between data. HyperText, as conceived in 1965 (Nelson, [1965](#)) was designed to keep connections between information and allow the computer to understand what linked information *is*. The version of hypertext we use today is much weaker than Nelson’s HyperText or Berners-Lee’s Semantic Web and does not achieve these goals, as the inventors agree (Ross, [2005](#); Nelson, [2006](#); Ziogas, [2020](#)). In the absence of connected networks of personal information and with people collecting more information than they discard (Whittaker and Hirschberg, [2001](#)), the 2000s saw software like Google Desktop Search (‘Google Desktop Search’, [2004](#)) and Infovark (‘Infovark Company Profile’, [2007](#)) emerge to try and discover users’ data files and unify access to them, with limited impact (Bergman *et al.*, [2008](#)). Around this time, Microsoft invented WinFS, a system to re-invent the modern day operating system to be based upon relational structured data rather than file storage, but sadly it was never released (‘WinFS’, [no date](#)). Paul Dourish *et. al.* proposed Placeless Documents, which relied on the idea of assigning user-specific properties to documents so that their could be arranged and recalled by their common properties rather than their location (Dourish *et al.*, [2000](#); Dourish, [2003](#)). *Metadata* – information about what the data *is* – is critical to information organisation (Foulonneau and Riley, [2008](#)). One of the more advanced networked PIM systems is the Networked Semantic Desktop, which recognises that critical metadata is lost when files are copied or emailed, and attempts to maintain metadata and traceability by integrating PIM with peer-to-peer (P2P) technology (Decker and Frank, [2004](#)). Tags, which emerged as a means to organise data through systems like del.icio.us (‘Delicious’, [2003](#)) and Flickr in the 2000s, are still widely used on social media and websites today, and are even available within macOS (Frost, [2019](#)). Tags can be seen as a continuation of attempts to attach metadata to personal data to give it meaning, even though the dream of “folksonomies” has not been fully realised (Abbattista *et al.*, [2007](#); Terdiman, [2008](#)).

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

computers to understand users' mental models (Nadeem and Sauermann, [2007](#)).

While folders have emerged as the dominant means to organise computer files and are effective because they allow you to arrange information according to its meaning to you (Bergman *et al.*, [2012](#); Bergman, [2013](#)), supporters of *temporal* PIM systems argue they are inadequate as an organising device. Freeman and Gelernter proposed Lifestreams, a PIM system based on the principle that storage should be transparent, archiving and compatibility should be automatic, and concise overviews of groups of related information should be available (Freeman and Gelernter, [1996](#)). Central to this system is the idea that personal data can most easily be navigated when viewed as a *timeline*, partly because almost all data can be associated to a specific time, but also because this maps onto the idea of relating personal information to human memory (Lansdale and Edmonds, [1992](#)). TimeSpace provides another model of a PIM system that organises personal information by both time and the user's own activities, to support interaction with a "continuously changing and evolving information space" (Krishnan and Jones, [2005](#)). Time-based PIM approaches also coincide with a drive to move beyond files as a system of information storage. Gelernter believed we should not have to put effort into organising files, and argued somewhat prophetically that commercial factors have skewed personal data systems design away from the realities of human lives (Steinberg, [1997](#)). In my own 2011 article "Why files need to die", I mapped out how a personalised timeline could allow better personal information organisation and retrieval (Bowyer, [2011](#)). Echoing this as well as Decker's desire to maintain an information trail for every piece of information, Siân Lindley *et. al.*, having called for time to become a subject of design research in its own right (Odom *et al.*, [2018](#)), explored the concept of the *file biography*, a concept which allows the history of information to be kept as the file is used and changed. File biographies tell a story, and help to reconfigure our thinking away from mindsets around copying, deleting and sharing, to view information as fluid (Lindley *et al.*, [2018](#)). Moving into the world of online information collaboration, *activity streams* can also be seen as a recognition of the importance of tracking data as it changes, and offer new affordances (Hart-Davidson, Zachry and Spinuzzi, [2012](#)).

In 1995, Barreau highlighted the importance of *context* to PIM; People need access to different information according to what they are doing (Barreau, [1995](#)) In 2000, Abowd and Mynatt highlighted the importance of paying attention to the user's context in order to offer access to the most relevant information and features, and they suggest context can be identified by considering the "5 W's" - who, where, what, when and why (Abowd and Mynatt, [2000](#)). *Context-aware computing* (Abowd *et al.*, [1999](#); Eliasson, Cerratto Pargman and Ramberg, [2009](#)) has subsequently emerged as a sub-discipline of research in its own right (Dey, [2001](#)) (see also section 2.3.2). Dourish identified that context is both a problem of representation, in that it is information that can be captured and represented, and of interaction, in that it is a relational property between objects or activities. He calls for *embodied interaction* - allowing users to create their own practices and meanings in the course of their PIM system interaction, noting that context is not objective and predetermined, it arises from the activity (Dourish, [2004](#)); you need different organisations of information in different contexts. This means that PIM systems need to support representing a given set of information in different ways (Lansdale and Edmonds, [1992](#)) - but more than that, that different information should be shown according to the current context; different *perspectives* are needed to segment your life. TimeSpace uses 'activity workspaces' to achieve this (Krishnan and Jones, [2005](#)), but Karger *et. al.*'s Haystack system refines the concept further, introducing the concept of *lenses*. Perspectives change which information records are included, whereas lenses allow you to focus on different attributes of what might be the same or different information (Karger *et al.*, [2005](#)). Using a similar premise, Jilek's "context spaces" system attempted a dynamically reorganising contextual sidebar, but is limited in flexibility as it uses rigid types for specific contexts (Jilek *et al.*, [2018](#)). Lindley observes that different information abstractions are needed for different audiences, from which we can infer that in a multi-user system, no single arrangement of information will suffice because in the same context two people may have different needs (Lindley *et al.*, [2018](#)).

This is why the sixth trait of PIM systems is important: *subjectivity*. Information organisation cannot be handled in a deterministic, objective manner. Any PIM system must be tailored to, and adaptable by, the user. Shipman and Marshall found that forcing users into explicit information models or workflows is harmful

to user experience, and that interactive systems have to address the challenge of being just explicit enough but still allowing for differences in individual mental models (Shipman and Marshall, [1999](#)). Bergman et al. (Bergman, Beyth-Marom and Nachmias, [2003](#)) proposed three principles for subjective PIM, and their 2003 assertion that these principles are not currently well implemented in PIM systems remains true today:

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

Teevan's take on PIM subjectivity is important: "The user should feel in control of the information". She argues that this can be done by "understanding what *conceptual anchors* the user creates and keeping them constant while the data changes." (Teevan, [2001](#)). With semantic PIM systems, we can see that a successful system (or at least, its designers) must understand a great deal about their users.

2.2.3 Personal Informatics & The Quantified Self

In the late '00s, researchers and enthusiasts took PIM beyond task management and turned PIM thinking toward the self. In pursuit of Bush's vision of augmenting human memory, Jim Gemmell and Gordon Bell in their MyLifeBits project at Microsoft (Gemmell, Bell and Lueder, [2006](#); Bell and Gemmell, [2009](#)) tried to capture an entire life electronically. This became known as *lifelogging*: gathering as much data as possible, so that the maximum possible context, detail and understanding can be gained about that individual. In 2007, tech writers Kevin Kelly and Gary Wolf set out a vision for what they called *the Quantified Self*, that is, to achieve increased self-knowledge through self-tracking, not just of physical metrics such as step counts, heart rates or calories burned, but almost any aspect of your own life that could be numerically recorded in a computer (Kelly and Wolf, [2007](#)). The Quantified Self movement (QSM) is now a world-wide community of enthusiasts who have developed hundreds of tools and techniques for self-tracking/lifelogging and monitoring themselves through data for the purposes of self improvement, and also has a non-profit organisation aiming to 'advance discovery through increasing access to data' ('About The Quantified Self', [no date](#)). Around 2009, researcher Ian Li began writing about what he called *personal informatics*, noting that it can be difficult to know ourselves due to incomplete self-knowledge, difficulties in monitor our own behaviours, and being too busy to introspect. He proposes that "Computers can help: They can store large amounts of data, analyse the data for patterns, visualise the data, and provide feedback at opportune times (Li, [2009](#)). Just as QSM has gained traction with enthusiasts in the general public, so personal informatics has grown as an area of research, development and study in academic circles. While QSM and lifelogging focus slightly more on capturing data about oneself and personal informatics focuses slightly more on the mechanisms of integrating and reviewing self-tracking data, there is so much overlap that all three can be considered the same field, which for convenience I will refer to by the shorthand *self informatics* (SI) throughout this thesis. SI can be seen as a distinct advancement from PIM because of its focus on *using* personal information for personal benefit. SI can be seen as the antithesis of corporate data-centric motives outlined in 2.1 - as here, data is gathered for the data subject's benefit rather than that of the data-gathering organisation.

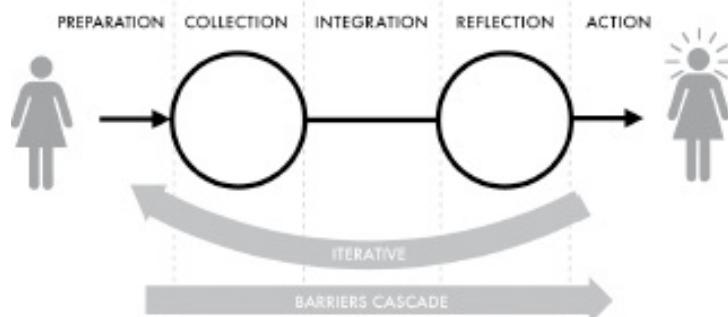


Figure 2: Li et. al.'s Stage-based Model of Personal Informatics Systems (Li, Dey and Forlizzi, 2010)

Li, Dey and Forlizzi conducted participatory research with SI practitioners and identified five stages of personal informatics systems (which can be seen as refinement of William Jones' list (W. Jones, 2011b) of the six activities involved in PIM). The five stages, illustrated in Figure 2, each of which can be driven by the user, the SI system or both, are:

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

Of these, reflection is perhaps the most important, as the capacity to gain new insight is the motivating reason to engage in SI. Reflective learning (Boud, Keogh and Walker, 1985) has been recognised as a valuable means of knowledge acquisition and improvement in a variety of contexts including education (Dewey, 1938), business (Beck *et al.*, 2001), and research (Lewin, 1946). In the context of the wisdom curve (see Figure 1 above), reflection can be seen as asking questions of data in order to acquire knowledge about oneself. Knowledge about oneself (a.k.a. self-insight (Hixon and Swann, 1993)) serves not only to satisfy curiosity (Li, Dey and Forlizzi, 2010) but can improve self-control (O'Donoghue and Rabin, 2001), increase self-awareness (Aslam *et al.*, 2016) and enable positive behaviours such as saving energy (Seligman and Darley, 1976).

Reflection can be facilitated in SI systems by enabling the tracking of subjective factors such as mood, health or activity, and can be triggered by means of notifications, or during more direct information exploration by the user as they recall or revisit experiences (Rivera-Pelayo *et al.*, 2012). To aid interpretation of data by SI users, *contextualisation*, enhancing information with additional facts to ease its comprehension. This can include social, spatial or historical context, subjective or objective metadata or external sources of information (e.g. weather) (Rivera-Pelayo *et al.*, 2012), or external devices (Dey, 2000). There are two phases of reflection, *discovery* and *maintenance*. During the initial discovery phase, typical questions that SI users ask concern the *history* of data changes, understanding the *context* of a datapoint, the *factors* that cause a pattern in data, and the identification of suitable *goals* to pursue. During the maintenance phase, these goals frame the questions asked, which concern *status* (how well you are doing at meeting your goals) and *discrepancies* (examining the difference between actual behaviour and desired behaviours).

In order for a SI user to successfully reach this maintenance phase where they can continue to reflect upon their actions and adjust their goals, they must have been able to successfully navigate each of the 5 phases illustrated in Figure 2; if they have not collected the right data, they cannot integrate it, if they have not been able to integrate the collected data in a meaningful way, they cannot reflect upon it, and so on. Li et. al. framed this the *barriers cascade* (Li, Dey and Forlizzi, 2010), and the pursuit of new ways to overcome these barriers has in effect been the major problem space for all SI approaches; this is especially evident in

the QSM (Choe *et al.*, 2014). While effortless SI is not yet a reality and many barriers still exist, progress in easing the SI journey through the barriers cascade is being made: in 2011, Jones had noted that people often postpone or don't have time for meta-level information management activities (W. Jones, 2011a), but by 2019 the increased automation around self-tracking and data collection was judged to have given people more free time and energy for reflection and managing their goals (Feng and Agosto, 2019).

2.2.4 The Emergence of Complex Digital Lives

As described in 2.1.2 above, the rise of data-centrism has meant that every aspect of our lives now involves digital service providers and products which process personal data. Smartphones put computers in everyone's pockets, and cheap cloud computing and an open web allowed every organisation to serve the population digitally through apps and websites. In 2010, broadband access was declared a legal right in Finland ('Finland: Broadband Access Made Legal Right In Landmark Law', 2010), and in 2011, the UK Supreme Court declared that Internet access was an "essential part of everyday living" and denial of Internet access for criminals such as sex offenders was ruled unlawful (Roche, 2011; Wagner, 2012). Everyone now required access to information and online digital services. "The boundary between real life and online [had] disappeared" (Burkeman, 2011). The promise that whatever you want to do "there's an app for that" had become true (Apple, 2009). During the late '00s and throughout the 2010s data-centric companies disrupted almost every industry: Amazon (shopping & books), Uber (taxis), Netflix (movie rental), Spotify (music), AirBNB (accommodation), Google (email, news & advertising), Facebook (social networking & advertising), Paypal/Revolut/Monzo (banking), match/Tinder (dating), Steam (computer games), Just Eat (takeaways), and many more (Levine, 2011; Carter, 2015). As a result, we now produce rich data trails simply by going about our daily lives, and this has become "the driving force for value creation" online (Symons *et al.*, 2017). More recently as we start the 2020s, the trend has accelerated, with the COVID-19 pandemic necessitating the move of both information work and social activities to online using platforms such as Zoom, Google Docs and miro (O'Donnell, 2020).

Throughout the transition to this information economy, the computing industry has delivered revolutionary new capabilities, but with every provider offering their own apps and websites, the information landscape has become hugely challenging for people to manage; information overload is now a serious problem that has been linked to increased anxiety, impaired critical thinking, exhaustion, and loss of willpower and focus (Hemp, 2009; Tunikova, 2018; Fu *et al.*, 2020). Our personal information is fragmented and a unified interface is needed: "We must launch multiple applications and perform numerous repetitive searches for relevant information, to say nothing of deciding which applications to look in (Karger and Jones, 2006)." In the silo-ed world of today's Internet, this has only got worse. Bergman's subjective principles (see above) imply that our data should be able to move and be referenced freely, but it cannot. Our ability to share and connect data is limited (Crabtree and Tolmie, 2018). Our data is trapped (Abitbol, André and Kaplan, 2015), not only because it is held by organisations without giving us effective access, but also by various practical means such as format incompatibilities, device restrictions, paywalls, and a lack of data portability. We need to free our data, as I expand upon (Bowyer, 2018).

It is clear that general-purpose computing has yet to provide people with the tools to manage their complex digital lives. There have been attempts to create general purpose interfaces for personal data, typically based around a timeline, such AllOfMe.com ('AllofMe Company Profile', 2007; 'AllofMe.com Teaser Clip', 2008) in 2008 and more myTimeline a decade later ('myTimeline', 2018), however none of these products have reached public availability. To date the closest market-successful tool that people have for general purpose information handling is Facebook, given that it can store personal information, handle asynchronous and instant messaging, news, photo sharing, some retail functionality, brand interaction & support, calendaring and event management, and group discussions. However, it is a closed system with no capability for customisation; none of its content is available outside the network and external content cannot be linked or interacted with except by import; as such it cannot be considered a PIM system. Its own Timeline feature, promoted at launch in 2011 as "the story of your life" and "a new way to express who you are" (Siegler, 2011) has been retired, along with many other tools designed to make information easier to

manage such as personal news feeds and friend lists (Perez, [2018](#)), a reminder that Facebook exists primarily to serve its advertisers, rather than the general public, as per the often-repeated saying “if you’re not paying for it, you are the product”. The most promising area for the development of interfaces for managing digital lives is the emerging “personal data locker” space, explored more in 2.3.4 below, which offer the promise of “a place for personal data”, as Jones imagined PIM should be (W. Jones, [2011a](#)), though these are still quite limited. As Abiteboul noted in 2015, “everyone should be able to manage their personal data with a personal information management system” (Abiteboul, André and Kaplan, [2015](#)), but as of yet, in any meaningful or holistic way, they cannot, because no general-purpose personal information management system for modern day digital lives exists.

2.2.5 Research Gap: The Data Beyond The Individual

In this section, I have detailed the ways in which personal information management systems have developed, and shown that they have not kept pace with the ever-more-complex needs of the Information Age. Most PIM systems treat data as a static resource to be filed and accessed much like you would a file in a 1970s office. Most digital services operate in isolation from each other, without any shared representation or co-operative understanding of an individual’s personal information. Where personal data access is provided, it is limited in usage to the delivery of the specific service on offer, it is treated as a property asset and the data is not *participatory*. As Katie Shilton writes, “Much of the social impact of participatory personal data will depend on how data are captured and organized; who has access; whether individuals consent and participate; and how (or whether) data are curated and preserved (Shilton, [2011](#)).” We need “fundamental changes in the way we represent and manipulate data” (Karger and Jones, [2006](#)); we need holistic representations of data that can be subjectively meaningful and which allow for the constant change and evolution of data over time.

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

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

2.3 Practical Human-centred Design

2.3.1 Human Computer Interaction Foundations

Up until the 1980s, the only reasons to consider the relationship between a human and the computer they were using were ergonomics, comfort and efficiency. People were shielded from the complexities of the machines they were using—the machine did the work and the human was just the *operator*. In the 1990s, the “first wave” of what is now known as *Human-Computer Interaction* (HCI) recognised humans as *actors* operating in groups, who had tasks to perform either using or assisted by technology (Bannon, [1995](#)). People were now *users* of technology. Design thinking shifted from machine-centric to *user-centric design* (UCD), motivated by the goal of helping the user to do their tasks better. In the personal computer revolution of the 1990s, people began to work in complex and varied multi-user situations, and observation and understanding of a user’s working environment provided empathy that enabled better design. There was a recognition that people use computers differently in different contexts. In the 2000s, as smartphones,

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

2.3.2 Data Transcendence, Context and Human-Data Interaction

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

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

2.3.3 Human-Centred Design: A Sociotechnical Challenge

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

Rogers (Rogers, 2006), a more humane interface is needed (Raskin, 2000), one which designs for the whole person. Judging the success of a user interaction can no longer be done by assessing task-completion efficiency (Abowd and Mynatt, 2000) but should consider the holistic needs of the individual at that moment in time.

Yet in the 2010s, there was a growing recognition that the world had lurched severely away from such goals. The design of information-consumption interfaces was having a detrimental effect upon people, not just in terms of the psychological impacts of information overload as detailed above in section 2.2.4, but also in terms of the impact on users' attention. This would become known as "the attention economy" (Croll, 2009; Cogran and Kinsley, 2012; Brynjolfsson and Oh, 2012). Social media technologies like infinite scrolling and smartphone notifications had created "a culture of perpetual distraction" (Timely, 2020) which "hijacks people's minds" (Harris, 2016). As Zeynep Tufekci put it in her TED talk, "we are creating a dystopia just to make people click on ads" (Tufekci, 2017). In 2013, Tristan Harris released a presentation calling on the tech industry to respect users' attention and minimize distraction (Harris, 2013a), which lead to the creation of the Center for Humane Technology (Harris, 2013b), a central group in this new movement to design for positive human values and to practice *value-sensitive design* (Friedman and Hendry, 2019). This focus beyond just supporting data interaction to understanding and enhancing the individual's lived experience can be seen as a central guiding tenet of *Human-centred design*.

