

Understanding and Improving Human Data Relations

Alex Bowyer

Contents

Frontmatter	5
Abstract	5
Acknowledgements	6
Lists of Tables and Figures	6
Lists of Tables by chapter	6
Tables in Chapter N	6
Lists of Figures by chapter	6
Figures in Chapter N	6
1 Introduction	6
1.1 Background and Motivation for this Research	6
1.1.1 Personal motivation and context	7
1.1.2 Research Objectives and purpose	9
1.2 Nature and Contributions of the thesis	10
1.2.1 An Understanding of what People want from Personal Data	10
1.2.2 Establishing a new field: Human Data Relations	11
1.2.3 Additional contributions in the Early Help and Civic Data Use context	13
1.2.4 Additional contributions in the context of GDPR and Everyday Data Access	13
1.3 Publications arising from and connected to this research	15
1.3.1 Pilot Study	15
1.3.2 Primary Case studies	15
1.3.3 Workshop papers & presentations	16
1.3.4 Publications from other work	18
1.4 The structure of this thesis	18
2 Literature Review	20
2.1 Understanding Data and its Role in Today's World	21
2.1.1 What is Data?	21

2.1.2	Personal Data & the Rise of Data-centrism	22
2.1.3	Legislating to Protect Personal Data: the Story so far	23
2.1.4	The Need for Practical and Effective Personal Data Access	25
2.1.5	Research Gap: The Human Experience of Personal Data	27
2.2	A History of Personal Data Interaction	27
2.2.1	Computers as General-Purpose Information Tools	27
2.2.2	Personal Information Management	28
2.2.3	Self Informatics	32
2.2.4	Overwhelmingly Complex Digital Lives	34
2.2.5	Research Gap: The Data Beyond The Individual	36
2.3	An Alternative to Data-centricism: Human-centricity	36
2.3.1	Human-Computer Interaction (HCI) Foundations	36
2.3.2	Data Transcendence, Context and Human-Data Interaction (HDI)	37
2.3.3	Human-Centred Design: A Sociotechnical Challenge	38
2.3.4	The Pursuit of Human-centred Personal Data Ecosystems (MyData and beyond)	39
2.3.5	Research Gap: Defining the Research Agenda for Achieving Human-Centricity in Practice	41
3	Methodology	43
3.1	Forming a Research Paradigm: Ontology & Epistemology	43
3.2	Research Approach: Participatory Action Research & Experience-centred Design	45
3.3	Research Objectives	50
3.3.1	Research Question 1 (RQ1): What is the human experience of personal data, and what do people want from their data?	50
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?	51
3.3.3	Research Question 3 (RQ3): What challenges and opportunities are relevant when attempting to establish these ideals for human data relations?	52
3.4	Overview of Research Contexts and Activities	53
3.4.1	Context One: Civic Data Use and Access to Data in the Early Help Context	54
3.4.2	Context Two: Accessing the Personal Data from your Digital Life using GDPR	56
3.4.3	Context Three: The Practical Pursuit of Better Human Data Relations	58
3.5	Methodologies Employed in Case Studies	60
3.5.1	Sensitisation of Researcher and Participants	61
3.5.2	Discussion and Exploration with Stimuli	66
3.5.3	Participatory Co-Design of Possible Solutions	68
3.5.4	Practical Data Experiments, Interface Design and Prototyping	70