We can see from the above that the design of human-centred personal data interaction is not purely a matter of designing better user interfaces, nor even of designing for the user's physical environment, but in fact a design challenge that exists at the sociotechnical (Bunge, 1999; Murton, 2011) level – it must take into account the social relationships of the individual (as detailed in 2.2.6) as well as the power imbalance that exists between data holders and data subjects (as detailed in 2.1.2). Andy Crabtree recognised the sociotechnical nature of the HDI challenge in his 2016 paper with Mortier on 'The Shifting Locus of Agency and Control' and highlighted particular aspects of this multi-party challenge around personal data, specifically being able to ensure the *privacy* of your data as well as the *accountability* data subjects require over data-processing algorithms and data-handling organisations (Crabtree and Mortier, 2016). These goals are now actively pursued through research into *privacy by design* (Cavoukian, 2010) and *Critical Algorithm Studies* (Gillespie and Seaver, 2016) respectively. In his subsequent work with Peter Tolmie, Crabtree focused on the particular HDI challenges around data-sharing, which must also be designed for (echoing Lindley's work on file biographies mentioned earlier) (Crabtree and Tolmie, 2018). These areas of pursuing a human-centric agenda within a sociotechnical context continue to be areas of active research today, as seen in projects such as Nesta's DECODE (Symons *et al.*, 2017), which focuses on individual empowerment, and UKRI's not-equal.tech (Crivellaro *et al.*, 2019), which focuses on *data justice* (Taylor, 2017).

2.3.4 The Emergent Human-Centred Personal Data Ecosystem

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

Taking a more technical slant on similar ideas, David Siegel outlined a vision of a personal data interface that would allow the ideas of VRM to be realised. He called this a *Personal Data Locker*, though the equivalent terms Personal Data Store, Personal Data Vault (PDV) and Personal Data Services are also

used. The concept is explained in his book (Siegel, [2010](#)) and video (Siegel, [2009](#)). He also coined the term *Pull-centric Computing* (where information is ‘pulled’ at your request rather than being pushed upon you). The WEF’s Rethinking Personal Data project (mentioned earlier) describes the potential for a *personal data ecosystem (PDE)* of “commercial entities, acting as trusted intermediaries, exchanging assets on behalf of individuals, following a clear set of principles and legally binding contracts” with the PDV being the technical means to place the individual at the centre of that ecosystem, the PDV provider would be “an intermediary collecting user data and giving third parties access to this data in line with individual users’ specifications” (Hoffman, [2010](#)). A 2010 report by nonprofit Mydex helps to contextualise the PDV, explaining that the PDV is a service to the individual that positions “individuals as information managers” at the “epicenter of a new ecosystem of PIM services” and that it will not just give access to data but “transform relationships between individuals and organisations” (Mydex CIC, [2010](#)); this to me is what substantially differentiates the PDE from traditional PIM systems - it is a response to the sociotechnical need outlined in the previous section. A 2012 report from Ontario’s Information Privacy Commissioner notes that the PDE collides with traditional concepts of ownership when it comes to data, that the PDE needs to “provide a collection of tools and initiatives aimed at facilitating individual control over personal information” wherever it is located; this is another way in which PIM within PDE can be differentiated from traditional PIM (Cavoukian, [2012](#)).

It was against this landscape that *Personal Information Management Services* (PIMS¹) became a business area in its own right, the basis for a *personal data economy*. PIMS is attempting to create a market for “tools that help individuals gather, manage and use personal information to make better decisions and manage their lives better”, with a potential market value (in the UK) of £16.5 billion, more than the automotive and pharmaceutical industries (Ctrl-Shift, [2014](#)). In 2016, a global network and non-profit initiative called MyData was founded, bringing together researchers, companies and public sector agencies in the PDE space, in pursuit of a “fair, sustainable and prosperous digital society, where the sharing of personal data is based on trust, and relationships between individuals and organisations are balanced” (MyData.org, [2018](#)). An important aspect of MyData is its aim to combine companies’ needs for data with individuals’ digital human rights. Through analysis of principles of PIMS, VRM and other related spaces (‘MyData Comparison of Principles document’, [2017](#)), the MyData declaration was produced, outlining a detailed vision for the PDE space to “empower individuals with their personal data, thus helping them and their communities develop knowledge, make informed decisions, and interact more consciously and efficiently with each other as well as with organisations.” (MyData, [2017](#)) MyData now has over 700 parties involved worldwide and provides a focal point to the PDE community.

The MyData declaration identifies data controllers’ transparency with data and data-handling practices as an essential means for individuals to gain agency and accountability, and puts forward the idea that the individual should be the point of integration of their own personal data ecosystem; in other words, “everything goes through me”; this is the embodiment of the human-centric ideal of individual empowerment but will also be a good way for data controllers to ensure awareness, accuracy and consent. They also introduce the idea of a *personal data operator* (also known as a *data trust*) which is a key part of the personal data ecosystem - a trusted third party which stores or transfers data on behalf of the data subject, but does not use it themselves. An example operator is digi.me, which has developed a PDV with a “private sharing” model that allows users to allow subsets of their data to be used by external organisations or apps with strictly controlled parameters (Firth, [2019](#)). The MyData/PDE space is very active currently, with many emerging businesses and startups having appeared in the last two to three years. Citizen.me (‘Our Values’, [no date](#)) is another company with a similar positioning. Other operators such as UBDI (‘Whose data is it anyway?’, [2019](#)) and datacy (‘About Us’, [no date](#)) are positioned under a different business model which aims to help individuals take control of their personal data for profit. Open Humans has a PDV optimised to allow people to share their data for the benefit of research (Price Ball, [no date](#)). Ethi is a PDV platform focused on providing individuals with deep insights from their data, and tools to more easily delete your personal data from data-holding organisations (Jelly, [2021](#)).

2.3.5 Research Gap: Defining the Research Agenda for Achieving Human-Centricity in Practice

In this section, I have shown how the emergent human-centric personal data ecosystem has developed from its roots in HCI, ubicomp and HDI. The call for designs and sociotechnical systems that empower individuals with their personal data arise from the power imbalance (Hoffman, [2014a](#)) that has emerged as a result of the datafication of modern life. In the third wave of HCI (Bødker, [2015](#)), user interface design's main consideration was "what does *the user* want to do?". Over the last decade, catalysed by the shift by the explosion of Internet culture and the shift from self-install software products to massive-scale cloud-based Internet services, there has been a gradual but perceptible shift away from the tenet that the user's needs should come first: the designs of commercial and civic web applications now more reflect the question (considered from the provider's perspective) "What do we want the user to do?". Users (people) and their individual needs have been left behind. The MyData community have clearly outlined the goals to address this problem, but much of the focus at present is on technology questions of how to build better PDVs and better PIM interfaces, or on indentifying an effective business model that will facilitate the transition to a PDE, which is a necessary but distracting question. My research is situated at the bleeding edge of this emerging human-centric personal data ecosystem and being non-commercial, is able to take a more purist human-centric stance. After uncovering the human experience of personal data (as detailed in 2.1.5) and the lived experience of personal data usage within people's wider digital life and relationships (2.2.5), I will seek to address a third research gap - to understand the technical, legal, policy, economic and social realities of the PDE landscape itself, sufficient to inform the design of PDE processes and systems. Thinking of the barriers cascade in the SI space (Li, Dey and Forlizzi, [2010](#)), what barriers exist that inhibit the building or adoption of PDE human-centric technologies? What opportunities might make it easier to overcome these barriers and to catalyse progress toward the human-centric agenda as envisioned in the MyData declaration? What are the key challenges faced when we attempt to build human-centric technologies in today's world? By applying learnings about human experiences and attitudes to the data-centric world to the practice of PDE design & development, can we more clearly map the road ahead and define a research agenda for the next step of tackling the PDE challenge?

By adopting both a participatory design and technical strategist's standpoint throughout this thesis, building on the theoretical foundations of effective data access, information management and human-centric data interaction, I aim to progress PDE / MyData thinking, using methods detailed in the next chapter, in pursuit of my primary research question, which is:

"What role should people's data play in their lives, what capabilities do they need, and how could these ideals be achieved?"

3 Methodology

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.

3.1 Forming a Research Paradigm: Ontology & Epistemology

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 through 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.

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

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 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).

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.

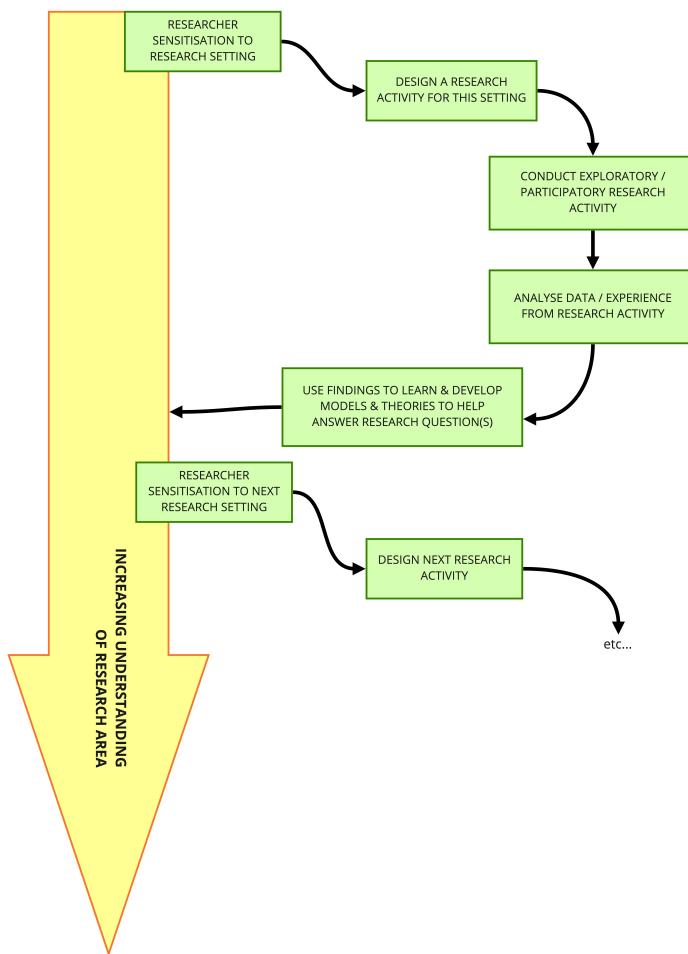


Figure 3: My action research approach

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 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.

3.3 Research Objectives

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

“What role should people’s data play in their lives, what capabilities do they need, and how could these ideals be achieved?”

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:

3.3.1 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.

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

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 socio-technical 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-centrality? 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.

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

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, and the individual desires and needs we uncover in RQ1 and RQ2, might be achieved in practice. This involves gaining understanding of the technical, economic, political and legal landscape that personal data interaction occurs within. This involves 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, Dey and Forlizzi, [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, 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. 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. Unlike RQ1 and RQ2, this strand of research will be explored not through strictly configured study research engagements but rather through a process of acculturation to the world of building data systems and developing my own knowledge through design, technical prototyping and pushing the boundaries of the systems that do exist so that they may be better understood. Ultimately these insights should allow me to achieve greater expertise, backed by the empirical findings from RQ1 & RQ2, to allow me to draw conclusions about how I believe the discipline of human-centred data relations should proceed in its future research and development.

3.4 Overview of Research Contexts and Activities

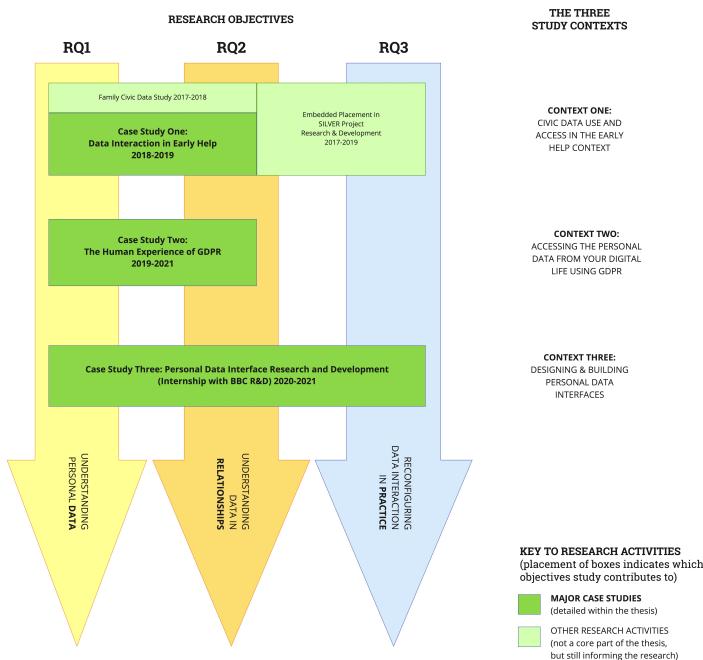


Figure 4: Research Activities and Contexts

As explained in the last section, the three sub-research questions RQ1, RQ2 and 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, and RQ3 as understanding how to reconfigure data interaction in practice. Figure 4 also illustrates how the three contexts of study and three major case studies, 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.

3.4.1 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:

3.4.1.1 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, 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 includes a short section [ADD REF TO CHAPTER 7 SUBSECTION] detailing my high level observations from participating in the project. The final report from the project is available at [ADD REF HERE WHEN AVAILABLE].

3.4.1.2 Understanding Family Civic Data study

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].

3.4.1.3 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.

3.4.2 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) but also crucially it enabled me to become a participant in my own research, enabling a deeper understanding of this research context.

Within this context, I carried out four research activities between 2016 and 2020:

3.4.2.1 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].

3.4.2.2 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].

3.4.2.3 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 to 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). A summary of observations and findings are presented in [INSERT APPENDIX REFERENCE HERE].

3.4.2.4 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.

3.4.3 Context Three: Designing and Building Personal Data interfaces

The third context for this PhD, which has remained a 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 designing and building of new interfaces.

In total four separate research activities between 2017 and 2021 took place within this practical research context:

3.4.3.1 Health Interface Development in the CHC 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. This aspect of the SILVER project is considered out of scope for this thesis, though reference is made to it in Chapter 7.

3.4.3.2 Reconfiguring Data Interfaces and Obtaining Data through Web Augmentation

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].

3.4.3.3 CASE STUDY THREE: Research Internship with BBC R&D Cornmarket Project

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. 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.

This case study is described in detail as Chapter 7 of this thesis.

3.5 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

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.

3.5.1 Sensitisation of Researcher and Participants

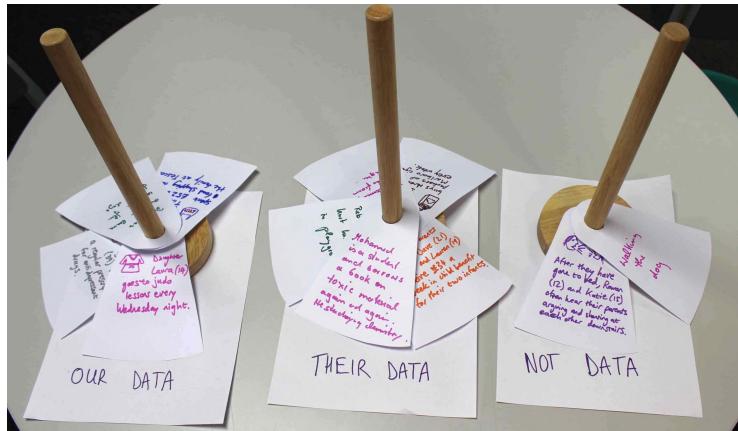


Figure 5: “Family Facts” – What is Data?

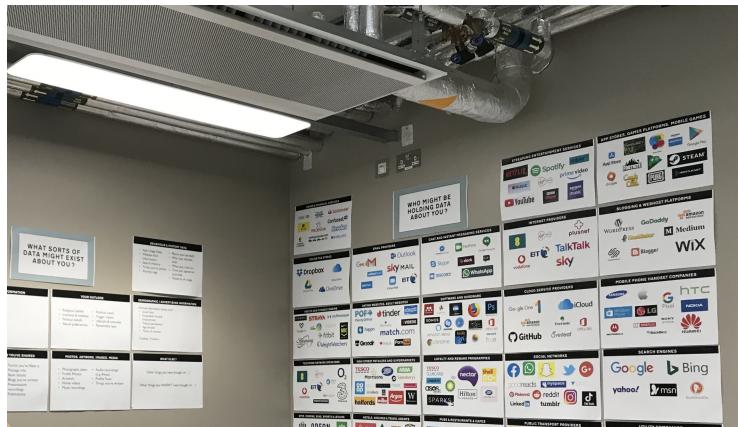


Figure 6: Walls of Data – Sensitising participants to the world of commercially-held data and GDPR

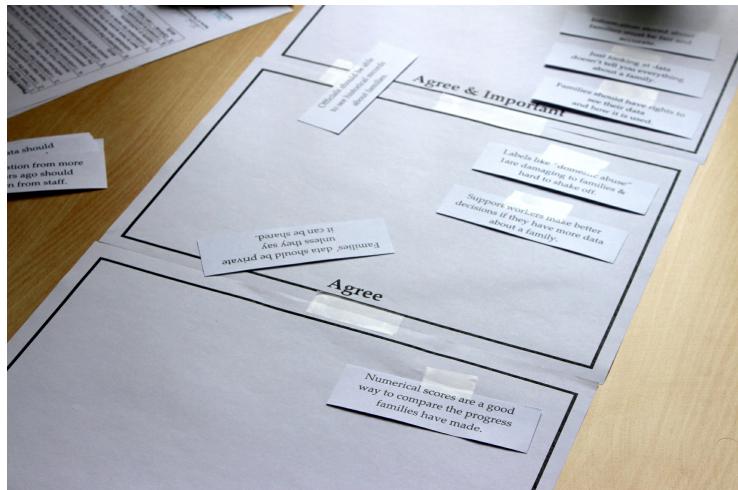


Figure 7: Sentence Ranking – Bringing staff and families to a shared problem space

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 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.

3.5.2 Discussion and Exploration with Stimuli



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



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

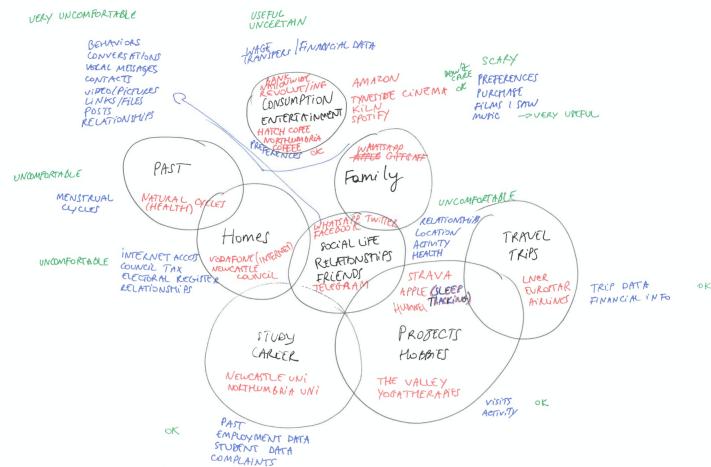


Figure 10: Sketching Dialogue – An example life sketch created by a participant and annotated during discussion

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, 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 Messeseter, 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 and 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 and M. 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.

3.5.3 Participatory Co-Design of Possible Solutions



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

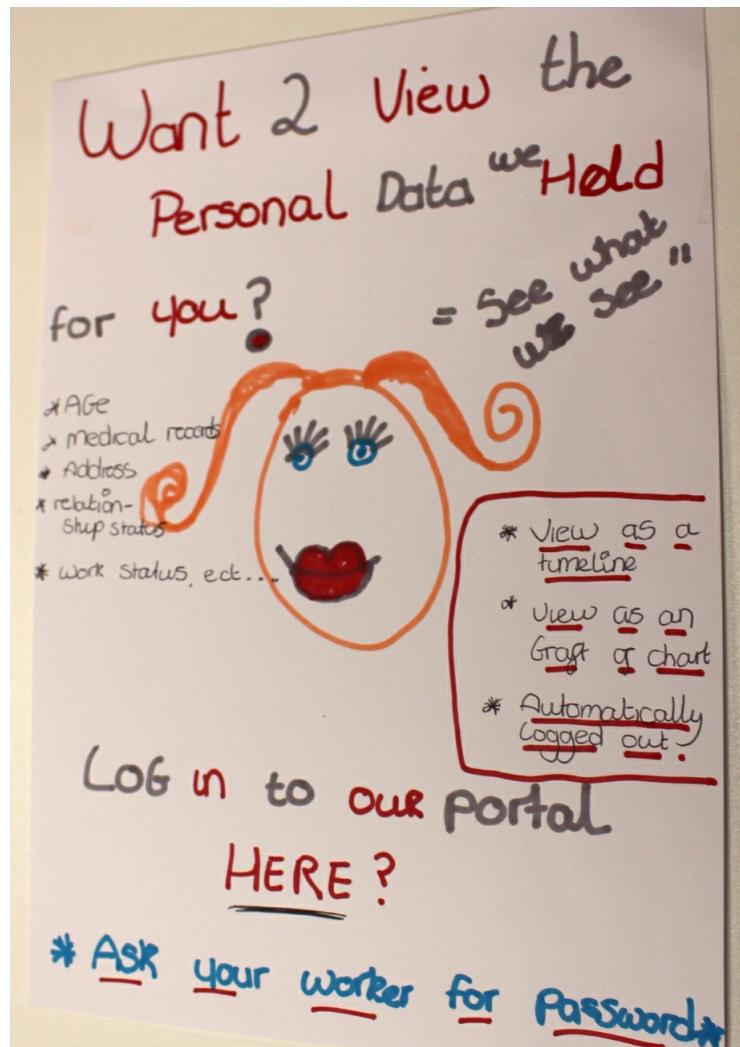


Figure 12: Group poster design – A participant-designed poster to advertise features of imagined data interface products



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

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 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.