3.5.5	Analysis, Modelling and Learning	73
3.5.6	Recruitment	75
3.5.7	Ethics	80
3.5.8	Thesis structure approach	81
4	Case Study One: Accessing and Using Civic Data in Early Help	81
4.1	Context: Data Use in Early Help	82
4.1.1	Data-centric Family Intervention in UK Social Care	82
4.1.2	Current Practice: Early Help Case Records as a Source of Truth	83
4.1.3	Rethinking the Role of Data in Early Help Support Relationships	85
4.2	Preliminary Explorations of Family Civic Data: Families' and Support Workers' Perspectives	86
4.2.1	What is Family Civic Data?	86
4.2.2	What is the Family Perspective on Their Civic Data and its Use?	89
4.2.3	What is the Staff Perspective on Family Civic Data and its Use?	89
4.2.4	Seeking an Equitable and Mutually Beneficial Data Use Model: Shared Data Interaction	90
4.2.5	The Workshops	90
4.2.6	Sensitising Participants and Discovering Shared Values	92
4.3	Thematic Findings	93
4.3.1	Themes & Subthemes	95
4.3.2	Theme 1: Meaningful Data Interaction for families	100
4.3.3	Theme 2: Giving a Voice to the Family	103
4.3.4	Theme 3: Earning Trust through Transparency	107
4.4	Discussion	110
4.4.1	The Value of Involving People With Their Data	111
4.4.2	Effective Data Access Requires Human Interaction	113
4.4.3	The Implications of Shifting the Locus of Decision-making Through Shared Data Interaction	116
4.5	Summation	119
5	Case Study Two: The Human Experience of GDPR	122
5.1	Context: Accessing Your Personal Data Using Your GDPR Rights	124
5.1.1	The Current Need for Data Access	124
5.1.2	Current GDPR Research and its Limitations	125
5.1.3	Human-Data Interaction: Towards a Human-centric Personal Data Ecosystem	126
5.2	Study Design and Configuration	127
5.3	GDPR Request Outcomes	130
5.3.1	Interview 1: GDPR Target Selection	130
5.3.2	Interview 2: Privacy Policy Review and Goal Setting	132
5.3.3	Interview 3: Reviewing the GDPR Response	133

5.3.4	Perceived Power and Trust	136
5.4	Thematic Findings	138
5.4.1	Themes & Subthemes	138
5.4.2	Theme 1: Many Companies are Evasive and People are “Still in the Dark”	143
5.4.3	Theme 2: People Struggle to Understand, Use and Control Their Data	145
5.4.4	Theme 3: Poor GDPR Handling Can Damage the Forced & Fragile Trust Relationship	150
5.5	Discussion	155
5.5.1	Implications for Policymakers: Compliance, Quality and Ongoing Access	155
5.5.2	Implications for Data Holders: Earn Trust by Opening Up Data and Enabling Users	157
5.5.3	Implications for Individuals: Becoming Aware of the Value and Power of Data, and Demanding More	158
5.6	Summation	159
6	Discussion I: An Understanding of Human Data Relations	162
6.1	Answering RQ1: What do people want in <i>direct</i> data relations?	163
6.1.1	Visible	163
6.1.2	Understandable	167
6.1.3	Useable	171
6.2	Answering RQ2: What do people want in <i>indirect</i> data relations?	174
6.2.1	Process Transparency	174
6.2.2	Individual Oversight	178
6.2.3	Involvement	181
6.3	Achieving Individual Empowerment	185
7	Discussion II: Designing and Pursuing Better Human Data Relations	187
7.1	Introduction & Background	187
7.1.1	External Research Settings	189
7.1.2	Attribution of Insights	189
7.1.3	Theories of Change	190
7.1.4	Better Human Data Relations as a Recursive Public	192
7.2	Refining and Defining ‘Human Data Relations’ (HDR)	193
7.2.1	The Role of Personal Data	193
7.2.2	Human Data Interaction or Human Information Interaction?	194
7.2.3	The Two Distinct Motivations for Human Data Relations	195
7.3	The landscape of opportunity for better HDR, with its obstacles and resulting insights	198
7.3.1	Data Awareness & Understanding	199
7.3.2	Data Usability	199
7.3.3	Ecosystem transparency	202
7.3.4	Ecosystem negotiability	204

7.3.5	Obstacles in the Solution Space	208
7.4	Four approaches to improving HDR, and their specific opportunities	212
7.4.1	Learn & Discover	214
7.4.2	Defending the Status Quo and Pushing for Compliance .	215
7.4.3	Create New Structures and Systems	216
7.4.4	Create New Information Interfaces	226
7.4.5	Create New Capabilities	229
7.4.6	Influence & Motivate	231
7.4.7	Educate & Empower	231
7.5	Thesis Conclusion	231