3.5.4 Practical Data Experiments, Interface Design and Prototyping



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

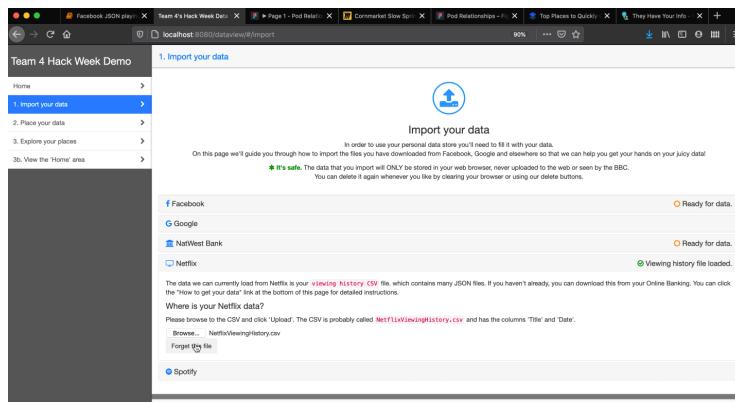


Figure 15: Prototype interface for GDPR Data Viewing – A working prototype that I developed during a hack week at BBC R&D

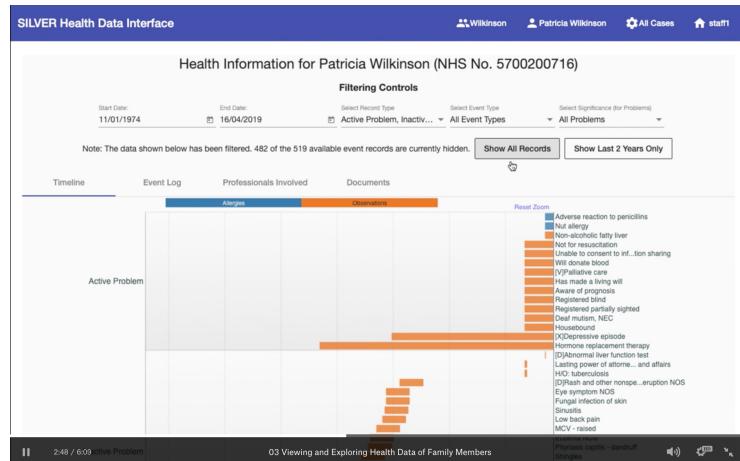


Figure 16: SILVER Health Data Viewing Interface – A working health data viewing interface for Early Help support workers that I developed as part of the SILVER project

In Case Study Three 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 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.

In Case Study Three, 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, but also provided me with copies of my own personal data. Within Case Study Three, 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.

3.5.5 Analysis, Modelling and Learning

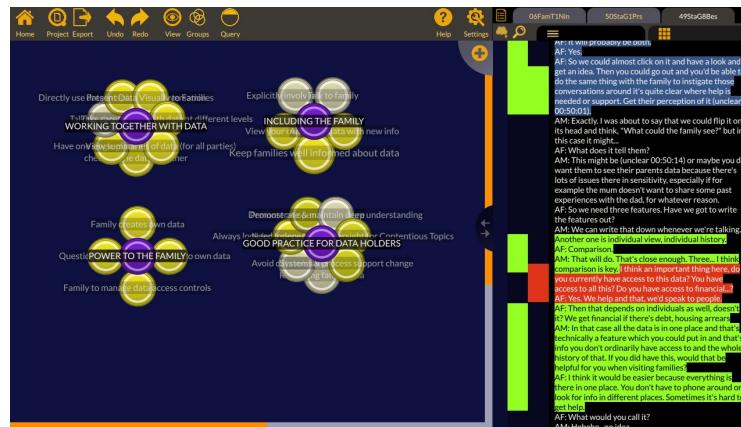


Figure 17: Thematic Analysis – A screenshot of thematic coding of qualitative data using Quirkos for Case Study One

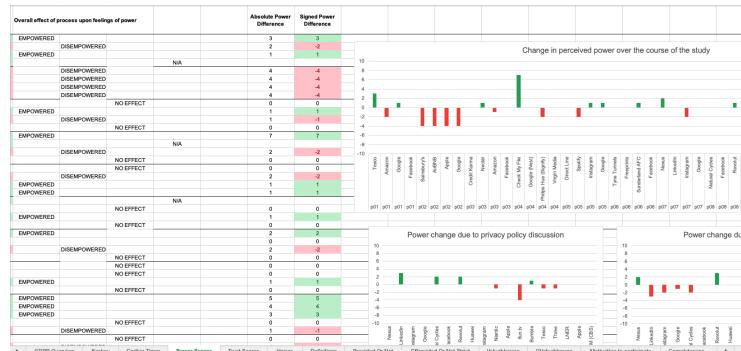


Figure 18: Quantitative Analysis – A screenshot of spreadsheet-based quantitative analysis of interview data from Case Study Two

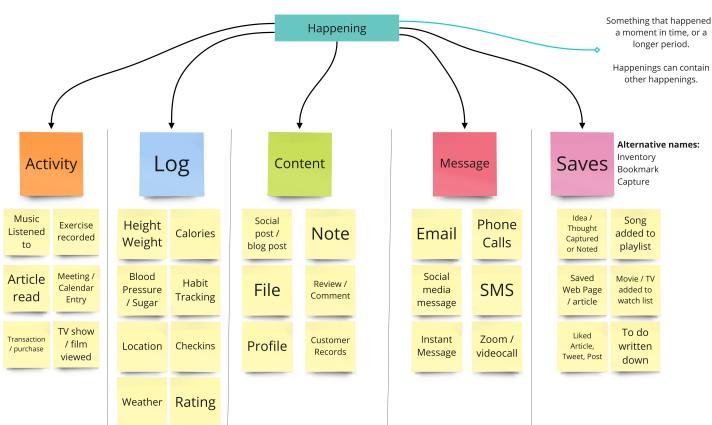


Figure 19: A Model for Personal Data – Developing a common model for personal data imported into a PDV a part of Case Study Three

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 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 Case Study Three, 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 PDV context while embedded at BBC R&D.

3.5.6 Recruitment

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

Research Activity	Engagement	Stage or Phase	Duration	Number of Participants	Recruitment Method
Understanding 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 support workers & related staff	Selected by Local Authority Care Services

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

Tables 1 and 2 summarises the participants involved in this research². 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 is 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

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

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

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

Research Activity	Engagement	Stage or Phase	Duration	Number of Participants	Recruitment Method
Smartphone Usefulness study	3 x 1-on-1 interview	preliminary	3 x 45 minutes	3 adults	Convenience sample
Digital Life Mapping study	5 x 1-on-1 interview	preliminary	5 x 2 hours	5 adults	Convenience sample
Main study (Guided GDPR)	11 x 1-on-1 interview (Life Sketching)	1	11 x 1 hour	11 adults	Convenience sample

Research Activity	Engagement	Stage or Phase	Duration	Number of Participants	Recruitment Method
Main study (Guided GDPR)	10 x 1-on-1 interview (Privacy Policy Reviewing)	2	10 x 1 hour	10 adults	Continuation from previous stage ³
Main study (Guided GDPR)	10 x 1-on-1 interview (Viewing GDPR returned data)	3	10 x 2 hours	10 adults	Continuation from previous stage

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 equally. 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 cases⁴ 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.

3.5.7 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.

3.5.8 Thesis structure approach

In writing up this thesis, I made a choice to foreground my three 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 each span two research questions (RQ1 and RQ2 - see Figure 4 in section 3.4) as they explore both people's relationship with data *and* the relationships people have that involve data. Case Study Three maps directly to RQ3, and is focused on designing human data relations in practice.

Because of the overlapping RQs 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 a short 'bridge' chapter which separately unifies RQ1 findings and RQ2 findings, so that they can be referenced in general terms as an understanding of people's experience of data (RQ1) and of relationships that use data (RQ2) without needing specific details from the first two case studies.
- Chapter 7 details Case Study Three in the context of RQ3, and drawing on the unified RQ1 and RQ2 findings in Chapter 6.
- Chapter 8 is the discussion chapter, covering all three research subquestions and the main research question.

4 Case Study One: Accessing and Using Civic Data in Early Help

In this chapter, I describe the first major case study of this PhD, in which I ran three 2 hour participatory co-design workshops involving local authority support workers and parents and children from supported families that had recently participated in Early Help programmes. The purpose of the research was to build upon prior explorations to gain deeper understanding of family and staff attitudes to civic data holding (in pursuit of RQ1) and to move beyond this and explore the role of data within the support relationship (in pursuit of RQ2). A particular area that I explored was to consider the possibility of shared data interaction, where supported families and their support workers would interact with data together and in person as part of the support engagement.

In section 4.1, I will provide background on the Early Help context in England. In 4.2[CHECK], I will review the prior findings from my own preliminary studies as well as that of others including Connected Health Cities, and show how these findings were used to establish a common ground within the sensitisation activities at the start of each workshop. In 4.3, I will describe the three themes discovered through

qualitative analysis: that families want to be given a voice (4.3.1), that trust can be earned through data and process transparency (4.3.2), and introduce the concept of meaningful data interaction for families (4.3.3). In section 4.4, I will discuss these findings in the context of prior literature, drawing insights into the value of involving people with their data (4.4.1), the need for human interaction to make data interaction effective (4.4.2), and the pros and cons of the shifting of the locus of decision-making towards the family that shared data interaction would bring about. In 4.5, I will summarise the case study in terms of how these insights expand our understanding of the research questions and their wider significance.

4.1 Context: Data Use in Early Help

4.1.1 Data-centric Family Intervention in UK Social Care

In the UK, the social care system been shaped by a history of efforts, initially under the *Every Child Matters* policy programme [REF], to improve the lives of children, especially those suffering the most. The *Contact Point* and *Common Assessment Framework (CAF)* programmes, were established with the aim to create universal digital tools to support co-ordination at a local level across public sector services, centred around around children and young people [REF Wilson et al 2011, Cornford et al 2013], later expanding to include their families [REF Malomo and Sena, 2017]. A change of government in 2010 saw many of the policies around children and families moved from a basis of universal access to a targeted provision. Programmes such as *Think Family* [REF Cornford et al 2013, Crossley XX] introduced a focus on *family intervention* as a primary approach; social workers learn about and get directly involved with the lives of targeted young people and their families in order to understand problems and to help empower them to overcome specific difficulties they face. The Troubled Families Programme (TFP), created in 2012 for England, was built upon a claim that £9 billion of civic spending was due to just 120,000 families and that a net saving of £11,000 could be achieved for each family that could be “turned around”. Local municipalities were required to work with partner agencies to identify *troubled families*⁵ – those ‘at risk’ families experiencing multiple issues from a list including unemployment, overcrowded housing, poor education, mental health issues, disability, low income, poverty, truancy, crime and domestic violence – and to work with such families to reduce these risk factors for them (Bate and Bellis, 2018). The TFP was set up in such a way that local authorities could claim central government funding for each family they had provably ‘turned around’, and as such encouraged extensive collection and use of data about each supported family to track and demonstrate progress and impact. This shift towards using data mirrors the societal rise of data-centrism described in section 2.1, but was also being seen across the public sector; under increasing pressure to demonstrate performance and deliver measurable, consistent results, all human services (including social care, health care and education) have become adept in the collection and use of data about their clients or service users. The use of data by the state as a means to represent and think about families is considered problematic (REF Cornford et al., 2013, Neves and Casimira 2018). For instance, from the perspective of the state, such data may include both objective facts from families’ lives such as address or family inter-relationships, as well as potentially more subjective information such practitioners’ observations or numerically-quantified measurements of risk. The risk of inaccurate data or unfair judgement is compounded by the fact that the clients of such services typically have limited access to this data. Although in theory families retain the ability to interact with services (and have access rights to data) the practitioners and the organisations for which they work become de-facto *gatekeepers* to the data about a family (REF). This is then played out in a policy context where data-driven approaches to family care are encouraged through policy and reports about improving quality of the sector (Bate and Bellis, 2018; Department for Education, 2018; Field, 2010; OFSTED, 2015).

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

Over the last decade, *Early Help* programmes have become a key social care offering from almost all local authorities. These programmes seek to pre-emptively help individual residents voluntarily before statutory intervention is needed. Early Help was quickly identified as a suitable setting to explore the use of *family civic data* (a term I introduce in (Bowyer et al., 2018)) and its impact on individuals in this data-centric policy context. Connected Health Cities’ SILVER project, a Department of Health and Social Care funded project

working across five local authority areas in North East England, aimed to improve Early Help support through improved use of family civic data. Through my embedded collaboration within this project, existing use of families' civic data by early help practitioners and front-line support workers was possible.

The need to produce data for use as evidence for schemes like the TFP led local authorities to update their Early Help processes; support workers would now carry out an 'early help assessment' (a guided enrolment questionnaire) to create an 'early help record' (EHR) for each supported individual and their family, which is then stored in a case management system such as CareFirst, LiquidLogic or eCAF. To help form a holistic perspective of a supported family's situation, a process of information gathering and family-centric inter-agency collaboration is adopted. The EHR is supplemented by data from other agencies reporting on an ad hoc or periodic basis (e.g. via emailed spreadsheets, phone conversations, and in-person meetings, such as the *Team Around the Family* (TAF) – a bespoke grouping with representatives from other agencies such as police, schools or housing agencies. This data is used to evaluate that family's situation and progress against the 'Common Assessment Framework' (REF). Support workers are encouraged to use data as evidence at all stages.

An Ofsted report into UK early help in 2015 found that early help services across the UK were too inconsistent and recommended that greater standardisation in assessment and evidence-based practice were needed. Consequently, Early Help schemes continue to seek more data about 'at risk' individuals to use as evidence and to inform their care. Support workers, if provided with better data, can in theory make better decisions as part of the care they provide, and this belief that the best evidence is data is baked into national policies: "IT systems are most valuable when practitioners use the shared [between agencies] data to make more informed decisions about how to support and safeguard a child." (Department for Education, 2018). Such central policies highlight that in the UK, early help work is a data-driven service.

Despite this policy goal, the technical reality has been far more complex. Many different IT systems are used for social care, even within the same local authority; teams work in isolation using different systems and applications. The information ecosystem that the care services fit within is vastly complex (Copeland, 2015) with each part of the system having its own ICT systems and limited arrangements being in place to facilitate information sharing across the different data-holding authorities (which sometimes include local charities with their own ICT systems to which care functions are sometimes outsourced). The existence of different administrative boundaries for different authorities and agencies further complicates the situation. This fragmented ecosystem has proliferated due to each local authority being responsible for procuring their own IT systems in the absence (despite recommendations (Harbird, 2006)) of any centralised systems or information sharing standards.

The reality of information sharing in this context today is that many barriers exist – for example care workers can rarely access health data from GPs and have to rely on school nurses, health visitors, specialists or the individual's own account. Where such information is shared, it is often in the form of emailed spreadsheets or reports, telephone conversations or committee discussions, and not supported by technical integration. No one team, agency or authority can have a full picture of an individual's data (Malomo and Sena, 2017). Different operating policies, consent agreements, privacy regulations, technical access levels, system functions and staff competences result in different interpretations and limitations about what data can be shared (Malomo and Sena, 2017). Data should flow freely through the system in the service of individual care, but it does not, the public sector has a closed and fragmented ecosystem (Pollock, 2011).

Processes such as TAF meetings and the attempt to unify all information onto a single EHR can be seen as a recognition of this failure in the system to produce a single source of truth or understanding of individuals from a 'whole life' perspective. In attempting to create and expand the EHR as a central representation of truth about the family in order to inform care decision making, we can see data-centric solutionism [REF] being applied to try and solve a problem that was created by a data-centric approach in the first place.

4.1.3 Rethinking the Role of Data in Early Help Support Relationships

While support workers often refer to data from the EHR, the families they are supporting have no access to the data records and are only aware of those aspects that support workers or TAF professionals choose to share with them; often such data is reported only in verbal form and would rarely be shown in its entirety. Critiques suggest more data may only consolidate more power in practitioners' hands and further undermine the families they are meant to be supporting (Neff, 2013; White and Wastell, Crossely). The scattering of data across so many different systems and organisations, combined with informal processes for sharing, provide a serious opportunity for privacy breaches or mishandling of people's personal data. At the most basic level, this might be a violation of consent – the passing of some data, collected for a specific purpose, to another authority for some new purpose without the data subject's explicit consent for such use. The creation of the EHR as a source of truth carries significant risk of disempowering families further and countering the empowerment goals of the programme itself: The possibility of errors in the personal data that goes into the EHR is high, and might result in prejudice or unfair decisions being made. In more serious cases, individual privacy may be violated, or individuals put at risk, if a domestic abuser or criminal gained access to the record. The failure of such case record systems to properly represent families (REF Cornford) produces further risk; information shared by one individual in confidence could be seen by another family member, and this could have extreme psychological consequences, such as an adopted child finding out they are adopted.

Data is not neutral (Gitelman, 2013; Neff, 2013), and collecting data within the context of the delivery of a specific service or intervention rather than as an objective collection of facts undermines local professionals' discretion and organisational agility to deliver the care that is needed ((Cornford et al., 2013); Lowe and Wilson 2018). This means that rather than improving the situation of a family the collection and use of data may be instead reinforcing the existing asymmetries of power that exist between data-holding organisations, the practitioners and the supported families (Cornford et al., 2013).

This context therefore provides an ideal opportunity to study the dynamics of data use and its impact upon service relationships, in service of RQ2. Following preliminary sensitisation research with both families and support staff (summarised in 4.3 below), a study was designed with the objective of investigating the role of data within the Early Help support relationship, from both the individual perspective of both parties (in so doing deepening our understanding of RQ1), but looking at the power balance and effectiveness of the relationship as a whole, remembering that the ultimate goal of Early Help is to empower families to build better lives for themselves and get them to a point where they no longer need support. A further objective in exploring RQ2 is to explore possible alternative models for the use of data within Early Help relationships, and to explore the viability and potential benefits of such models with participants in pursuit of better and more effective support relationships and more empowered citizens. The approach taken to this objective is to conduct participatory research separately with supported families and with support workers to understand their separate perspectives, concerns and needs, and then to identify common goals and bring both parties together in further participatory work to explore and design solutions that would improve the relationship effectiveness for all in pursuit of those common goals.

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

4.2.1 What is Family Civic Data?

As outlined in section 3.5.1, the first step in designing a study like this is to sensitise oneself as researcher to the study context. In this case, there were three things to familiarise myself with - the type of data being stored, the family perspective on the storage and use of that data, and the support workers' perspective on the same. Importantly, I needed to understand how families and support workers understood and talked about this data, so that I could represent and refer to it in ways that made sense to them. To do this, I collaborated with colleagues in the SILVER project and at local authorities to see anonymised examples of what data was used by TAF/Early Help teams or mentioned by support workers as being of interest. I adopted the term *Family Civic Data* to refer to these types of data (further detailed in (Bowyer et al., 2018)) and organised these into different groupings and categories to create a taxonomy. I then created a

taxonomic model of these data types, as shown in table 3:

Table 3. Example Categories of Family Civic Data.

Category	Type of data	Examples/Details
Family	Personal details	Date of birth, address, telephone number.
	Relationships	Marital status, ex's, step-parents, living arrangements.
	Children	Parentage, adoption, fostering, childcare.
Education	School Records	Attendance (truancy), special needs.
	Academic Results	SATs, reports, exam failures, training courses.
Welfare	Social Support	Social worker visits & notes, details of family crises, interventions, allegations.
	Welfare Benefits	Jobseeker's Allowance, child support, Disability Living Allowance, tax credits
Money/Work	Family Finances	Salary, savings, credit cards, spending, debt
	Employment	Job history, periods of unemployment, performance at work, NI, PAYE, pensions.
Civil	Housing data	Council house provision, eligibility criteria.
	Legal documents	Birth/marriage/death certificates, citizenship/immigration status, work permits.
Crime	Criminal records	Arrests, cautions, offenders' registers, prison time, speeding tickets, spent convictions.
	Court orders	Restraining orders, lawsuits, custody, ASBOs.
Medical	Domestic Violence	Allegations made, medical records, social/legal interventions, victim support.
	GP records	GP's notes, prescriptions, tests, referrals.
	Hospital records	Operations, hospital stays, emergency care.
	Medical conditions	Diagnoses, diseases, allergies, blood type.
	Mental health	PTSD, breakdowns, depression, sectioning.

Category	Type of data	Examples/Details
	Addictions	Substance abuse, gambling, rehab, crime.
Leisure ⁶	Library Usage	Books/CDs borrowed, computer access.
	Sports & Health	Gym usage, class attendance.
	Shopping Habits	Loyalty cards, store & online purchases.
	Transport Data	Buses used, ANPR tracking, walking patterns.

Early research recruitment attempts revealed that data is seen as an abstract concept in people's daily lives; a dry, technical topic that many families feel unqualified to talk about. We needed to make these data concepts relatable. Drawing on the work of Brandt and Messeter (Brandt and Messeter, [2004](#)) in creating *design games*, which observes that game pieces can be used to create common ground and as "things-to-think-with" (Papert, [1980](#); Brandt and Messeter, [2004](#)), I created a set of data cards (shown in Figure 8 in the previous chapter), that would serve as a visual and tangible representation of Family Civic Data. By using these as boundary objects (Star, [2010](#); Bowker *et al.*, [2015](#)) the aim was to bring researcher and participants' worlds closer together and to approach the concepts of data by directly starting with individual life experiences. A Data Card was created for each category in Table 3, including a summary and meaningful examples, so that the cards would be easy to digest, yet still contain sufficient detail to stimulate thinking. Keeping child-friendliness in mind, bright colors were a key element of the design. The cards were printed on high-quality, thick card with a glossy finish using a business card printing service to make them appealing and fun.

These cards were then used as research stimuli (see 3.5.2) within a preliminary study in which I met with four families in their homes⁷ and conducted a variety of participatory design activities and design games in order to explore family attitudes to family civic data. This study has been published at CHI (Bowyer *et al.*, [2018](#)) where its full findings are detailed, and these findings serve as researcher sensitisation to inform the main Case Study One. We found that once families had understood data as "stored information about their lives" they were able to very effectively engage and talk about it. The use of the games and the cards was very successful, keeping a light and playful environment and making the topic relatable. The topics on the cards served as a focal point that allowed families to talk freely about their own lives and views without feeling personally interrogated, as they were dissociated from the participants' lives.

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

The families we spoke to did care very much about what happened to their civic data, contrary to the expectations of some of our peers, and perceived a variety of risks due to data mishandling including identity fraud, criminal targeting and psychological harm. Families felt that data could easily misrepresent them through errors, prolonged storage of data beyond its need, or the recording of unfair judgements and opinions. Families wanted to view the data stored about them. They wanted a set of basic rights - to be informed, involved and accurately represented, with the ability to see, explain and correct their data to ensure it is fair and accurate. They wanted to know that their data will be handled sensitively and only by those that need to know, and they believe that having these capabilities would help them to be able to work together with representatives of the state in a more positive relationship.

As well as the need for families to be given such rights, other implications we were able to draw from these

findings were that family civic data is currently used as a proxy for them in decision making, which cuts families out of the loop, and that families should be given the opportunity to have a relationship with their data and also the opportunity to co-operate and have agency in the stewardship of their data. Further findings and insights are published in (Bowyer *et al.*, 2018).

Through my embedded involvement with the SILVER project (see 3.4.1.1) I was able to complement my understanding of the family perspective on civic data use in Early Help, but also was able to acquire an understanding and sensitisation to the staff/local authority perspective on that same data use. SILVER conducted qualitative interviews with supported families, and the findings from these reinforced this need for greater inclusion of families in data handling, having identified that while families were willing to consent to their information being shared in order to improve their care, they had very little understanding of how it was used and could not be deemed to have given informed consent to the way their data is currently used.

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

SILVER conducted a series of “Amy’s Page” (Wilson, Wilson and Martin (2020)) focus groups/workshops with support workers and other local authority representatives, through which I learned that staff had a desire for greater access to health information, particularly mental health indicators. These staff revealed a desire to gather as much data as possible about the families they were working with. The workers viewed the collection of data as a useful raw material that enabled them to do their job better.

Collectively the findings from my own research and from SILVER showed a conflict between the desires from families and support workers – with families wanting more involvement and less reduction to data but support workers wanting to amass more and better data. In part due to its solutionist framing, the SILVER project gave priority to the support worker perspective and continued to pursue the building of a richer data interface for support workers. This was the point at which my research objectives and those of the SILVER project diverged, as I was not ready to ‘take sides’ and wanted to explore whether it might be possible to satisfy the needs of both parties.

4.2.4 Seeking an Equitable and Mutually Beneficial Data Use Model: Shared Data Interaction

In searching for an approach to civic data use in Early Help that might help both families and support workers that could meet both parties’ needs while also addressing our research focus of increased data interaction within Early Help, I began to explore the idea of shared data interaction; instead of the support worker being the gatekeeper controlling and limiting the family’s access to data, and accessing data ‘behind the scenes’ at their offices, what if data could be looked at, examined, and updated together, during the face-to-face encounters between families and their support workers? This could potentially bring all the benefits of human-data interaction (increased agency, negotiability and legibility) (Mortier *et al.*, 2014) to families (and also to workers), while also serving as a boundary object that might improve the relationship itself (Bowker *et al.*, 2015). In theory it would allow families to gain some access to currently inaccessible data while also making it easier for support workers to ‘fill in the gaps’ in the data they already have by simply asking questions.

This concept formed the basis for the main research engagement within this study, so that we would not only exploring RQ1 and especially RQ2 in the context of current practice, but also ask participant to imagine a different set of practices that might potentially serve their needs better and in doing so, assess whether the proposed model of shared data interaction might address both groups’ needs and whether or not it would be perceived to benefit the early help support relationship as a whole. Regardless of whether this particular model was a preferred solution, such an exploration would put participants in a speculative, co-design mindset that would elicit deeper insights about how civic data *should* be used, not just opinions on how it was used currently.

4.2.5 The Workshops

Table 4 - Case Study One Group Design Workshops.

Workshop	Engagement	Phase	Number of Participants	Activities
Workshop A	Design Workshop for Families	1	8 adults and 9 children from 5 supported families	- Data Card Sorting - Sentence Ranking - Ideation Grids - Poster Design - Scenario Discussion
Workshop B (2 instances)	Design Workshop for Staff	1	36 support workers & related staff	- Data Card Sorting - Sentence Ranking - Ideation Grids - Poster Design - Scenario Discussion - Interface Discussion
Workshop C	Combined Staff and Parents' Design Workshop	2	3 support workers and 4 parents from supported families	- Sentence Ranking - Storyboarding Practice - Scenario-based Storyboarding

During the summer of 2018, we conducted four two-hour co-design workshops, with two phases, as detailed in Table 4. In phase 1, the initial objective was reconfirm the findings of early work and gain a deeper understanding of both parties' (families and staff) perspectives on data within the support relationship, by working with each group separately. A further objective was to learn about existing data practices and whether they work, or need improving (and where they do, to identify what the issues were). In phase 2, the objective was to work collectively with representatives from both groups to design imagined data practices and interactions for the shared data interaction model and to understand how *in practice* staff and families would imagine themselves using data together in the support relationship. Across both phases, a variety of participatory methods were used to explore these topics, as described in section 3.5.2 and 3.5.3. All workshops were audio recorded and transcribed. These transcripts were then analysed thematically, and in some cases quantitatively, as described in section 3.5.5. Refer to section 4.3 below for the major themes discovered.

4.2.6 Sensitising Participants and Discovering Shared Values

Prior to the main exploratory activities, it was important to ensure that all participants arrived at a common understanding which they would use to approach their 'design brief'. Also, there was a need to validate whether prior findings about the perspectives of staff and families held true for these participants too. To address both of these goals, a sensitisation (see section 3.5.1) and data-gathering activity called 'Sentence Ranking' was conducted, where participants were asked to consider a number of 'opinion statements' and rank them according to (a) whether they agreed, disagreed or were neutral on that statement and (b) whether or not they felt that statement was important. These statements, such as "*Families should always be able to talk to someone about their data*" (more examples in Figure 21 below and the complete list of sentences are included in [INSERT REF TO APPENDIX SECTION HERE]) were collated from family and staff perspectives observed during the above preliminary study, from the SILVER projects own research findings, and from my own observations through interacting with local authorities as part of my embedded role within the SILVER project. In discussing and reaching consensus on these opinions, families and staff would be in effect 'agreeing requirements' that could inform their thinking during design activities. By conducting this same activity across all participant groups and across both phases, this would also allow comparison between the different groups to identify differences and find shared values.

Within each workshop, groups of participants sat at tables of 4 to 6 people, and each table provided its own sentence rankings. This produced numerical ranking data which was analysed as follows:

1. Sentence rankings were encoded on two scales. Sentences which contained a negative statement were inverted so that disagreement with them could be considered as agreement with a positive statement.
 - a. *Agreement*: neutral (0) -> agree (+1.0)
 - b. *Importance*: not important (0.0) -> important (+1.0)
2. Rankings from different groups within workshops were aggregated, using mean averaging, with a weighting to ensure each workshop contributes equally regardless of attendance.
3. This gave four values for each sentence, for each participant group (families only, staff only, and combined). *Variance* can be understood as ‘unanimity of opinion’: i.e. variance 0.0 indicates total agreement and 1.0 would indicate disagreement.
 - a. *Mean agreement*
 - b. *Variance of agreement*
 - c. *Mean importance*
 - d. *Variance of importance*.
4. Prioritising variance in agreement over variance of importance, the four dimensions were reduced to three to allow a visualisation to be produced.

The data table for this analysis is shown in [INSERT REFERENCE TO APPENDIX]. The visualisation of these findings on shared values is shown in Figure 21. As the figure shows, there was universal agreement that:

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

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

Participants showed considerable contention over whether or not support workers should be able to access historical family records, about how families would feel about the collection of data about them and about having responsibility to managing access to it. Most other sentences received moderate agreement.

AGREE & NOT IMPORTANT	AGREE	AGREE & IMPORTANT
<p>STRONG AGREEMENT "Families don't want to be responsible for looking after their data."</p>	<p>UNANIMOUS "Families should always be able to talk to someone from the authorities about their data." "Public sector officials can not make good judgements just by looking at families' data."</p> <p>MODERATE AGREEMENT "Support workers make better decisions if they have more data about a family." "Support workers should be able to see family medical records." "A family's data should not be joined up and looked at together." "Labels like 'domestic abuse' are damaging to families and hard to shake off." "Numerical scores are not a good way to compare the progress families have made." "Families will be willing to spend time checking their data is correct."</p>	<p>UNANIMOUS "Councils should treat families like people, not records in a database." "Families should have a right to see their data and how it is used" "Information stored about families must be fair and accurate." "Just looking at data doesn't tell you everything about a family." "It is important for support workers to know mental health details."</p> <p>STRONG AGREEMENT "Asking families for consent to share data just once at the start is not enough."</p> <p>MODERATE AGREEMENT "Families' data should be private unless they say it can be shared."</p>
NOT IMPORTANT	CONTENTIOUS	IMPORTANT
<p>NEUTRAL & NOT IMPORTANT</p> <p>MODERATE AGREEMENT "The police should be able to see all of a family's data."</p>	<p>LOW AGREEMENT "Families find setting privacy preferences to be annoying and tedious." "Families won't mind lots of data being collected about them if they can see it." "Officials should be able to see historical records about families."</p>	
	NEUTRAL	NEUTRAL & IMPORTANT
	<p>MODERATE AGREEMENT "Any personal information from more than 5 years ago should be hidden from staff."</p>	

Figure 21: Participants' Shared Values Deduced from Sentence Rankings Data

[TODO: update the diagram to indicate (e.g. via a family symbol and a "support worker symbol" together with either a + or - sign and coloring in green/red) disagreement / agreement by the different stakeholder parties] [TODO: update the diagram so it doesn't look like rows 2 and 3 are in the wrong order]

Having completed the sentence ranking sensitisation activity, participants went on to carry out the other co-design activities as detailed in Table 4. Findings from the analysis of these activities' transcripts are presented in the next section.

4.3 Thematic Findings

The transcribed corpus from audio recordings of workshops A, B and C (approximately 120,000 words) was divided by activity, group, and family or staff focus into 85 different source texts. Each text was thematically coded and analysed through four cycles of analysis using the Miles and Huberman approach (Huberman and Miles (2002)). During this reductive process, participant creations, activity outputs and ranking data were referenced to add additional context to the interpretation. In this section, the qualitative findings from the thematic analysis of transcripts of workshops A, B and C are presented. In 4.3.1 the three main themes and subthemes are introduced, then each theme is further detailed in sections 4.3.2 to 4.3.4, including participant quotes⁸.

4.3.1 Themes & Subthemes

Given that our conversations with participants were framed as explorations of data use within the early help relationship, our findings are expressed as desirable best practices, some of which involved the proposed model of shared data interaction, within three core areas that participants see as beneficial to the early help relationship and ultimately to the family being supported: **Meaningful Data Interaction** (Theme 1), **Giving**

a Voice to the Family (Theme 2), and **Earning Trust through Transparency** (Theme 3). From explicit and implicit statements from participants, contextual clues, and accumulated knowledge from being embedded in this context, we were able to judge whether the discussed best practices were commonly in use (“*current*”), happening occasionally/partially (“*emergent*”) or not yet occurring at all (“*imagined*”). Tables 5, 6 and 7 shows the subthemes within these themes, along with illustrative participant quotes, and indicates the current, emergent or imagined status for each subtheme. Structuring the themes in this way facilitates the functioning of these findings as constructive, actionable input for Early Help (or other social care) system and process designers.

Table 5. **Theme 1 - Meaningful Data Interaction for Families.** Subthemes & Participant Quotes.

Subtheme	Description & Quote	Status
Understandable Information Summaries	<p>To maximise understanding, simple summaries of the information within families' data should be available to both families and support workers. Visualisations should be used to ease comprehension, and information should be contextualised at different levels (individual, family, community).</p> <p><i>“There's so much data that's stored. For me, for a parent [I want] to understand that through a text or email but just in point form. [...] The less written, the better for the parent. [What we need is] a small synopsis [...] like a summary view.”</i> [Parent, SQ44]</p> <p><i>“Some families will go, ‘Well you know that information because it's all there somewhere.’ We're like, ‘Yes, but we don't want to trawl back to eight years ago.’ There's reams and reams and reams of it [data].”</i> [Worker, SQ40]</p>	Emergent
Interact With Data Together	<p>Support workers should work to actively counter the knowledge imbalance by informing families what their data says. They should make use of specific datapoints as talking points to aid planning conversations.</p> <p><i>“You could have a table, you'd look at where they are and where they could be. [You could say] ‘This is where you are now but if you [take these specific steps], even though you've got a criminal record, you could progress to this level.’”</i> [Worker, SQ29]</p>	Emergent / Imagined
Direct and unified data access	<p>Individuals should be able to directly access their civic data through a personal interface; this should be a single, common place where all of an individual or family's data is brought together to give a complete and consistent overview to all parties with a need to know.</p> <p><i>“[I'm imagining an] online database of personal family info accessible [only] by people, practitioners that have permission [...] I would say that it's only who you want [to give access to, that can see it]. You would have your private code which you could hand out, like the doctors give you appointments.”</i> [Parent, FQ8]</p>	Imagined
Ongoing Data Access and Support	<p>It is not sufficient simply to give access to data. Families should be able to access information in their own time and should be supported in understanding it. Most importantly they should be able to ask questions, challenge data records or start a conversation to discuss their data at will.</p> <p><i>“[The families would have] a little app which they can log into and read all their information - what's recorded about themselves, [...] who we share the information with [...]. If they're not happy [...] they can fire off an email to us and let us know what they disagree with or if they want their information taken down or their consent.”</i> [Worker, SQ51]</p>	Imagined

Table 6. **Theme 2 - Giving a Voice to the Family**. Subthemes & Participant Quotes.

Subtheme	Description & Quote	Status
People not Records	<p>Support workers must always treat people like individuals, that are more than a data record. They should review family data before contact, but must always engage at a human level too, avoiding making any judgements based solely on data.</p> <p>Worker A: <i>"You should never make a judgement on data... that data could be wrong.</i></p> <p>Worker B: <i>"It takes individuality, working with that person as well, doesn't it?"</i> [SQ11]</p>	Current / Emergent
Review Data Together	<p>Families should be explicitly invited to review, discuss, check, correct and approve data records. Data recording should be visible, and workers and families should check data together.</p> <p><i>"[The parent could] countersign. [The worker would] say, 'I feel that we've talked about this today so I'm going to write that down. I'm going to show you. Can you sign and me sign if you're happy and I'm going to share this.' That's a bit different [better]."</i> [Parent, FQ12]</p>	Emergent / Imagined
Changing Lives Means Changing Data	<p>Recognising that families' lives are in constant flux, routine reviews of data should occur, and they should be invited to regularly review their choices regarding data collection, keeping and sharing. All systems and processes should treat data as fluid and flexible, not static unchanging facts. Feeds of recent changes should be available to both parties.</p> <p><i>"[There's] this perception of something sticking with you even after you've potentially reformed. [...] That's something that happened a long time ago and that judgement is still there but [you'd be wondering] 'Okay, is it [true]?'"</i> [Worker, SQ61]</p>	Imagined
Family-sourced Data	<p>Individuals should be able to create or contribute their own data to tell their own story and annotate particular datapoints with their own explanations.</p> <p>Worker A: <i>"If you read information [...] about me, you wouldn't expect to meet the person you meet."</i></p> <p>Worker B: <i>"That's it. It's the same for everybody."</i></p> <p>Worker A: <i>"[...] It just [has] basic things in most of the time, doesn't it [...]. You're not a person [in the data record] are you really?"</i></p> <p>Worker B: <i>"I'd like it if you could] give your bit of personal data, your own story."</i></p> <p>Worker A: <i>"Yes, because everybody makes mistakes and there's probably thousands of people out there who have got a criminal record and have never done anything since. [They're] getting judged by having one thing [but they should be able to write] 'Yes, I did this because of this situation but this is what I've done to make myself [better]...'"</i> [FQ10]</p>	Imagined
Granular Access Controls	<p>Families should be given controls to manage access to their data and configure and change preferences at a fine-grained level.</p> <p><i>"[Families need to] feel they're being involved. [...] [We need to be able to] sit together and say, 'Right, that's the information I'll allow you to share. I don't want that bit shared. But this bit, because it will help me and the family [...]. Say in this [scenario] family, she might have been married before and had domestic violence so she doesn't want that bit shared, that's in the past. So it's [only] certain up-to-date information about the family [that would be shared] because this [the family suggested by the data] isn't her family."</i> [Parent, SQ76]</p>	Imagined

Table 7. **Theme 3 - Earning Families' Trust Through Transparency**. Subthemes & Participant Quotes.

Subtheme	Description & Quote	Status
Respectful Data Handling	<p>Support workers should treat families' data with the utmost respect, keeping it safe, ensuring it is not used beyond its intended purposes, shared without consent or put at risk. When talking to families about data, it is helpful to focus on positives and strengths and not use it as a means to criticise.</p> <p><i>"There was a time where I was at the doctors' and they asked how many units of alcohol I drank, and I said, probably about three bottles a week, at the time, not any more but later on [the support worker] pulled me up on it and they had it down as three bottles a day. That could have caused an issue was anyone ever to ask."</i> [Parent, CQ7]</p>	Current
Always Seek and Demonstrate Greater Understanding	<p>Support workers should always assume that their understanding from data is incomplete and should seek to learn about individuals and build a more complete picture of their lives. By showing this effort and their growing understanding, they will engender trust.</p> <p><i>"You don't want to reduce them to this number in a database. You want to understand their actual experiences and support them in getting better."</i> [Worker, SQ74]</p>	Emergent
Pro-actively challenge data-centric norms	<p>Support workers and agencies can recognise that current systems and processes are data-centric and imbalanced, and can strive to change this through their actions: being as open as possible about how families' data will be handled, ensuring that proper oversight mechanisms exist for data handling especially in the sake of contentious issues, and that data is shared openly but consensually between authorities.</p> <p><i>"It hasn't been explained properly to this [scenario] family that their information will be shared with other professionals. So, they've been left feeling really let down and probably quite angry about it. So, although that information does need to be shared, they [the support workers involved] ought to make the family properly aware that information will be shared."</i> [Worker, SQ18]</p>	Imagined

4.3.2 Theme 1: Meaningful Data Interaction for families

Through our discussions with families and support workers we gained a deep understanding of what sort of data interactions were considered ideal for a family. Setting aside interface considerations, which were not the main focus of our enquiry, and focusing on the wider sociotechnical context around the data and its access, the key requirement we uncovered was that in order to maximise understanding for all parties, data interaction needs to be *meaningful* – this is the first theme of these findings. Encompassed within this concept are the need for understandable and effective summaries and visualisations, the need for direct and ongoing data access with human support, and the recommendation for families and support workers to interact with data together within the support interaction.