Bibliography	231
--------------	-----

Frontmatter

Abstract

Technologies including PCs, smartphones, and cloud computing have transformed the world: In our daily lives we interact with many businesses and public services who (to reduce costs) increasingly seek to rely on data collection and processing rather than face-to-face user interactions to inform their decisions. This creates an *imbalance of power* between those who hold data and the individuals about whom data is stored, who cannot easily see their personal data or how it is used. This *Digital Civics* PhD research explores, from a pragmatic, constructivist perspective, the topic of *Human Data Relations*. Through two qualitative case studies across public and private sectors, it answers the question, “*What relationship do people need with their personal data?*”. Case Study One focuses on *Early Help* social care: Through four workshops with supported families, social workers and staff, a deep understanding of the individual perspective on civic personal data use is established. *Shared data interaction* is explored as a means to shift the balance of power towards the individual while maintaining an effective care relationship. Case Study Two is a three-month study exploring 10 participants’ experience of using *GDPR data access rights* to view their own data, resulting in insights into individual needs and the challenges of data-centric service relationships, and recommendations for improvement of policies and practices. With reference to literature from the fields of *Personal Information Management*, *Human Data Interaction* and *MyData* personal data ecosystems, these case studies contribute to a unified understanding of *six core needs* that people have in Human Data Relations. In the final chapter, the thesis discusses the *practical pursuit* of these goals, drawing on first-hand knowledge acquired from expert participation in industrial research projects at BBC R&D and Hestia.ai/SITRA, mapping out the *landscape for future research and innovation*.

Acknowledgements

Lists of Tables and Figures

Lists of Tables by chapter

Tables in Chapter N

- Table N.n - Description goes here

Lists of Figures by chapter

Figures in Chapter N

- Figure N.n - Description goes here

1 Introduction

“My data is everywhere, and I am nowhere.” – Imogen Heap (musician and digital rights advocate), speaking at MyData 2019.

1.1 Background and Motivation for this Research

The collection, storage and exploitation of personal data about individuals has been a driving force in shaping the **data-centric** world we inhabit today. Every aspect of our lives involves the direct or indirect use of computer systems that attempt to **capture and represent individuals as data**, so that organisations can serve more customers at scale, by reducing costly human interactions and relying increasingly (often algorithmically or in aggregate) on the **interpretation of those datapoints to make decisions** - decisions that affect our everyday life in myriad ways, from determining eligibility to access particular services, benefits or products to choosing which advertisements or recommendations to serve to those individuals in an attempt to better persuade them to act. Data about people has become ‘*the new oil*’ (Toonders, 2014), the central resource underpinning the **surveillance capitalism** (Zuboff, 2019) driving profit for many businesses today. This splintering (Lemley, 2021) of our digital selves across hundreds of different organisations’ computer systems has not only created “*a chaos of multiplicity in terms of technologies, use situations, methods and concepts*” (Bødker, 2015) where data becomes **trapped** (Abiteboul, André and Kaplan, 2015) and hard for people to manage, but is **causing real harm both to individuals** - increasing overload, anxiety and distraction (Fu *et al.*, 2020; Timely, 2020) - **and to society** - manipulating our attention, radicalising people and distorting democracy (Thompson, 2011; Chan, 2019). There is **a power imbalance** “*in the amount of information about individuals held by industry and governments, and the lack of knowledge and ability of the same individuals to control the use of that information*”(Hoffman, 2014b).

In this context, it is clear that it is very hard for people to have **an effective relationship with their own data** – they lack *agency*, *negotiability*, and *legibility* (Mortier *et al.*, 2014). Put simply, once the inevitable sacrifice of personal data has occurred, there is a ‘*point of severance*’ (Luger and Rodden, 2013) beyond which individuals are cut out of the loop. Yet nonetheless, **people need to understand** the changing ways in which they are seen through data, or **face risks of unfair treatment, or physical or psychological harm** (Bowyer *et al.*, 2018). More than this, people need to be able to exert influence over their data, to express or adjust their consent dynamically as things change. Left unexplored and unchallenged, the situation will not improve, as the datafication of society grows, and user agency continues to diminish.

This power imbalance and lack of agency & negotiability is the problem space which this thesis delves into. Clearly, this is a critical problem for society, and in order to understand and validate these problems, and design possible solutions, there is a need for research into how people relate to data in today’s world, what capabilities they need from personal data, and how they would like service providers to handle their data. While many, especially in the field of Human-Computer Interaction (HCI), have researched the specific mechanisms by which humans interact with data, few have looked at the sociotechnical nature of this problem of humans existing in a world where they have limited access to or agency over their personal data.

I will now explain my personal motivation for conducting research in this problem space before outlining the objectives and structure of the research.