4.3.2.1 Understandable Information Summaries

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

"In that example, depression, ten years ago, that shouldn't be on there for the support worker. All they should get is if Social Services have been involved and it should just be, 'Please contact for more information.' [...] [The system should stop workers from] getting a list of all the kids who have ever missed dental appointments or when you were depressed ten years ago. [...] There needs to be a thing where it's,

sort of [...] key trigger words, where if the word comes up a lot of times, it spots the patterns. Whereas, if [a problem] is mentioned once, it should only be [shown] at the highest level.” [Parent, CQ10]

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

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

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

4.3.2.2 Interact With Data Together

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

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

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

Participants particularly recognise the value of referencing data points over time (such as a record of welfare scores that support workers have previously given them), for example to track progress [SQ29, shown above in Table 5]. This could motivate and reinforce progress [SQ6] by relating behaviours to consequences [SQ32] – essentially facilitating data-based decision making. Reviewing historical data is preferable to verbal description: “*Whenever you go through stuff like that [verbally], especially historic stuff, they can be quite remote so [having the data in front of you] would be good for that.*” [Worker, SQ33].

4.3.2.3 Direct and Unified Data Access

Despite the reality that families currently have no direct access to their civic data, family participants all eagerly described designs including apps, intranet terminals, online chat facilities, and self-service webpages, all offering individuals the ability to view their own data; there is a clear demand for *personal data interfaces*, which could empower families to use their own data: “*they could quickly tap onto the app [...] and show somebody else where they’re at.*” [SQ54]

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

Workers and families shared a desire for one single point of access for data, usable by all parties [SQ25, SQ26], though families '*don't want to be responsible for looking after all our data*'[FQnn]. Bringing together data from multiple sources would allow patterns to be spotted by correlating data from different sources, which workers perceived would help their preparation: "*[This imagined interface] would provide individual histories but you could also pull them all together so you can prepare, so for instance if mum was having some significant issues with mental health, you might be able to correlate the [child's] school attendance alongside that and find out why that's happening.*"[SQ8]

4.3.2.4 Ongoing Data Access and Support

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

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

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

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

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

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

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

4.3.3 Theme 2: Giving a Voice to the Family

The second theme of these findings is that there is a need to update processes and systems, which currently rely largely excessively upon the 'facts' within the data record, need to be updated to give the family an empowered role within their civic information ecosystem. The purpose of an early help intervention is to obtain more information for a better understanding of the family's situation and to make evidence-based plans and decisions to improve the situation, so seeking objective truth is clearly central; impressions of that truth can be formed either by reading the data or by talking to the family. We uncovered benefits and dangers of relying solely on either source. Families should become agents in the data ecosystem, and this involvement should lead to both greater empowerment and better evidence-based decisions.

4.3.3.1 People not Records

We found evidence, consistent with literature (Gitelman (2013)) and my earlier study (Bowyer *et al.* (2018)), that data can never represent absolute truth - it is often biased or incomplete, and this can mislead [SQ59 (shown in Table 6 above), FQ11A]. For example, a lack of mental health information could make an individual look like a poor parent [SQ12]. Families may be less willing to ‘open up’ if they feel they may be judged unfairly [SQ14]. Therefore, developing a strong relationship between worker and all family members is key to understanding the full picture [FQ1]; to ensure fairness, data must be current and complete [SQ13], but this state can only be achieved with the family’s cooperation. Looking at data will never provide support workers with a complete understanding. Yet, workers often ‘*tend to just trust that everything that has been put down is right*’[CQ1], allowing the data perspective to dominate. Such assumptions should be avoided [SQ10]; processes must recognise maintaining human face-to-face dialogue as a priority. Data should only provide supplementary insight: “*You should never make a judgement on data... that data could be wrong. It takes individuality, working with that person as well, doesn't it?*”[SQ11]. All participants presented with the sentence “*Public sector officials can make good decisions just by looking at a family's data*”[S18] disagreed with it.

In spite of the warnings above, the data record is undeniably useful; over 80 comments from workers believe the current practice of reviewing a family’s data prior to meeting in person is beneficial, because it provides useful background that will help them identify support needs. For example: “*I had a family where trying to unpick what had happened, over ten years, to the child, was really difficult. So, I went away, got the information and came back and if you have [...] that picture of how the family works [when you meet them], [that helps].*”[SQ1] Additional benefits identified included safeguarding workers [SQ3] or giving them an ability to ‘check the family’s claims’ so that they might constructively challenge individuals [SQ4]. Supported families echoed the value of workers reviewing data [FQ1A], and saw benefits included ‘not having to repeat your story’[SQ5].

The compromise that participants identified over the use of data is that workers should avoid making judgements based solely on data. While sometimes providing essential background to a worker [FQ11B, SQ62], historical data in particular often leads to inadvertent prejudice, especially where labels are used [SQ9]. No participant disagreed with the sentence “*Labels like ‘domestic abuse’ are damaging to families and hard to shake off*”[S15], and workers recounted experiences of being uncertain how to judge historical issues: “[*There's] this perception of something sticking with you even after you've potentially reformed. [...] That's something that happened a long time ago and that judgement is still there but [you'd be wondering] ‘Okay, is it [still true]?’”[Worker, SQ61]*

Many participants concluded that only ‘relevant’ information should be available, to those who ‘need to know’, but the wide range of opinions we saw expressed suggest that this is a highly subjective judgement that would be difficult to determine. A cut-off period before which workers should have no right to look was suggested [Parent, CQ15], but the sentence ‘*Officials should be able to see historical records about families*’[S17] was contentious. Some workers feared any restriction in access might mean they miss important background on an individual’s past, such as sexual abuse or mental health issues [Worker, SQ77]. The solution to this dilemma is unclear, but transparency about what is in the data would seem to be a critical ingredient (see 4.3.4).

4.3.3.2 Review Data Together

- E. Review data together
- F. Families can get their data changed/corrected.
- G. Record data visibly and get family sign-off.
- H. Workers and families checking data together (for accuracy).

Another option arose at the intersection of relying on dialogue and having families contribute data, suggesting workers visibly recording data in front of families on a tablet or 2-in-1digital device and then asking them to ‘*approve*’the accuracy [FQ12, SQ67]. Participants believed this would help to build trust

between the support workers and families.

Families imagined going beyond verbal explanations and being able to visibly review records that their support workers were creating and signing them off [FQ3]. If a family begins to feel powerless, they may disengage [SQ35]. Even minor involvement, such as signing off approval of data records [FQ12], could make the support relationship more productive.

I. Support changing lives

One reason for reviewing historical data and for requiring dialogue with the family to gain an up-to-date picture, is that the truth changes over time. People are not static, and families' lives are always changing given marriages, divorce, birth, death, house moves, etc... Inaction will cause data to become out-of-date, as it is inherently static – data does not change, but people do [SQ61, SQ63]. Accordingly, it is not just the content of the data, but the family's consent over what happens to that data that needs to be reviewed regularly:

Worker: "*If [consent] has to be all or nothing I can't see that is ever going to work [...] most people will have something they don't want anyone else to know.*" Researcher: "So, you might agree [to data sharing] as a family now but after a break-up, what happens?" Worker: "Yes, or the kids reach a certain age and they might say, 'Hang on a minute, I don't want you looking at all my medical records.'" [CQ16]

I. Regular reviews of data and consent with family.

This is important to prevent unwelcome surprises about how family data is handled [CQ2, CQ17], which can damage trust and hinder co-operation.

I. Make a 'feed' of family data changes available to both parties.

Several participants imagined ways how technology could support this, with the databases issuing notifications or update feeds for families and support workers showing significant events or data updates. Support workers currently get notified of police incidents, safeguarding concerns and hospital admissions, but including data across the care ecosystem could provide useful triggers for reviews or discussions [SQ64, SQ65].

I. Give individuals agency

When considering their relationship to their civic data, families (and support workers) imagined the family being able to act alone. This unlocked additional thinking, for example families helping to fill gaps in data [SQ57] or contribute new data that may not otherwise be recorded [SQ58]. Giving families the ability to contribute new data could be very empowering as it would allow them to '*tell their own story*' [FQ10]. This is vital because, not only because families' civic data will never be enough to adequately represent complex lived experience (Bowyer et al., 2018) and it is important to ensure the data is accurate: If something goes wrong, families must be able to do something about it. Without a cycle of feedback involving individuals as stakeholders having the ability to review and correct data, data will quickly become inaccurate (Pollock, 2011). A role in data reviewing, production and recording is empowering, as families are then able to hold information holders to account to ensure fair treatment, something that is currently very difficult to achieve. "*I think that [families] have got a right to know what is held about them and what is said about them.*" [Worker, SQ50] "*I just generally want to see [what is stored about me] just to know what people are saying and then obviously if it's wrong, I can correct them on it.*" [Parent, CQ14]

I. Family-sourced Data

II. Create or contribute own data.

Many participants envisaged that rather than solely relying on dialogue with families, families might provide new data more directly, e.g. through a 'family network app' for information contributions, which could also increase their sense of data ownership:

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

Some participants noted that families are better placed than anyone else to identify inaccuracies or gaps in their civic data, and that more accurate data could be obtained through families providing corrections. This does not mean free editing of records (as, for example, fears and/or self-interest could lead to families misrepresenting themselves in data (Bowyer et al., 2018)) but rather taking a role in reviewing, annotating, explaining, filling gaps [SQ57, SQ58] or requesting changes. This was imagined as a collaboration between workers and family members:

"[There would be an] individual view where each person within the family would have their own section [...] you could sit with them [...] and go through the data that we have got which would enable them to change anything that they want taken out." [Worker, SQ66] With new ways for self-expression, families could add context for support workers [FQ9, SQ55], unlocking new support topics [SQ56]. The overriding sense from both groups is that families having the ability to annotate or explain their data would allow them to hold authorities to account, and empower them to tell their story and '*show the real me*' as illustrated in [FQ10, shown in Table 7 above].

I. Granular Access Controls

II. Enable families to manage access controls to their data. Participants identified that it is important to consider that different individuals within the family would have different roles, access and summaries, in order to respect individual privacy [SQ52, SQ48]. This could extend beyond just giving consent to managing more fine-grained access controls:

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

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

We can see that giving families a role in the creation and stewardship of their data-selves has great potential to unlock new capabilities and a sense of empowerment for families.

close: Systems and processes support change.

4.3.4 Theme 3: Earning Trust through Transparency

The findings above clearly suggest that in seeking best possible understanding, families must be engaged in a fact-centric way, which requires trust in the support worker and in the system - A good relationship with the support worker is critical [FQ1]. Our findings show that transparent and open data handling and decision-making processing are key and can allow support workers to earn the trust of supported families. The key topic of trust arose directly or indirectly in almost all participant conversations. Currently, families are typically unaware of what data is held about them and what discussions about them are being had. Even if the data itself would be considered uncontroversial, this lack of transparency and how data informs judgements can cause great worry to families:

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

- C. Respectful Data Handling
- D. Respect family and individual privacy.

The current approach, which relies on the support workers mentioning data that they consider relevant, can result in expectations being broken by accidental sharing of information for which sensitivity is not recognised:

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

Lack of transparency and trust can lead to an atmosphere of suspicion [SQ17] where families have 'a totally overwhelming feeling of people checking up on them' [SQ71] and apply greater scrutiny to what they are told: *"You can get families who [...] don't believe what's being said about them."* [Worker, SQ73]. Fearful of consequences [SQ72], families may withhold information:

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

Ultimately an open and respectful approach is not just rooted in decency but in practicality as a co-operative family is easier to support: *"Because if someone is feeling judged or stressed or angry or whatever, then they can stop the conversation"* [Parent, CQ5]. It can also help with accountability and accuracy:

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

- C. Explain how data will be used and shared.

To avoid damaging negative spirals of emotion, deliberate openness is needed from support workers (and the entire care system) [SQ18] as to what information is held, and how it will be used and shared, in order to alleviate fears of data being used against families, instead giving them confidence that their interests are being protected, thus putting them at ease [SQ20]. To our understanding, this is only done once in very loose terms during initial engagement for the purposes of collecting informed consent.

- C. Avoid data mishandling and unexpected uses.
- D. Use a strength-based approach when referencing families' data.
- E. Always Seek and Demonstrate Greater Understanding It is clear that the support worker must always be seeking to form a completer and more up-to-date picture of the family. Demonstrating a deep understanding of the family, and that they are prioritising a family's lived reality above what the database says can be a key principle for workers to build trust: *"You don't want to reduce them to this number in a database. You want to understand their actual experiences and support them in getting better."* [Worker, SQ74]. It is important that families understand workers' good intentions when accessing data about them [FQ15]. However, having to show all the data to families could make it challenging to make that case convincingly, *"because literally [the data we have] is like everything, isn't it? So I don't know how I would feel..."* [Worker, SQ21]. In addition to avoiding breaches of expectations (see Theme 2 above), a transparent approach ensures that the privacy of families is respected, because data is not viewed without the chance for explanation:

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

- E. Show and maintain deep understanding of families' lives, not just their data.

Participants also indicated that the desired transparency is not just about reporting data usage. Dialogue and engagement is needed. Support processes need to change to better recognise the role of dialogue, rather than just consultation of a database, as the best way to achieve a rich and nuanced understanding.

- E. Always seek a more complete picture.
- F. Pro-actively challenge data-centric norms
- G. Proactively counter the knowledge imbalance.

Workers could easily imagine explaining data in greater detail than they currently do [SQ41] and clearly there is a need for proactive action by workers to counter the inherent knowledge imbalance of data being collected into systems that they are gatekeepers for.

However, workers lack control over the quality, coverage and timeliness of the family data and see this as a systemic issue they would not always be able to address. From our collaboration with multiple local authority early help services through the SILVER project (Connected Health Cities ([2017](#))) it became clear that while support workers can see more data than most, they have far from the complete picture; in fact, there is no one organisation or individual with visibility of the entire family-information ecosystem. (I) Be as open as possible with families about their data. Some participants suggested that openness to increase trust could go as far as browsing new information together [AS MENTIONED ABOVE] rather than just having the support worker get it first and report its content:

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

- I. Independent oversight of data handling for contentious issues.
- E. Ensure that consent is never assumed.

In current practice, consent is seen as a one-off formality at the start of the support process.

A common phrase that emerged here and in our prior research (Bowyer et al., 2018) is that data should only be seen by those that '*need to know*'.

[but who judges]

Without transparency of data handling, a family cannot verify whether this is happening. This is especially important given that some support workers expressed a belief that their right to access families' data could overrule families' consent:

“I think to enable us to work with families, we need to have as much information to give them the best possible service. So, I think we should be able to [access their information] regardless of what families say.” [Workers, SQ22]

We also found evidence of other reasons why accountability is important such as the difficulty of deciding what parts of a medical history are '*relevant*'[SQ23], arbitrating situations where legal duties may require the breaking of consent [SQ24], and being able to identify and address situations where recorded information may not tell the full story [CQ8].

We see evidence that transparency and openness are key to building trust around data access and decision-making as requirements for a healthy support relationship. Shared data interaction could be one way to achieve this, bringing benefits in accountability, accuracy, simplicity [SQ25, SQ26] and consent.

[close](#) Explicitly involving families in processing their data. [(I) Promote an open data-sharing culture.]

4.4 Discussion

[Target 3,000 words or maybe more?] [Then comes the ‘Interpret’ part. Where you assess how these things matter, do they confirm or disconfirm theory, relating it back to the literature. This part heavily references theory from your lit review. Frames the themes/subthemes as knowledge contributions.]

4.4.1 The Value of Involving People With Their Data

Through our analysis of attitudes to data usage in the UK early help context we have shown that data about supported individuals and their families is already providing great value to building up a more complete picture of a family’s life, in service of better support and decision-making. However, this comes at a cost to the family’s autonomy and we have identified a number of problems with the prevalent mindset in the care system – which is that families’ civic data is considered as a resource to be utilised. Such a mindset (sometimes known as ‘dataism’) carries an implicit assumption that data is an objective source of truth, however this would require trust in the independence and integrity of the data-collecting and data-holding institutions (van Dijck, 2014), which we have shown is often absent. Supported families lack awareness of what data is held about them and how it is used: this can lead to false expectations and surprises and in the worst cases, this can feed feelings of fear or suspicion which can harm the effectiveness of the overall care relationship. Stored data can often serve as a proxy for their involvement (Bowyer *et al.*, 2018), and without any involvement of the family in checking data accuracy, the current system is susceptible to inaccuracies and errors of judgement due to out-of-date, incorrect or missing data, which can directly affect supported families in the form of prejudice, discrimination, or privacy violations. A key finding we have uncovered is that trust is critical to a support relationship – trust in the support worker, and trust in the system as a whole, and that the current usage of families’ data is not conducive to trust. The best way for a support worker to build trust with a family is to show that they have, and are continually striving to develop, an ongoing and deep understanding of the family as individuals, whose perspective is more important than *‘what the computer says’*. The more they are treated as people, not *‘objects to be administered’* [(REF Cornford *et al.*, 2013)], and the greater awareness and access they have to data records and data handling and decision-making processes, the greater the trust they can have in the system and the more effective the relationship will be. Shared data interaction practices such as checking data together, visible data recording, family sign-off, or contribution of their own perspectives as data, give the family direct evidence that they are being listened to and that their viewpoint is important even when it contradicts the digital record, and this would be very powerful in building trust. Furthermore, transparency of processing allows accountability – something that is currently all but impossible, and this would further empower families by allowing them to gain confidence that they are being treated fairly and that data about them is accurate. It is evident from our findings that a trustworthy care system requires the direct involvement of the individual(s) being cared for and that the mechanisms of shared data interaction offer specific shapes in which that involvement could take place. Consistent with field studies such as the World Health Organisation’s decision-making tool [(Johnson *et al.*, 2010)], we found evidence that staff and supported families believe they would be able to collaborate more efficiently through shared data interaction as it would be more evidence-based (see [REF what was 8.2.3] above). This has the potential to remove inefficiencies such as spending time correcting misunderstandings or repairing damaged relations caused by misjudgement, and the emergent practices of using data to track progress are already proving to be an effective and tangible way for families to improve their situation; giving them the ability to track this data outside of the support engagement would empower them even more to be self-sufficient. A digital health innovation project in South Africa echoes our findings on the importance of trust, agency and involvement of the individual: *“The user must feel or experience trust, have to change behaviour, feel that they can control and increase their own access to a system. Their uptake and use are essential for such a [digital ecosystem] to work or to be regarded as a sustainable solution.”* (Herselman *et al.*, 2016) Viewing data as a shared resource to be curated together would also solve the problem that the current system is in effect lacking a true consent mechanism, since the initial consent is in practice, a handover of power that gives the care authority carte blanche to collect and use data about the individuals (see [REF] 8.2.2). In effect, the ongoing access to and direct use of data by families (see [REF] 8.2.3) would serve as a practical implementation of a ‘dynamic consent’ model [REF (Kaye *et al.*, 2015; Williams *et al.*, 2015)]. Instead of consent being seen as the acquisition of a formal permission that has to be certified, stored, reviewed and modified, adopting simple practices such as talking

families through their data and carrying out regular checks together could provide a practical but less bureaucratic guarantee that families are on-board with the way their data is being used, since their ongoing awareness and absence of complaint can be taken as satisfaction. If implemented in a robust manner, this approach has the potential to greatly simplify the consent challenge for authorities, requiring less process and reducing liability. Families will be happier with the use of their data if they can see it, notice and speak up when they feel something is amiss. Additionally the sharing of responsibility for data stewardship between both parties can reduce the liability for support workers; some were fearful of missing something important when given access to large amounts of families' data – in this model where conversations are more focussed upon data, relevant information can be identified more quickly while at the same time mistakes can be spotted sooner. With families involved in checking and shaping their own data, that data can become more reliable and accurate, which goes some way to addressing the problems described by Cornford et al. of the state forcing families to be represented through data models that are not up to the task of representing the complexity of their lives [REF (Cornford et al., 2013)]. This need to give the user a role in understanding and influencing the life of their own data is identified as a key ingredient of moving towards a more progressive model of digital citizenship. In 2016, Bridle explained:

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

Perhaps the greatest benefit to the care organisation of shared data interaction approaches would be the inclusion of supported families to a much greater degree as a stakeholder in their 'case'. Instead of the care worker using their authority to pass judgement and deliver advice, the care worker becomes an ally, with the family member(s) empowered as an agent in their own self-care, with a greater ability to take action and drive things forward than they had previously (see REF Theme 2). Supported families would be able to trust that their interests are being looked out for and that through their ability to contribute to and access their 'data self', to take part in informed decisions that could improve their lives, and to use their data in new ways to serve their own ends.

4.4.2 Effective Data Access Requires Human Interaction

Our findings reveal that the current inequality over families' civic data will not be solved simply by opening up databases to families and giving them access. They must be able to meaningfully comprehend the data and meaningful effect change based on what they learn from it. This involves the translation of raw data into meaningful information – through summaries, visualisations and explanations (the challenge of who can or would create these information representations is an area for future study). The information available to the individuals must be – as described in one of the central tenets of Human-Data Interaction – legible (Mortier et al., 2014) but also their access must be effective (Gurstein, 2011). This includes providing suitable opportunities for access – not just within the support meetings, for example – and addressing technology, literacy, mental or physical handicaps and means that the information should be supported by a human relationship – one where someone can both explain the data as well as answer questions about it (see [REF] 8.2.3). It is the combination of effective data access and human-to-human interaction that makes data access meaningful, and the former without the latter will not empower the individual concerned. Access to data must be supported by a conversation. The system needs to have a human face that the individual may put their trust in and to which they can address their questions; simply giving access to raw data would be inadequate (see 8.2.3), and limiting [REF (Cornford et al., 2013)]. As a sub-field of Human-Computer Interaction, it is easy to assume that HDI is simply about interaction in the traditional sense of users interacting with data through an interface, but our work, which has, guided by our participants, focussed less upon layout and screen interaction and more upon the wider sociotechnical context of the support relationship, suggests that HDI can be more effective sought when the word 'interaction' is considered in an interpersonal sense. By focussing on the human aspect of the proposed use of data within

the support relationship, we can see that as well as improving accuracy, consent and trust, shared data interaction can bring practical benefits by facilitating a better interpersonal interaction. By physically bringing data into the interaction – be it a printout of a table or graph, or a tablet or 2-in-1 device – rather than just reporting it verbally, this representation serves as a focal point for discussion, bringing both parties to the same topic space faster and more efficiently than abstract discussion would. The data records here function as a boundary object [REF (Bowker et al., 2015; Star, 2010, 1989)] – the families understand it because it relates to their life, and the support workers understand it because they are familiar with the systems it came from. As such, it can become a valuable tool for encouraging families to open up. Many of our participants talked about how looking at data would provide a discussion stimulus or serve as a conversation starter (see [REF] 8.2.1). Also, it provides support workers an opportunity to be less adversarial, by positioning themselves as equals looking at the data together (*'let's make sure this data is right'*) rather than appearing as if they side with the data by being the ones who voice it (*'Our records say that you have...'*). The effectiveness of having data representations as “things to think with” that can establish common ground is discussed in our prior work (Bowyer et al., 2018) and is also echoed in the methods in this study. In particular in our third workshop, which brought support workers and supported family members together, used storyboarding action cards in specific fictional scenarios. These cards provided a focal point for discussions and helped the participants to quickly imagine a realistic situation, again serving as boundary objects. The yellow (for families) and blue (for staff) borders on the cards helped ensure that both parties owned a piece of the puzzle: We had given no direction about who would place which cards, but we observed parents feeling confident to place yellow cards and support workers keen to place blue cards, because the card helped them identify with the corresponding role in the scenario and feel ownership over the choice of options that would be available to them. Similarly, the green bordered cards (which corresponded to those actions involve both parties) almost always resulted in both parties discussing and agreeing a view before the card was placed. If we relate this to an imagined discussion of actual data records, we can envisage that the presentation of the data as being “yours” or “ours” would have a noticeable effect upon how the families would engage with it, and the strength with which they would perceive the power of the data holder over them. Having access to the data within the context of the support relationship is a key enabler of the Human-Data Interaction properties of agency and negotiability (Mortier et al., 2014) for the family members; an ability to interact with and correct or comment on the data directly would give them some agency that do not currently have, but if there is no ability for their comments or corrections to the data to actually influence the support discussion and the work being done, then they have no negotiability - their data access is not really part of the system, it would be tangential to the actual support process. Therefore, efforts to deliver effective HDI capabilities in future should focus on interpersonal interaction, and the role of the human in the information system, as a data interface is limited by its operational context as to its ability to truly empower a data subject (indeed, even the term ‘data subject’ which persists even in progressive data paradigms such as the EU’s General Data Protection Regulations [REF (European Parliament, 2016)] embodies the prevalent problematic stance, evoking as it does imagery of a medieval king looking down upon his subservients). As our participants all strongly agreed, supported families *‘should be treated like people, not database records.’* [REF (see S4 and Table REF3, in section REF8.1)]. This framing can inadvertently become problematic in early help practice focusing upon child welfare: *“children [can be seen as] the objects of a variety of concerns which need to be acted upon rather than agents of their own lives”* [REF (European Commission, 2014)]. Analysis of the Child Index, an early warning electronic information for child welfare in the Netherlands, drew a similar conclusion on the importance of maintaining a compassionate human aspect in family-state relations: “Taking into account that [care] professionals’ first love is the best interest of and care for a child, it is recommended for policymakers to provide enough room for the ‘love’ between future technologies and their social actors to flourish.” [REF (Lecluijze et al., 2015)]

4.4.3 The Benefits and Implications of shifting the Locus of Decision-making

Through our discussions and activities with support workers and supported families, we have explored the possibilities of shared data interaction and our concept of shifting the locus of decision-making – the idea that power concentrates close to the point where data is accessed, and that by shifting the point where data

is access to the centre of the support interaction, this would necessarily shift decision-making closer to the individuals and thus swing the power balance in their favour (see section [REF]7.2 above for a full explanation and refer to Figure 1 for an illustration of the concept). In this section we examine the potential benefits and implications of such a shift. The introduction of data interaction into the interpersonal interaction of the support relationship can be seen as a change to the nature of the support relationship, in that some of the work that was previously done solely in the domain of the data holder (specifically, data maintenance and the direct use of data to inform judgements and plans) is now taking place in a different context – the two-party context of the support engagement itself. So at a basic level, the power is shifted by the new approach. The use of data in current practice is limited because any data must flow through the support worker as gatekeeper – the removal of the gatekeeper role redistributes the power to interpret, select and judge data much more equitably between the two parties. The potential benefits of this shift in terms of empowering families are significant. As detailed above it would give them a role to play as agents in the life of their data, and a new ability to create and curate their own ‘data self’ – the representation of them that is seen by the state – so that it is as fair, accurate and representative as possible. But more than that, given the increased visibility of the metrics by which their progress is judged, they are now empowered to take steps to influence any poorer metrics by making improvements in their own life that would result in those metrics improving visibly, which then could then use as evidence to prove their achievements – a positive feedback cycle that was previously inaccessible. By shifting the locus of decision-making, families can take more responsibility for their own lives, through an increased ability to reflect and make plans – an important element of harnessing one’s personal data for self-improvement [REF (Abiteboul et al., 2015; Li et al., 2010)], thus ‘encouraging the family to take full accountability for their own responsibilities’ as one support worker put it [SQ75]. The perceived benefit of individuals directly using digital interfaces for health and wellbeing are already accepted, with 93% of doctors believing that apps can improve health outcomes [REF (Kostkova, 2015)]. The above are benefits to the supported individual, which of course can be seen as benefits to the care provider as well, given that the function of the early help service is to help the supported family improve their situation as effectively as possible. But shifting the locus of decision making also carries practical benefits for the care provider too. If the family are involved in the stewardship of their data, this reduces the burden and responsibility upon the authority to look after that data – instead, the responsibility for ensuring completeness, accuracy and fairness is now a shared responsibility. And if responsibility is shared, this must surely also reduce the likelihood of complaints or litigation, because it can transform the way that families think of the care provider away from ‘us and them’ thinking towards a more equitable stance. An additional advantage of a cooperative approach to data stewardship is that the consent problem is solved; the scope for non-consent is reduced because at every single meeting (and perhaps even outside those meetings if individual personal data interfaces are available) the supported families are involved in a conversation that directly enables them to voice their approval or concerns for the ways their data is being used. However, implementing such a change to the system would not be without its challenges. There would be significant costs: New equipment such as tablets or 2-in-1 devices might need to be purchased if support workers do not already have these. New software interfaces would need to be commissioned, developed and purchased. The existing configuration of IT systems in the public sector (see section [REF]) is not well suited to the creation of such unified data interfaces due to its fragmented nature [REF (Copeland, 2015)]. Identity management in this context is already very challenging to negotiate [REF (Wilson et al., 2011)]. Support workers would need additional training both on software and hardware. The need to increase digital skills across health and social care has already been identified as a current issue in the UK [REF (Honeyman et al., 2016)] and in other countries such as Poland where it is deemed critical [REF (Soja, 2015)]. This will become particularly important in a system where the care workers are also the ones who would be helping individuals to make sense of digital information. The use of computer-based communication and information approaches would need particular care with child welfare [REF (Tregeagle and Darcy, 2008)]. Local authority business processes would need significant overhauls to recognise the individual members of the public as an important part of the system – which would likely carry with it new considerations for system access controls, technical support and public liability insurance. In particular the provision of personal data interfaces to the public, and new communication channels for public enquiry, would carry with it a large human resource burden to manage and support those channels and usages.

While the creation of a direct communication channel between supported individuals and support services does on the face of it have the potential to carry some savings for the state in terms of reducing the amount of “in-the-home” contact necessary – which is particularly challenging and costly to deliver in rural areas far from major towns [REF (Kriisk and Minas, 2017)] – the idea of the data access being supported by human contact, and of making more decisions together, may ultimately require a greater investment of manpower in communicating with supported families. Measures would have to be put in place for when things go wrong: dispute resolution procedures and additional legal and information governance support would be likely to be needed. It is also possible that giving more power to families could create new challenges: it is not impossible that particular individuals, for whatever motivation, might try to be destructive, manipulative or otherwise challenging to the system, and they might try and use their new powers against the state (for example, hiding criminal activity, misleading workers etc. for personal gain). While very unlikely to be a mainstream issue, this is a fringe possibility that must still be considered and planned for. It would be fair to criticise this model of human-centred state interaction in that it would be not be cheap or scalable; in essence this model creates mechanisms for families to have more interactions with the state, which means that every case would take more worker time in a system that is already overburdened and underfunded [REF (Copeland, 2015 ADD REF Local Government Association)]. The state has increasingly adopted a data-centric approach to citizen interaction because it cannot manage to provide human relationships with individual citizens. But now this approach has become ingrained into government approaches to citizen relations – *“it is no longer a technological necessity but it has become a political intention”* [REF (Bridle, 2016)]. What we have identified is that there is a need to reverse this trend, if people’s interests are to be best served, and if a welfare state is to be truly [REF ‘enabling’ (Miettinen, 2013)]. By taking a more innovative approach to digital policy, it is possible that governments could be more effective in helping to involve those citizens that have become disadvantaged by the current system – a more human-centred approach could help to combat the digital divide [REF (Kalvet and Tan, 2008; Steyaert and Gould, 2009)]. In our model that shifts the locus of decision-making, we have not sought to provide an implementable solution that could be rolled out at scale, rather we consider our model to be a useful mental model to stimulate further discussion - our contribution’s value is that it shines a light on the positive and negative impacts of current procedures on relationship effectiveness, and identifies imagined practices that could be preferable and more efficient than current practice. Our findings serve as a challenge to the status quo, that should encourage early help providers to question their priorities when it comes to the use of people’s civic data and when they consider reconfiguring their services.

4.5 Summation

[Target 1,500 words or maybe less?] [Finally after the TSEI structure for each theme, there is a combination/conclusion part of the chapter, bringing the three themes together.] [In 4.5, I will summarise the case study in terms of how these insights expand our understanding of the research questions and their wider significance.]

[TO REDRAFT AND SHORTEN POSSIBLE REF THAT COULD BE ADDED AT
https://miro.com/app/board/o9J_kvbFr24=/?moveToWidget=3074457347460416545&cot=13]

Through our participatory co-design workshops with supported families and support workers in North-East England, we have highlighted five major problem areas which our participants perceive to exist with current personal data practices: 1. A power imbalance – Families’ personal civic data is collected by care organisations and viewed as a resource to be utilised by the support workers, creating a structural power imbalance against families which is further emphasised by the authority, influence and network centrality of the support service with each family’s data landscape. 2. A closed and opaque data ecosystem – Families lack awareness of what data is held about them and how it is used, with support workers (who themselves have limits to their access) functioning as gatekeepers to what families will be told about. 3. Ineffective, meaningless consent – The current consent model, while legally satisfactorily, is ineffective, as it is viewed as a one-time initial hurdle after which support workers can do whatever they deem necessary with families’ data and those families are never again given any meaningful choices about what happens to their data. 4.

No accountability and limited trust – Without any transparency or ability to request or demand changes to data or data practices, families have no ability to hold data handlers to account. The lack of visibility makes families' trust in the system hard to earn and fragile to maintain. 5. A lack of agency or true empowerment – With families having no ability to shape the way they are represented in data or see themselves in data, opportunities are missed to truly empower families to be better represented and to better themselves.

Through our explorations of a shared data interaction and personal data interfaces, we have found both a need and a desire for a new approach. We have shown that a model in which support services are deliberately open with families' data and bring it to the heart of their face-to-face consultations could address all five of these problems. The removal of the gatekeeper role would shift the power balance towards the family as it would give them a role in the stewardship of their own data. Providing families with a transparent view of stored data, and with clear visibility of data recording and usage, would enable accountability, which has previously been absent, which in turn can help to improve trust. With the family involved at every stage and able to see their data at any time, the consent problem would be largely solved – because families would be able to immediately speak up at any point should their wishes change in the light of new developments or new information. With the family becoming truly involved in data-informed support conversations that can make better decisions, and being more able to influence the way they are represented, they would be more empowered to make changes in their own lives and could achieve a previously unattainable level of agency. We have also uncovered additional benefits of a shared data interaction approach. Data visualisations and summaries could be very effective as conversation starters and as boundary objects, potentially leading to more effective conversations. The ability to reference specific data points over time can provide an objective measure against which to track progress – whose primary value is not to the support organisations (where they are used to measure service effectiveness) but in fact to the families themselves, who are now able to directly see the effects of their own actions in their data. The shift from support workers reporting what the data says to 'looking at data together' helps to shift the dynamic of the support interaction away from 'us and them' thinking towards a more collaborative approach and would be less adversarial. The inclusion of individuals in the stewardship of their own data would lead to more accurate data, because in reality the truth lies somewhere between what the data says and the family's own perspective, and can only emerge through dialogue. Individual family members will be able to notice mistakes or gaps, and contribute explanations, context or additional data to enrich the picture. By ensuring the discussions are based on more accurate data, the quality of decision-making would naturally improve and conversations will be likely to be more effective and efficient as they will be more grounded in reality. In particular, we have shown that giving the family a role could be very powerful, because the ability to contribute their own data or have visibility of data recording would provide them with direct evidence that they are being listened to and that their perspective is seen to matter more than 'what the computer says'. The ability to ask questions about their data, and to explain or clarify things seen in the data places more respect upon the family than the purely data-and-technology-based approach of the state-citizen service infrastructure experienced on the whole by non-supported families. The ability to act independently, in their own time and in contexts outside of the support interaction, allows individuals to alleviate concerns quickly and maintain confidence that their data selves, the version of themselves used by the state to inform decisions, remain fair and accurate, but also opens up new opportunities to individuals for using their data for their own ends in ways that were not previously possible. In exploring the usage of data in its full sociotechnical context, not just from the provider's perspective or citizen's perspective, we have shown that merely providing people with access to data would be insufficient to properly address the identified problems, and that in applying the principles of Human-Data Interaction we need to consider interaction in an interpersonal sense. Capabilities – or their absence – matter more than the nature of the data interaction. Data interfaces are limited by their operating context as to how much they can offer, but considering the wider human-facing relationship between the individual and the representative of the state allows us to imagine a more holistic solution that can better address any situation arising. It is vital that the human perspective be given the highest priority, so that professionals' flexibility is not limited, but also because data cannot adequately represent the complexities of human life – people are more than just data, and you have to talk to them to make sense of their lives and to avoid excluding them. The usage of data must always be supported with dialogue and engagement. It is the need to focus on the human aspect that

explains why trust underpinned nearly every single problem imagined by our participants – without an open system that encourages dialogue and discussion it is very hard not to close doors, create suspicion and harm trust. Through our sentence ranking exercises we have been able to offer a snapshot overview of what this sample of support workers and supported families think about data, and where they agree and disagree (see Table 3). Our detailed analysis of our workshop transcripts has provided an understanding of the positive and negative impacts on the support relationship of current civic data practices within early help, and through our qualitative analysis we have been able to identify [CHANGE]38 specific practices, many of which are currently imagined or only just emerging, which participants believe would improve families engagement and the support they receive (see Table 4). We believe these suggestions can serve as a challenge to the status quo that could inform policymakers attempting to reform care services or digital citizenship offerings. There would be significant challenges in adopting our proposed changes, in cost, training, manpower and emergency planning, as with any systemic practice change in an organisation, but our findings suggest that such an approach may get closer to the heart of the real issue of empowering ‘left-behind’ (disempowered) families than a purely state-centred approach to problem solving, and that this may offer part of a route to a more enabling welfare state. More generally our work serves as a reminder that as we move into the data-driven age it is important that data stays close to the people it is about, rather than those that use the data to provide services, and that service practice and processes should remain human-centric rather than data-centric. This is a general principle that could be equally applied to other domains including education, healthcare, democracy and commerce, and our emphasis upon individual capability over interface design is a useful mindset that can be applied to many human-computer interaction and design endeavours.

5 Case Study Two: The Human Experience of GDPR

[Target x words]

6 Bridge: An Understanding of Human Data Relations

This is not a data chapter, it unifies the findings of Chapter 4 and Chapter 5 in the context of RQ1 and RQ2 to provide a common set of findings.

[Target x words]

7 Case Study Three: Personal Data Interface

Design & Development

[Target x words]

8 Discussion: Designing Better Human Data Relations

[Target x words]

Bibliography

Abbattista, F. *et al.* (2007) 'Shaping personal information spaces from collaborative tagging systems', in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. (PART 3), pp. 728–735. doi: [10.1007/978-3-540-74829-8_89](https://doi.org/10.1007/978-3-540-74829-8_89).

Abiteboul, S., André, B. and Kaplan, D. (2015) *Managing your digital life with a Personal information management system*. 5. ACM, pp. 32–35. doi: [10.1145/2670528](https://doi.org/10.1145/2670528).

'About The Quantified Self' (no date). Available at: <https://quantifiedself.com/about/what-is-quantified-self/> (Accessed: 22 March 2021).

'About Us' (no date). datacy. Available at: <https://datacy.com/personal-about> (Accessed: 31 March 2021).

Abowd, G. D. (2012) 'What next, ubicomp?: celebrating an intellectual disappearing act', in *Proceedings of the 2012 acm conference on ubiquitous computing*. New York, New York, USA: ACM Press, pp. 31–40. doi: [http://dx.doi.org/10.1145/2370216.2370222](https://doi.org/10.1145/2370216.2370222).

Abowd, G. D. *et al.* (1999) 'Towards a better understanding of context and context-awareness', in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, pp. 304–307. doi: [10.1007/3-540-48157-5_29](https://doi.org/10.1007/3-540-48157-5_29).

Abowd, G. D. and Mynatt, E. D. (2000) 'Charting Past, Present, and Future Research in Ubiquitous Computing', *ACM Transactions on Computer-Human Interaction*, 7(1), pp. 29–58. doi: [10.1145/344949.344988](https://doi.org/10.1145/344949.344988).

Ackoff, R. L. (1989) 'From data to wisdom', *Journal of Applied Systems Analysis*, 16(1), pp. 3–9.

Adams, R. (2017) 'Michel Foucault: Discourse'. Available at: <https://criticallegalthinking.com/2017/11/17/michel-foucault-discourse/> (Accessed: 7 May 2021).