1.1.1 Personal motivation and context

This PhD and this thesis represent the culmination of my lifelong passion to help people get more value from our computers. Over 30 years ago, I learned from an early age about computers by programming my Acorn Electron, one of the many 1980s home computers that taught their users that the computer was a tool to be exploited, one that you could master and bend to your will. In my formative years at University and beyond, I lived through the birth of the public Internet and marvelled at the ability for computers to connect people across the world, empower individuals as creators, innovators and broadcasters, level the playing field and transform the way people interact. Keenly tracking and embracing the Web 2.0 revolution while observing the digitisation and disruption of so many industries, I became fascinated with the ways in which humans were shaping computer systems which in turn were shaping our habits and our society. As a graduate software engineer at IBM in the 2000s, I podcasted about new ways to be more productive with computers, and participated in an innovation club with colleagues imagining new ways to relate to digital information, and I gradually moved from back-end development to front-end development to user experience, getting closer to a place where I could help end users benefit from technology. From 2009-2011, while working in Canadian startups, I founded and

was a lead writer on a blog called Human 2.0¹, examining the inter-relationship between society and emerging technology. I was witness to a changing world, where we were gaining new capabilities, but also, through the digitalisation of businesses and the shift to data-centric cloud-centric business models, losing our agency to harness computers for our own ends. I presented short talks on my developing ideas about these changes and what better human data interaction might look like four times at Bitnorth conferences² and had essays published at O'Reilly Radar(Bowyer, 2011) and in print (Bowyer, 2012). Despite seeing further potential for smarter, more helpful computer systems through my participation in the Semantic Web community and being a senior developer of semantic text analysis software at Open Text, by 2014 it was beyond doubt to me that the software industry had lost its way, prioritising business goals over user agency, reducing features and creating technology designed to limit and corral users to behave in certain ways. Web 2.0's revolutionary potential of a 'people's internet' was squashed and withered away in the face of new data giants Google, Facebook, Apple and Amazon and their reshaping and usurping of Internet and smartphone technologies. Against a backdrop of a social media revolution which was literally breaking society and democracy to further the pursuit of profit (Tufekci, 2017; Hall, Tinati and Jennings, 2018), I took the leap to escape corporate, for-profit IT in order to seek ways to research, design and help to build a better digital future, with the objective of making computers useful again. This led me, via a web science architect position at citizen science platform Zooniverse that gave further understanding of user motivation and of the power of collaboration around data, to join the Digital Civics CDT programme (Open Lab, 2014), where I was finally able to work full-time on this most important of problems – understanding and improving Human Data Relations. It has been a tremendous privilege to spend six years understanding in great detail the nature of the problems facing our data-centric society, to map those impacts into tangible needs, and to be able to map out the landscape for improving the way we relate to data. As well as allowing me to discover grounded evidence to quantify and qualify the losses of agency I had observed and theorised, this programme has given me space to experiment with using both GDPR and web-scraping to access data and push boundaries, and to design and prototype new models and views of data and of information which have transformed the way I look at digital information and how we relate to it, and which I hope can help others in the same way. Looking forward, this opportunity has opened doors that have enabled me to begin to put these learnings into action, working on important projects [see 7.1.1] with Connected Health Cities, BBC R&D, and Hestia.ai to explore how data interaction reforms can be realised in practice, and how we can come not just innovators but social data activists to begin to have an impact and to build that better future. It is the journey of a lifetime, and also one that is in many ways just beginning. I hope that my work and this thesis can, in some small way, contribute to a better, more human-centric digital

¹Archived at <https://web.archive.org/web/20111231165329/http://www.human20.com/>

²<http://bitnorth.com/shortbits/>

world, and I can't wait to see where this leads.

1.1.2 Research Objectives and purpose

Given the societal importance of this problem as outlined at the start, the goal of this research and of this thesis is **to produce knowledge and insights** that can enable researchers, innovators and activists to make progress in **redressing the power balance between individuals and data holders** and in delivering **increased agency and negotiability** to individuals. Informed by a constructivist ontology and a pragmatist epistemology (further detailed in Chapter 3), and employing a multi-disciplinary *Digital Civics* (Vlachokyriakos *et al.*, 2016) approach, this objective will be approached by conducting participatory research in relevant contexts, to understand and synthesise a clear model of how people relate to data, how they understand and use it, and what they need from it in order to thrive and to meet their own goals.

To do this, two key settings in which individuals have some remit to '*look behind the curtain*' of the previously opaque data-centric organisations they interact with have been identified, creating spaces where individuals can be interviewed and probed to uncover their attitudes and desires:

1. Supported families, who meet with '*Early Help*' support workers whose role is to access civic data to understand and empower those individuals to improve their lives - forming the context for Case Study One of this thesis [Chapter 4]; and
2. Ordinary British and European citizens, who gained new legal rights via 2018's *GDPR* legislation, enabling them to request copies of held personal data along with other relevant information from the companies and service providers in their lives - this forms the context for Case Study Two of this thesis [Chapter 5].