- 'AllofMe Company Profile' (2007). Available at: <https://www.crunchbase.com/organization/allofme> (Accessed: 23 March 2021).
- 'AllofMe.com Teaser Clip' (2008). YouTube. Available at: <https://www.youtube.com/watch?v=JWyt4WL6xE> (Accessed: 21 March 2021).
- Andrews, R. (2005) 'GTD : A New Cult for the Info Age', *Wired*. Available at: <https://www.wired.com/2005/07/gtd-a-new-cult-for-the-info-age/>.
- Apple (2009) 'iPhone 3G Commercial: "There's an app for that"'. YouTube. Available at: <https://www.youtube.com/watch?v=mFIITzqRBWY>.
- Aslam, H. *et al.* (2016) 'Harnessing Smartphones as a Personal Informatics Tool towards Self-Awareness and Behavior Improvement', *Proceedings - 2016 IEEE 14th International Conference on Dependable, Autonomic and Secure Computing, DASC 2016, 2016 IEEE 14th International Conference on Pervasive Intelligence and Computing, PICoM 2016, 2016 IEEE 2nd International Conference on Big Data*. IEEE, pp. 467–474. doi: [10.1109/DASC-PICoM-DataCom-CyberSciTec.2016.92](https://doi.org/10.1109/DASC-PICoM-DataCom-CyberSciTec.2016.92).
- Bannon, L. J. (1995) 'From Human Factors to Human Actors: The Role of Psychology and Human-Computer Interaction Studies in System Design', *Readings in Human-Computer Interaction*, pp. 205–214. doi: [10.1016/b978-0-08-051574-8.50024-8](https://doi.org/10.1016/b978-0-08-051574-8.50024-8).
- Barreau, D. K. (1995) 'Context as a factor in personal information management systems', *Journal of the American Society for Information Science*, 46(5), pp. 327–339. doi: [10.1002/\(SICI\)1097-4571\(199506\)46:5<327::AID-AS14>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1097-4571(199506)46:5<327::AID-AS14>3.0.CO;2-C).
- Barreau, D. and Nardi, B. A. (1995) 'Finding and reminding', *ACM SIGCHI Bulletin*, 27(3), pp. 39–43. doi: [10.1145/221296.221307](https://doi.org/10.1145/221296.221307).
- 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>.
- Beck, K. *et al.* (2001) 'The Agile Manifesto'. Available at: <http://agilemanifesto.org/>.
- Bell, G. and Gemmell, J. (2009) *Total recall: how the E-memory revolution will change everything*. Dutton (09), pp. 47–5062–47–5062. doi: [10.5860/choice.47-5062](https://doi.org/10.5860/choice.47-5062).
- Bergman, O. (2013) 'The Effect of Folder Structure on Personal File Navigation', *Journal of the American Society for Information Science and Technology*, 64(July), pp. 1852–1863. doi: [10.1002/asi](https://doi.org/10.1002/asi).
- Bergman, O., Beyth-Marom, R. and Nachmias, R. (2003) 'The user-subjective approach to personal information management systems', *Journal of the American Society for Information Science and Technology*, 54(9), pp. 872–878. doi: [10.1002/asi.10283](https://doi.org/10.1002/asi.10283).
- Bergman, O. *et al.* (2008) 'Improved search engines and navigation preference in personal information management', *ACM Transactions on Information Systems*, 26(4). doi: [10.1145/1402256.1402259](https://doi.org/10.1145/1402256.1402259).
- Bergman, O. *et al.* (2012) 'How do we find personal files?: The effect of OS, presentation & depth on file navigation', *Conference on Human Factors in Computing Systems - Proceedings*, pp. 2977–2980. doi: [10.1145/2207676.2208707](https://doi.org/10.1145/2207676.2208707).
- Bjerknes, G. *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.
- Boud, D., Keogh, R. and Walker, D. (1985) *Reflection: Turning experience into learning*. Routledge.
- Bowker, G. C. et al. (2015) *Boundary objects and beyond : working with Leigh Star*. MIT Press, p. 548. Available at: <https://books.google.co.uk/books?hl=en&lr=&id=nmSkCwAAQBAJ&oi=fnd&pg=PR5&dq=Boundary+Objects+and+Beyond:+Working+with+Leigh+Star+Objects+and+Beyond%3A+Working+with+Leigh+Star&f=false>.
- Bowyer, A. (2011) 'Why files need to die'. Available at: <http://radar.oreilly.com/2011/07/why-files-need-to-die.html>.
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1191/1478088706qp063oa).
- Brooks, D. (2013) 'The Philosophy of Data'. Available at: <https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html>.
- Brown, D. (2015) 'Here's what 'fail fast' really means'. Available at: <https://venturebeat.com/2015/03/15/heres-what-fail-fast-really-means/>.
- Brynjolfsson, E. and Oh, J. H. (2012) 'The attention economy: Measuring the value of free digital services on the internet', *International Conference on Information Systems, ICIS 2012*, 4, pp. 3243–3261.
- Bunge, M. (1999) *Social Science Under Debate: A Philosophical Perspective*. University of Toronto Press. Available at: <https://books.google.co.uk/books?id=-MLjZzJLbpkC>.
- Burkeman, O. (2011) 'SXSW 2011: The internet is over'. Available at: <https://www.theguardian.com/technology/2011/mar/15/sxsw-2011-internet-online> (Accessed: 23 March 2021).
- Bush, V. (1945) 'As we may think', *The Atlantic Monthly*, 3(2), pp. 35–46. doi: [10.2307/227181.227186](https://doi.org/10.2307/227181.227186).
- Bødker, S. (2006) 'When second wave HCI meets third wave challenges', *ACM International Conference Proceeding Series*, 189(October), pp. 1–8. doi: [10.1145/1182475.1182476](https://doi.org/10.1145/1182475.1182476).
- Bødker, S. (2015) 'Third-wave HCI, 10 years later—participation and sharing', *Interactions*, 22(5), pp. 24–31. doi: [10.1145/2804405](https://doi.org/10.1145/2804405).
- Campbell, P. L. (2011) 'Peirce, pragmatism, and the right way of thinking', *Sandia National Laboratories, Albuquerque*. Citeseer.
- Carter, J. (2015) 'Who are the digital disruptors redefining entire industries?' Available at: <https://www.techradar.com/uk/news/world-of-tech/who-are-the-digital-disruptors-redefining-entire-industries-1298171> (Accessed: 23 March 2021).
- Cavoukian, A. (2010) 'Privacy by design: the definitive workshop. A foreword by Ann Cavoukian, Ph.D', *Identity in the Information Society*, 3(2), pp. 247–251. doi: [10.1007/s12394-010-0062-y](https://doi.org/10.1007/s12394-010-0062-y).

- Cavoukian, A. (2012) 'Privacy by Design and the Emerging Personal Data Ecosystem', (October), pp. 1–39.
- 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](https://doi.org/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.
- Choe, E. K. *et al.* (2014) 'Understanding quantified-selfers' practices in collecting and exploring personal data', in *Proceedings of the 32nd annual acm conference on human factors in computing systems - chi '14*. New York, New York, USA: ACM Press, pp. 1143–1152. doi: [10.1145/2556288.2557372](https://doi.org/10.1145/2556288.2557372).
- Chung, C. F. *et al.* (2016) 'Boundary negotiating artifacts in personal informatics: Patient-provider collaboration with patient-generated data', *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 27, pp. 770–786. doi: [10.1145/2818048.2819926](https://doi.org/10.1145/2818048.2819926).
- Cogran, P. and Kinsley, S. (2012) 'Paying Attention: towards a critique of the attention economy', *Cultural Machine*, 13.
- Connected Health Cities (2017) 'SILVER Project: Smart Interventions for Local Residents'. Available at: <https://www.connectedhealthcities.org/research-projects/troubled-families/> (Accessed: 14 May 2021).
- Coughlan, T. *et al.* (2013) 'Methods for studying technology in the home', in *CHI'13 extended abstracts on human factors in computing systems*, pp. 3207–3210.
- Coughlan, T. *et al.* (2013) 'Current issues and future directions in methods for studying technology in the home', *PsychNology Journal*, 11(2), pp. 159–184.
- Crabtree, A. and Mortier, R. (2016) 'Personal Data, Privacy and the Internet of Things: The Shifting Locus of Agency and Control', *SSRN Electronic Journal*, pp. 1–20. doi: [10.2139/ssrn.2874312](https://doi.org/10.2139/ssrn.2874312).
- Crabtree, A. and Tolmie, P. (2018) 'The practical politics of sharing personal data', in *Personal and Ubiquitous Computing*. Springer-Verlag (2), pp. 293–315. doi: [10.1007/s00779-017-1071-8](https://doi.org/10.1007/s00779-017-1071-8).
- Crivellaro, C. *et al.* (2019) 'Not-equal: Democratizing research in digital innovation for social justice', *Interactions*, 26(2), pp. 70–73. doi: [10.1145/3301655](https://doi.org/10.1145/3301655).
- Croll, A. (2009) 'The Three Economies of Online Currency'. Available at: <https://solveforinteresting.com/the-three-currencies-of-the-online-economy/>.
- Ctrl-Shift (2014) 'Personal Information Management Services: An analysis of an emerging market'. Nesta, p. 38. Available at: <https://www.nesta.org.uk/report/personal-information-management-services-an-analysis-of-an-emerging-market/>.
- 'Data' (no date). Grammarist. Available at: <https://grammarist.com/usage/data/>.
- Decker, S. and Frank, M. (2004) 'The Networked Semantic Desktop', *WWW Workshop on Application Design, Development and Implementation Issues in the Semantic Web*. doi: [10.1108/eb057368](https://doi.org/10.1108/eb057368).
- 'Delicious' (2003). Available at: [https://en.wikipedia.org/wiki/Delicious_\(website\)](https://en.wikipedia.org/wiki/Delicious_(website)).
- 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. (1938) 'Experience and education'.
- Dewey, J. and Archambault, R. D. (1964) 'John Dewey on education: Selected writings'.
- Dey, A. K. (2000) *Providing Architectural Support for Building Context-Aware Applications*. PhD thesis.
- Dey, A. K. (2001) 'Understanding and using context', *Personal and ubiquitous computing*, pp. 4–7. Available at: <http://dl.acm.org/citation.cfm?id=593572>.
- 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](https://doi.org/10.7551/mitpress/8732.003.0007).
- Dourish, P. (2001) *Where the action is: the foundations of embodied interaction*. MIT press.
- Dourish, P. (2003) 'The appropriation of interactive technologies: Some Lessons From Placeless Documents', *Computer Supported Cooperative Work*, 12(4), pp. 465–490.
- Dourish, P. (2004) 'What we talk about when we talk about context', *Personal and Ubiquitous Computing*, 8(1), pp. 19–30. doi: [10.1007/s00779-003-0253-8](https://doi.org/10.1007/s00779-003-0253-8).
- Dourish, P. et al. (2000) 'Extending document management systems with user-specific active properties', *ACM Transactions on Information Systems*, 18(2), pp. 140–170. doi: [10.1145/348751.348758](https://doi.org/10.1145/348751.348758).
- Eliasson, J., Cerratto Pargman, T. and Ramberg, R. (2009) 'Embodied interaction or context-aware computing? An integrated approach to design', in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. Springer, Berlin, Heidelberg (PART 1), pp. 606–615. doi: [10.1007/978-3-642-02574-7_68](https://doi.org/10.1007/978-3-642-02574-7_68).
- Engelbart, D. C. (1962) 'Augmenting human intellect: A conceptual framework'. Menlo Park, CA, USA: Stanford Research Institute.
- Etzel, B. (1995) 'New strategy and techniques to cope with information overload', in *IEE colloquium on information overload*. IEE (223), pp. 2–2. doi: [10.1049/ic:19951427](https://doi.org/10.1049/ic:19951427).
- European Union Agency for Fundamental Rights (2020) 'Your Rights Matter: Data Protection and Privacy 2020', p. 20. doi: [10.2811/031862](https://doi.org/10.2811/031862).
- 'Facebook–Cambridge Analytica Data Scandal' (2014). Available at: https://en.wikipedia.org/wiki/Facebook–Cambridge_Analytica_data_scandal.
- Feng, Y. and Agosto, D. E. (2019) 'Revisiting personal information management through information practices with activity tracking technology', *Journal of the Association for Information Science and Technology*, 70(12), pp. 1352–1367. doi: [10.1002/asi.24253](https://doi.org/10.1002/asi.24253).
- 'Finland: Broadband Access Made Legal Right In Landmark Law' (2010). Available at: https://www.huffpost.com/entry/finland-broadband-access_n_320481 (Accessed: 23 March 2021).
- Firth, E. (2019) 'Personal data has value in so many different ways'. digi.me. Available at: <https://blog.digi.me/2019/09/04/personal-data-has-so-much-more-value-than-pure-cash/>.
- Foulonneau, M. and Riley, J. (2008) *Metadata for digital resources : implementation, systems design and interoperability*. Chandos Pub, p. 203.
- Fowler, M. and Highsmith, J. (2001) 'The agile manifesto', *Software Development*. [San Francisco, CA: Miller Freeman, Inc., 1993-, 9(8), pp. 28–35.
- Freeman, E. and Gelernter, D. (1996) 'Lifestreams: A Storage Model for Personal Data', *SIGMOD Record*

(ACM Special Interest Group on Management of Data). Association for Computing Machinery (ACM), 25(1), pp. 80–86. doi: [10.1145/381854.381893](https://doi.org/10.1145/381854.381893).

Friedman, B. and Hendry, D. G. (2019) *Value Sensitive Design: Shaping Technology with Moral Imagination*. MIT Press (The mit press). Available at: <https://books.google.co.uk/books?id=8ZlWDwAAQBAJ>.

Friedman, R. L. (2006) ‘Deweyan Pragmatism’, *William James Studies*, 1. Available at: <https://williamjamesstudies.org/deweyan-pragmatism/>.

Frost, A. (2019) ‘Forget Folders: The Best Ways to Organize Your Files with Tags and Labels’. Available at: <https://zapier.com/blog/how-to-use-tags-and-labels/>.

Fu, S. et al. (2020) ‘Social media overload, exhaustion, and use discontinuance: Examining the effects of information overload, system feature overload, and social overload’, *Information Processing and Management*, 57(6). doi: [10.1016/j.ipm.2020.102307](https://doi.org/10.1016/j.ipm.2020.102307).

Gelernter, D. (1994) ‘The cyber-road not taken: Lost on the info-highway? Here’s some stuff that could really change your life.’, *The Washington Post*, 3.

Gemmell, J., Bell, G. and Lueder, R. (2006) ‘MyLifeBits: A personal database for everything’, *Communications of the ACM*, 49(1), pp. 88–95. doi: [10.1145/1107458.1107460](https://doi.org/10.1145/1107458.1107460).

Gillespie, T. and Seaver, N. (2016) ‘Critical Algorithm Studies - A Reading List’. Available at: <https://socialmediacollective.org/reading-lists/critical-algorithm-studies/>.

Gitelman, L. (2013) *Raw data is an oxymoron*. Edited by Lisa Gitelman. MIT Press, p. 182. Available at: <https://mitpress.mit.edu/books/raw-data-oxymoron>.

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](https://doi.org/10.1145/1858171.1858189).

‘Google Desktop Search’ (2004). Available at: <https://en.wikipedia.org/wiki/Google/Desktop>.

Guba, E. G. (1990) ‘The alternative paradigm dialog’, *The paradigm dialog*. Sage Publications, Inc, pp. 17–30. Available at: <http://www.jstor.org/stable/3340973>.

Gurstein, M. (2003) ‘Effective use: A community informatics strategy beyond the digital divide’, *First Monday*, 8(12). doi: [10.5210/fm.v0i0.1798](https://doi.org/10.5210/fm.v0i0.1798).

Gurstein, M. B. (2011) ‘Open data: Empowering the empowered or effective data use for everyone?’, *First Monday*. First Monday, 16(2). doi: [10.5210/fm.v16i2.3316](https://doi.org/10.5210/fm.v16i2.3316).

Harris, T. (2013a) ‘A Call to Minimize Distraction Respect Users’ Attention’. Available at: <http://www.minimizedistraction.com/>.

Harris, T. (2013b) ‘Who We Are: Center for Humane Technology (CHT)’. Available at: <https://www.humanetech.com/who-we-are>.

Harris, T. (2016) ‘How Technology Hijacks People’s Minds — from a Magician and Google’s Design Ethicist’. Available at: <https://www.tristanharris.com/2016/05/how-technology-hijacks-peoples-minds-from-a-magician-and-googles-design-ethicist/> (Accessed: 22 March 2019).

Hart-Davidson, W., Zachry, M. and Spinuzzi, C. (2012) ‘Activity streams: Building context to coordinate writing activity in collaborative teams’, in *SIGDOC’12 - proceedings of the 30th acm international conference on design of communication*. New York, New York, USA: ACM Press, pp. 279–287. doi: [10.1145/2379057.2379109](https://doi.org/10.1145/2379057.2379109).

- 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](https://doi.org/10.1145/1993060.1993065).
- 'HDI Lab, Heerlen' (2020). Available at: <https://hdilab.com/>.
- 'HDI Network Plus, University of Glasgow' (2018). Available at: <https://hdi-network.org/>.
- Hemp, P. (2009) 'Death by Information Overload'. Available at: <https://hbr.org/2009/09/death-by-information-overload> (Accessed: 23 March 2021).
- Hendler, J. and Berners-Lee, T. (2010) 'From the Semantic Web to social machines: A research challenge for AI on the World Wide Web'. doi: [10.1016/j.artint.2009.11.010](https://doi.org/10.1016/j.artint.2009.11.010).
- Hixon, J. G. and Swann, W. B. (1993) 'When Does Introspection Bear Fruit? Self-Reflection, Self-Insight, and Interpersonal Choices', *Journal of Personality and Social Psychology*, 64(1), pp. 35–43. doi: [10.1037/0022-3514.64.1.35](https://doi.org/10.1037/0022-3514.64.1.35).
- 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.
- Hoofnagle, C. J., Sloot, B. van der and Borgesius, F. Z. (2019) 'The European Union general data protection regulation: What it is and what it means', *Information and Communications Technology Law*. Taylor & Francis, 28(1), pp. 65–98. doi: [10.1080/13600834.2019.1573501](https://doi.org/10.1080/13600834.2019.1573501).
- Hosch, W. L. (2017) 'Web 2.0'. Available at: <https://www.britannica.com/topic/Web-20> (Accessed: 26 April 2021).
- Hotho, A., Nürnberger, A. and Paaß, G. (2005) 'A brief survey of text mining.', in *Ldv forum*. Citeseer (1), pp. 19–62.
- Huberman, M. and Miles, M. B. (2002) *The qualitative researcher's companion*. Sage.
- 'Human Data Interaction Project at the Data to AI Lab, MIT' (2015). Available at: <https://hdi-dai.lids.mit.edu/>.
- Hutton, D. M. (2012) 'Turing's Cathedral: The Origins of the Digital Universe'. Emerald Group Publishing Limited.
- Hwang, E. (2021) 'Sketching Dialogue : Incorporating Sketching in Emphatic Semi-structured Interviews for HCI'.
- 'Information' (no date). Available at: <https://en.wikipedia.org/wiki/Information>.
- Information Commissioner's Office (2014) 'Data controllers and data processors: what the difference is and what the governance implications are', p. 20. Available at: <https://ico.org.uk/for-organisations/guide-to-data-protection/introduction-to-data-protection/some-basic-concepts/>.

- Information Commissioner's Office (2018) 'Your data matters - Your rights'. Available at: <https://ico.org.uk/your-data-matters/>.
- 'Infovark Company Profile' (2007). Available at: <https://www.crunchbase.com/organization/infovark>.
- Jelly, M. (2021) 'The Mission'. ethi.me. Available at: <https://www.ethi.me/the-mission> (Accessed: 31 March 2021).
- Jenkins, H. (2006) *Convergence Culture: Where Old and New Media Collide*. New York, USA: New York University Press. doi: [10.7551/mitpress/9780262036016.003.0012](https://doi.org/10.7551/mitpress/9780262036016.003.0012).
- Jilek, C. et al. (2018) 'Context spaces as the cornerstone of a near-transparent and self-reorganizing semantic desktop', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11155 LNCS, pp. 89–94. doi: [10.1007/978-3-319-98192-5_17](https://doi.org/10.1007/978-3-319-98192-5_17).
- Jones, T. (2011) 'Designing for second screens : The Autumnwatch Companion'. Available at: <https://www.bbc.co.uk/blogs/researchanddevelopment/2011/04/the-autumnwatch-companion---de.shtml>.
- Jones, W. (2011a) 'The Future of Personal Information Management Part I: Our Information, Always and Forever'.
- Jones, W. (2011b) 'The Future of Personal Information Management Part I: Our Information, Always and Forever', p. 72.
- Karger, D. R. et al. (2005) 'Haystack: A customizable general-purpose information management tool for end users of semistructured data', in *2nd biennial conference on innovative data systems research, cidr 2005*, pp. 13–27. Available at: <https://s3.amazonaws.com/academia.edu.documents/46870765/haystack.pdf>.
- Karger, D. R. and Jones, W. (2006) 'Data unification in personal information management', *Communications of the ACM*, 49(1), p. 77. doi: [10.1145/1107458.1107496](https://doi.org/10.1145/1107458.1107496).
- Kelly, K. and Wolf, G. (2007) 'What is the quantified self'. Available at: <https://web.archive.org/web/20100507215130/http://www.kk.org/quantifiedself/2007/10/what-is-the-quantifiable-self.php>.
- Kelly, R. (2020) 'The Biggest ICO Fines Ever Issued'. Available at: <https://digit.fyi/data-protection-2020-the-biggest-fines-ever-issued-by-the-ico/>.
- Kensing, F. and Blomberg, J. (1998) 'Participatory design: Issues and concerns', *Computer supported cooperative work (CSCW)*. Springer, 7(3), pp. 167–185.
- Klein, B. et al. (2004) 'Enabling flow - A paradigm for document-centered personal information spaces', in *Proceedings of the eighth iasted international conference on artificial intelligence and soft computing*, pp. 187–192. Available at: <https://www.semanticscholar.org/paper/Enabling-flow%3A-%7BA%7D-paradigm-for-document-centered-Klein-Agne/22be4a7b25e75de235e5d96bad6ab4ab4583daac>.
- Krishnan, A. (2010) 'Pervasive Personal Information Spaces'. University of Waikato. Available at: <https://researchcommons.waikato.ac.nz/handle/10289/4590>.
- Krishnan, A. and Jones, S. (2005) 'TimeSpace: Activity-based temporal visualisation of personal information spaces', *Personal and Ubiquitous Computing*, 9(1), pp. 46–65. doi: [10.1007/s00779-004-0291-x](https://doi.org/10.1007/s00779-004-0291-x).
- Lansdale, M. and Edmonds, E. (1992) 'Using memory for events in the design of personal filing systems', *International Journal of Man-Machine Studies*, 36(1), pp. 97–126. doi: [10.1016/0020-7373\(92\)90054-O](https://doi.org/10.1016/0020-7373(92)90054-O).
- Lansdale, M. W. (1988) 'The psychology of personal information management', *Applied Ergonomics*, 19(March 1988), pp. 55–66. doi: [10.1016/0003-6870\(88\)90199-8](https://doi.org/10.1016/0003-6870(88)90199-8).

- Larsson, S. (2018) 'Algorithmic governance and the need for consumer empowerment in data-driven markets', *Internet Policy Review*, 7(2). doi: [10.14763/2018.2.791](https://doi.org/10.14763/2018.2.791).
- Leprince-Ringuet, D. (2021). Available at: <https://www.zdnet.com/article/gdpr-fines-increased-by-40-last-year-and-theyre-about-to-get-a-lot-bigger/>.
- Levine, R. (2011) 'How the internet has all but destroyed the market for films, music and newspapers'. Available at: <https://www.theguardian.com/media/2011/aug/14/robert-levine-digital-free-ride> (Accessed: 23 March 2021).
- Lewin, K. (1946) 'Action Research and Minority Problems', *Journal of Social Issues*, 2(4), pp. 34–46. doi: [10.1111/j.1540-4560.1946.tb02295.x](https://doi.org/10.1111/j.1540-4560.1946.tb02295.x).
- Lewin, K. (1951) 'Problems of research in social psychology', *Field theory in social science: Selected theoretical papers*, pp. 155–169.
- Li, I. (2009) 'Designing Personal Informatics Applications and Tools that Facilitate Monitoring of Behaviors', *Uist*. Available at: <http://citeseerv.ist.psu.edu/viewdoc/summary?doi=10.1.1.232.8536>.
- Li, I., Dey, A. and Forlizzi, J. (2010) 'A stage-based model of personal informatics systems', *Proceedings of the 28th international conference on Human factors in computing systems CHI 10*. New York, New York, USA: ACM Press, p. 557. doi: [10.1145/1753326.1753409](https://doi.org/10.1145/1753326.1753409).
- Lindley, S. E. et al. (2018) 'Exploring new metaphors for a networked world through the file biography', *Conference on Human Factors in Computing Systems - Proceedings*, 2018-April, pp. 1–12. doi: [10.1145/3173574.3173692](https://doi.org/10.1145/3173574.3173692).
- Malone, T. W. (1983) 'How do people organize their desks?: Implications for the design of office information systems', *ACM Transactions on Information Systems*, 1(1), pp. 99–112. doi: [10.1145/357423.357430](https://doi.org/10.1145/357423.357430).
- Mannay, D. and Morgan, M. (2015) 'Doing ethnography or applying a qualitative technique? Reflections from the 'waiting field'', *Qualitative research*. Sage Publications Sage UK: London, England, 15(2), pp. 166–182.
- Marshall, C. C. and Jones, W. (2006) 'Keeping encountered information', *Communications of the ACM*, 49(1), pp. 66–67. doi: [10.1145/1107458.1107493](https://doi.org/10.1145/1107458.1107493).
- McCarthy, J. and Wright, P. (2004) 'Technology as experience', *Interactions*, 11(5), pp. 42–43. doi: [10.1145/1015530.1015549](https://doi.org/10.1145/1015530.1015549).
- Millar, S. (2002) 'UK singled out for criticism over protection of privacy'. Available at: <https://www.theguardian.com/technology/2002/sep/05/security.humanrights>.
- 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](https://doi.org/10.1145/1240624.1240832).
- Mortier, R. et al. (2013) 'Challenges & opportunities in human-data interaction', *University of Cambridge, Computer Laboratory*. Citeseer. doi: [10.5210/fm.v17i5.4013](https://doi.org/10.5210/fm.v17i5.4013).
- Mortier, R. et al. (2014) 'Human-data interaction: The human face of the data-driven society', Available at SSRN 2508051. doi: [10.2139/ssrn.2508051](https://doi.org/10.2139/ssrn.2508051).
- Murton, D. (2011) 'A Brief History of the Evolution of Social Technology'. Available at: <https://www.scottmonty.com/2011/04/brief-history-of-evolution-of-social.html>.
- MyData (2017) 'Declaration - MyData.org'. Available at: <https://mydata.org/declaration/> (Accessed: 8 November 2019).

- 'MyData Comparison of Principles document' (2017). Available at: <http://bit.ly/pd-principles>.
- MyData.org (2018) 'MyData - Who we are'. Available at: <https://mydata.org/about/>.
- Mydex CIC (2010) 'The Case for Personal Information Empowerment : The rise of the personal data store', *World*, pp. 1–44.
- 'myTimeline' (2018). Available at: <https://www.timelineinc.com/> (Accessed: 23 March 2021).
- Nadeem, D. and Sauermann, L. (2007) 'From Philosophy and Mental-Models to Semantic Desktop Research: Theoretical Overview'.
- Neef, D. (2015) *Digital exhaust: what everyone should know about big data, digitization and digitally driven innovation*. Pearson Education.
- Negroponte, N. and Bolt, R. A. (1978) *Spatial data management system*. MASSACHUSETTS INST OF TECH CAMBRIDGE ARCHITECTURE MACHINE GROUP.
- Nelson, T. (2006) 'Lost in hyperspace', *New Scientist*, 191(2561). doi: [10.1002/elsc.200620112](https://doi.org/10.1002/elsc.200620112).
- Nelson, T. H. (1965) 'Complex information processing', pp. 84–100. doi: [10.1145/800197.806036](https://doi.org/10.1145/800197.806036).
- Norman, D. A. and Draper, S. W. (1986) 'User Centered System Design; New Perspectives on Human-Computer Interaction'. L. Erlbaum Associates Inc.
- Odom, W. et al. (2018) 'Time, Temporality, and Slowness', pp. 383–386. doi: [10.1145/3197391.3197392](https://doi.org/10.1145/3197391.3197392).
- O'Donnell, B. (2020) 'Zoom, the office and the future: What will work look like after coronavirus?' Available at: <https://eu.usatoday.com/story/tech/columnist/2020/09/07/zoom-work-from-home-future-office-after-coronavirus/5680284002/>.
- O'Donoghue, T. and Rabin, M. (2001) 'Choice and procrastination', *The Quarterly Journal of Economics*. MIT Press, 116(1), pp. 121–160.
- Organisation for Economic Co-operation and Development (1980) *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. Available at: <https://www.oecd.org/digital/ieconomy/oecdguidelinesonthe protectionofprivacyandtransborderflowsofpersonald ata>
- 'Our Values' (no date). Available at: <https://www.citizenme.com/about/our-values> (Accessed: 31 March 2021).
- 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>.
- Perez, S. (2018) 'Facebook is shutting down Friend List Feeds'. Available at: <https://techcrunch.com/2018/08/09/facebook-is-shutting-down-friend-list-feeds-today/>.
- 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.
- Pór, G. (1997) 'Designing Knowledge Ecosystems for Communities of Practice', in *Advancing organizational capability via knowledge management*.
- Price Ball, M. (no date) 'About Us'. Available at: <https://www.openhumans.org/about/> (Accessed: 31 March 2021).
- Puussaar, A., Clear, A. K. and Wright, P. (2017) 'Enhancing Personal Informatics Through Social Sensemaking', *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI*

'17. Association for Computing Machinery, 2017-May, pp. 6936–6942. doi: [10.1145/3025453.3025804](https://doi.org/10.1145/3025453.3025804).

Raskin, J. (2000) *The humane interface: new directions for designing interactive systems*. Addison-Wesley Professional.

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.

Rivera-Pelayo, V. et al. (2012) 'A framework for applying Quantified Self approaches to support reflective learning', *Proceedings of the IADIS International Conference Mobile Learning 2012, ML 2012*, pp. 123–131.

Roche, M. (2011) 'Full internet ban for sex offenders ruled unlawful'. Available at: <https://ukhumanrightsblog.com/2011/08/12/full-internet-ban-for-sex-offenders-ruled-unlawful/> (Accessed: 23 March 2021).

Rogers, Y. (2006) 'Moving on from Weiser's Vision of Calm Computing: Engaging UbiComp Experiences', *LNCS*, 4206, pp. 404–421. Available at: [http://www.inf.ufg.br/\\$sim\\$vagner/courses/mobilecomputing/docs/papers/03-Rogers_Ubicomp06.pdf](http://www.inf.ufg.br/simvagner/courses/mobilecomputing/docs/papers/03-Rogers_Ubicomp06.pdf).

Ross, G. (2005) 'An introduction to Tim Berners-Lee's Semantic Web'. Available at: <https://www.techrepublic.com/article/an-introduction-to-tim-berners-lees-semantic-web/>.

Saha, D. and Mukherjee, A. (2003) 'Pervasive computing: A paradigm for the 21st century'. IEEE. doi: [10.1109/MC.2003.1185214](https://doi.org/10.1109/MC.2003.1185214).

Sauermann, L., Bernardi, A. and Dengel, A. (2005) 'Overview and outlook on the semantic desktop', in *CEUR workshop proceedings*.

Schumacher, K., Sintek, M. and Sauermann, L. (2008) 'Combining fact and document retrieval with spreading activation for semantic desktop search', in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, pp. 569–583. doi: [10.1007/978-3-540-68234-9_42](https://doi.org/10.1007/978-3-540-68234-9_42).

Searls, D. (2008) 'The Intention Economy: What Happens When Customers Get Real Power'. Available at: https://web.archive.org/web/20101226073246/http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/2009_05/searls_intention_economy.pdf

Searls, D. (2012) *The intention economy: when customers take charge*. Harvard Business Press (04), pp. 50–2168–50–2168. doi: [10.5860/choice.50-2168](https://doi.org/10.5860/choice.50-2168).

Seligman, C. and Darley, J. M. (1976) 'Feedback as a means of decreasing residential energy consumption', *Journal of Applied Psychology*, 62(4), pp. 363–368. doi: [10.1037/0021-9010.62.4.363](https://doi.org/10.1037/0021-9010.62.4.363).

Shannon, C. E. (1948) 'A mathematical theory of communication', *The Bell system technical journal*. Nokia Bell Labs, 27(3), pp. 379–423.

Shilton, K. (2011) 'Participatory Personal Data: An Emerging Research Challenge for the Information Sciences', *Advances in Information Science*.

Shipman, F. M. (. and Marshall, C. C. (1999) 'Formality Considered Harmful : Experiences , Emerging Themes , and Directions on the Use of Formal Representations in Interactive Systems', pp. 333–352.

Shneiderman, B. (1996) *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualisations*.

Siegel, D. (2009) *Pull: The power of the semantic web to transform your business*. Penguin.

Siegel, D. (2010) 'Personal Data Locker Vision Video'. Available at: <https://vimeo.com/14061238>.

Siegler, M. G. (2011) 'Facebook Unveils Timeline: The Story Of Your Life On A Single Page'. Available at:

- <https://techcrunch.com/2011/09/22/facebook-timeline/> (Accessed: 21 March 2021).
- Simon, H. A. and Newell, A. (1958) 'Heuristic Problem Solving: The next advance in operations research'. doi: [10.1057/978-1-349-94848-2_792-1](https://doi.org/10.1057/978-1-349-94848-2_792-1).
- 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](https://doi.org/10.1080/15710882.2017.1310466).
- Spencer, D. and Warfel, T. (2004) 'Card sorting: A definitive guide', *Boxes and arrows*, 2(2004), pp. 1–23.
- Spiekermann, S. and Korunovska, J. (2017) 'Towards a value theory for personal data', *Journal of Information Technology*, 32(1), pp. 62–84. doi: [10.1057/jit.2016.4](https://doi.org/10.1057/jit.2016.4).
- 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](https://doi.org/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](https://doi.org/10.1177/0162243910377624).
- Steinberg, S. G. (1997) 'Lifestreams', *Wired*. Available at: <https://www.wired.com/1997/02/lifestreams/>.
- Symons, T. et al. (2017) 'Me, my data and I: The future of the personal data economy', *DECODE (DEcentralised Citizen Owned Data Ecosystems) Report*, (732546), p. 88. Available at: <https://media.nesta.org.uk/documents/decode-02.pdf>.
- Taylor, L. (2017) 'What is data justice? The case for connecting digital rights and freedoms globally', *Big Data and Society*, 4(2). doi: [10.1177/2053951717736335](https://doi.org/10.1177/2053951717736335).
- Teevan, J. et al. (2004) 'The perfect search engine is not enough: A study of orienteering behavior in directed search', in *Conference on human factors in computing systems - proceedings*, pp. 415–422. Available at: <http://people.csail.mit.edu/teevan/work/publications/papers/chi04.pdf>.
- Teevan, J. B. (2001) 'Displaying dynamic information', in *Conference on human factors in computing systems - proceedings*, pp. 417–418. doi: [10.1145/634067.634311](https://doi.org/10.1145/634067.634311).
- Terdiman, D. (2008) 'Using tags to improve the Flickr experience'. Available at: <https://www.cnet.com/news/using-tags-to-improve-the-flickr-experience/>.
- The European Parliament and the Council of the European Union (2016) 'Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data', pp. 16–32. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679> <https://eur-lex.europa.eu/legal-content/EN/PDF/?uri=CELEX:32016R0679&from=ES>.
- 'The GDPR: Does it Benefit Consumers in Any Practical Way?' (2020). Atebits.com. Available at: <https://www.atebits.com/the-gdpr-does-it-benefit-consumers-in-any-practical-way/>.
- 'The personal computer revolution' (no date) in *Britannica*. Available at: <https://www.britannica.com/technology/computer/The-personal-computer-revolution>.
- Timely (2020) 'The attention economy: what it is, what it's doing to you'. Available at: <https://memory.ai/timely-blog/the-attention-economy>.

- Toonders, J. (2014) 'Data Is the New Oil of the Digital Economy'. Available at: <https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/>.
- Tufekci, Z. (2017) 'We're building a dystopia just to make people click on ads'. TED. Available at: https://www.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads.
- Tunikova, O. (2018) 'Are We Consuming Too Much Information?' Available at: https://medium.com/@tunikova_k/are-we-consuming-too-much-information-b68f62500089 (Accessed: 23 March 2021).
- US Department of Health Education and Welfare (1973) 'Records Computers and the Rights of Citizens'.
- Various Authors (2018) 'Our Digital Lives', in *TED talks*. TED. Available at: https://www.ted.com/playlists/26/our_digital_lives.
- 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](https://doi.org/10.1145/2851581.2886436).
- Wagner, A. (2012) 'Is internet access a human right?' Available at: <https://www.theguardian.com/law/2012/jan/11/is-internet-access-a-human-right> (Accessed: 23 March 2021).
- Wallace, D. P. (2007) *Knowledge management: Historical and cross-disciplinary themes*. Libraries unlimited.
- Weiser, M. (1991) 'The computer for the 21st century', *Scientific American*, 265(3), pp. 94–105. doi: [10.1145/329124.329126](https://doi.org/10.1145/329124.329126).
- Weiser, M. and Brown, J. S. (1996) 'The coming age of calm technology', *Beyond Calculation: The Next Fifty Years of Computing*. Available at: <http://www.teco.edu/lehre/ubiq/ubiq2000-1/calmtechnology.htm> http://link.springer.com/content/pdf/10.1007/978-1-4612-0685-9_6.pdf%5Cnpapers2://publication/uuid/F86D6ECE-A71E-4D20-A47B-9AF86A84923D.
- Wellisch, H. H. (1996) *Abstracting, indexing, classification, thesaurus construction: A glossary*. American Society of Indexers.
- Whittaker, S. and Hirschberg, J. (2001) 'The Character, Value, and Management of Personal Paper Archives', *ACM Transactions on Computer-Human Interaction*, 8(2), pp. 150–170. doi: [10.1145/376929.376932](https://doi.org/10.1145/376929.376932).
- 'Whose data is it anyway?' (2019). 04: UBDI. Available at: <https://www.ubdi.com/blog/whose-data-is-it-anyway> (Accessed: 31 March 2021).
- Wilson, L., Wilson, R. and Martin, M. (2020) *Health and Care Practitioner Insights: Understanding Information Sharing in Constellations of Care - Report on Amy's Page workshop series*. Great North Care Record. Available at: <https://www.greatnorthcarerecord.org.uk>.
- 'WinFS' (no date). Available at: <https://en.wikipedia.org/wiki/WinFS>.
- 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](https://doi.org/10.1007/978-3-030-73426-8_22).
- Zichichi, M., Ferretti, S. and D'Angelo, G. (2020) 'On the Efficiency of Decentralized File Storage for Personal Information Management Systems'. Available at: <http://arxiv.org/abs/2007.03505>.
- Zins, C. (2015) 'What is the meaning of "data", "information", and "knowledge"?' , *Institute of Knowledge Sharing*, 3(1).
- Ziogas, G. (2020) 'The Inventor of the World Wide Web Says the Internet Is Broken'. Available at:

<https://medium.com/digital-diplomacy/the-inventor-of-the-world-wide-web-says-the-internet-is-broken-fbce1c8bf6cf>.

Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile. Available at: <https://books.google.co.uk/books?id=W7ZEDgAAQBAJ>.

Zúñiga, H. de, Garcia-Perdomo, V. and McGregor, S. C. (2015) ‘What is second screening? Exploring motivations of second screen use and its effect on online political participation’, *Journal of communication*. Oxford University Press, 65(5), pp. 793–815.

-
1. The usage of the abbreviation PIMS here is not to be confused with its use to refer to “Personal Information Management Systems” in traditional PIM terminology.[↩](#)
 2. Note that Case Study Three (Designing Personal Data Interfaces) involved no participants which is why it does not have its own table in this section.[↩](#)
 3. One participant withdrew from the study after the first interview of the Guided GDPR study due to COVID-19. The other 10 participants took part in all three stages.[↩](#)
 4. (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.)[↩](#)
 5. The term ‘troubled families’, popularised by the TFP, has fallen from use, as it was considered to be negative and judgemental. A latter term ‘vulnerable families’ has also been criticised for being disempowering. Most councils now refer simply to ‘families’ or sometimes ‘supported families’, and the rest of this thesis adopts this convention.[↩](#)
 6. Some leisure categories (namely Shopping and Transport) were included that are not strictly civic data, as these are useful for exploring issues of ethics and helping participants to have a reference point when discussing the “big data” benefits of data linking.[↩](#)
 7. The first of these interviews was a ‘trial run’ with a couple selected by convenience sample, and conducted in a University meeting room not their home at the participants’ request.[↩](#)
 8. The notation used for the quote references is as follows:
 - FQnn = Family Quote - a quote from the families-only workshop (A)
 - SQnn = Staff Quote - a quote from a staff-only workshop (B)
 - CQnn = Combined Quote - a quote from the combined workshop (C).
 - Sn = Sentence n - a sentence from the Sentence Ranking exercise, see [INSERT REF TO SENTENCES IN APPENDIX]

The number after FQ/CQ/SQ provides a unique identifier for each quote, which can be used to look up the referenced quote in [INSERT REF TO APPENDIX SECTION HERE]. Individual speakers are identified only by their role: within each quote, or in brackets afterwards, the speakers are identified as Worker, Parent, Child, or Researcher.[↩](#)