These Case Studies were designed to enable the collecting of interview transcripts and other data as primary data sources that can be examined in order to attempt to answer two key research questions (RQ):

- **RQ1. What is the human experience of personal data, and what do people want from their data?**
- **RQ2. What role does data play in people's service relationships, and how could relationships involving data be improved?**

Having gathered data and insights from these settings, the further objective is to look for commonalities across the two settings that can serve to validate each other and be distilled into clear insights about people's relationships to personal data that can serve as answers to the two RQs. This synthesis and analysis of interview data will form the core empirical element of this PhD research.

This research is situated in the Human-Computer Interaction (HCI) discipline, which means that **design** (both participatory co-design and expert-informed user-centred design 3.5) forms a key part of the approach to exploring the

problem space. However, despite initial consideration, it was explicitly decided that **field evaluation of particular interface designs or processes would not be done as part of this PhD**. Given the broad and far-reaching nature of the problem space described, such work would be technically difficult and, highly resource-intensive and time-consuming. The formation of clear underlying models around data attitudes and needs is viewed as a more critical first step that should and must precede any implementation of systems or processes aiming to improve the way in which people can acquire improved agency and negotiability '*in the wild*'. Nonetheless, given this research's purpose of enabling others to have meaningful impact on the world, a significant piece of this thesis (Chapter 7) will be dedicated to **sharing the designerly thoughts, models and ideas** I formulate both from talking to participants of the Case Studies, and from parallel research and development activities I will conduct alongside the primary empirical research of this PhD. Therefore, while this thesis will not be able to be definitive about which precise interfaces, processes or practices will succeed in combatting the power imbalance over data, it will not shy away from offering untested theories, ideas and insights, originating from both **participant-expressed needs** and **my own research-and-practice-informed expertise**, which are presented so that future researchers and innovators might explore or evaluate them and build upon the contributions of this thesis.

This dual approach of focusing the PhD on empirical grounded academic research, while also operating outside of the PhD as a researcher and developer working on real-world innovation in the same space, will produce a **powerful feedback loop** where practical realities of tackling the problem space can inform the participatory research explorations of the space, and vice-versa, feeding into the **action research cycle** [3.2] of the PhD through more than just the participatory empirical research itself.

I will now explain the primary contributions of this thesis [1.2], and the publications resulting from the research [1.3], before explaining the structure of the thesis across the subsequent chapters [1.4].

1.2 Nature and Contributions of the thesis

This section lists the contributions (**Cn**) of this thesis: specifically:

- an understanding of what people need when they relate to data [1.2.1];
- the establishment of the field of *Human Data Relations* [1.2.2]; and
- additional contributions specific to the Case Study contexts of
 - Early Help [1.2.3], and
 - GDPR/everyday data access [1.2.4].

1.2.1 An Understanding of what People want from Personal Data

C1: An understanding of what People want in Direct Data Relations

Through the concluding sections of Chapters 4 and 5, the reader will be able

to see that research participants across both studies (and the pilot study) shared common issues around personal data. In section 6.1 of Chapter 6, those commonalities that address RQ1 *what people need in direct data relations* are specifically expressed in answer to that question as three specific needs:

- for data to be **visible** [6.1.1],
- **understandable** [6.1.2], and
- **useable** [6.1.3].

C2: An Understanding of what People want in Indirect Data Relations

Similarly, in section 6.2 of Chapter 6, those commonalities that address RQ2 – *what people need when they have an indirect relationship to their data because it is held by someone else* (such as their service provider) – are specifically expressed to answer RQ2 as three specific needs:

- **process transparency** [6.2.1],
- **individual oversight** [6.2.2], and
- **involvement** in processes and decision-making [6.2.3].

1.2.2 Establishing a new field: Human Data Relations

C3: The synthesis and formulation of the field of Human Data Relations (HDR).

At the highest level, the contribution of this thesis is to establish and map out a new field of research and innovation, **Human Data Relations** (HDR). This begins with a broad literature review in Chapter 2 of prior areas of research and established thinking that contribute to this field, specifically the problems of data-centricism and limited access to data [2.1], a review of prior work in personal information management and interaction [2.2], and of existing research and innovation around human-centric perspectives on data [2.3]. The HDR field is then explored and understood through the two research questions RQ1 [3.3.1] and RQ2 [3.3.2]. Both RQs are explored through participatory research and qualitative data analysis across the two contexts of Chapter 4 and Chapter 5, contributing to a synthesis in Chapter 6 of what people want in direct data interaction [RQ1, 6.1] and in relationships that involve the use of personal data by the other party [RQ2, 6.2]. Finally in Chapter 7, the field of HDR is refined [7.2], and a landscape of possible approaches to improve HDR is mapped out, including the identification of specific obstacles to progress [7.3] and possible approaches that could be explored [7.4].

C4: A clear delineation of two primary motivators for individuals seeking better HDR

In 7.2.3, informed by both participatory research within this thesis and by the research and design activities conducted in external research settings 7.1.1, I outline a first top-level perspective on the HDR space, that there are two key reasons why people need good data relations:

- *Life Information Utilisation*, and
- *Personal Data Ecosystem Control*.

C5: A map of the HDR landscape, identifying obstacles, insights and opportunities

The goal of this thesis is to set the stage for future research and innovation in the newly-defined space of Human Data Relations. While evaluating methods and approaches ‘in the wild’ was well-beyond the scope of this thesis, my involvement in external research settings allowed a broad and grounded understanding of the HDR landscape and its practicalities to be formed, such that the landscape can be mapped from multiple perspectives.

In 7.3 the specific wants mentioned above in C2 and C3 are reduced to four simple objectives for effective HDR:

- data awareness and understanding
- data useability
- ecosystem awareness and understanding
- ecosystem negotiability

The same section then continues to map out eight obstacles to better HDR that exist in these four areas, as well as four obstacles that exist in the solution space across all four:

1. The Personal Data Diaspora
2. Illegible Data
3. Data that isn’t free
4. Unmalleable and non-interrogable data
5. Hegemony through data holding
6. A trend of actively diminishing of users’ agency
7. Closed, insular & introspective practices
8. The inaccessible data self
9. A lack of HDR demand from individuals
10. A lack of HDR demand from organisations
11. A lack of interoperability
12. Insufficient machine understanding of human data.

To begin to address these obstacles, seven insights are offered that could seed future research and innovation towards tackling these obstacles:

1. Life information makes data relatable.
2. Ecosystem information is an antidote to digital life complexity.
3. Life & ecosystem information should be useable as a material.
4. Data needs provenance.
5. Data holders exploit four levers of power to manipulate the digital landscape.
6. Semantic analysis and information standards can transform data storage and facilitate human-centric interface building.
7. New life capabilities and pain relievers drive user demand.

8. Better HDR can deliver business value through increased accuracy and consent, and decreased liability.

1.2.3 Additional contributions in the Early Help and Civic Data Use context

C6: Validation and enumeration of supported families' attitudes and needs around civic data

The early research carried out of my pilot study [1.3.1], continued in Case Study One [Chapter 4], served an important purpose to validate that people do feel the effects of the data records about them, and do want access. Prior to this research, colleagues speculated that people would not really care about their data, but in fact these studies found evidence that people want **continuing rights, control and visibility over their personal data**, so that it remains **fair, accurate and meaningful**. Furthermore the lived experiences of supported families show how **data can become a proxy for human involvement**, and that this can be harmful and disempowering. In particular, my research shows that:

- Supported families need **meaningful interaction** with and through data,
- They need to be **given a voice** to explain, challenge or add context to data, and
- **Transparency over data can improve trust** in support services.

C7: Shared Data Interaction: A proposed model for more efficient and empowering social support relationships that embraces human-centricity.

Given the conflicting goals of support service providers wanting to be more data-centric to improve accuracy [4.1.2, 4.2.3] while supported families want a more human, less data-centric treatment, I created a model that has the potential to address both parties' needs and enhance the support relationship: *Shared Data Interaction* [4.2.4]. While this was not evaluated in the field, it is consistent with emergent practices 4.3.1, and when explored thoroughly by both parties in Case Study One - especially in phase 2 [Table 3.1] - was perceived to be beneficial. The benefits (and challenges) of such an approach are explored thoroughly in [4.4.2]. At the level of multi-party collaboration, further evidence for the effectiveness of bringing people together around representations of data is found at the meta level of this research - specifically in the success of the methodologies [3.5.2] I used in both Early Help studies, echoing other researchers' work on *boundary objects* (Star, 1989) and '*things to think with*' (Brandt and Messeter, 2004).

1.2.4 Additional contributions in the context of GDPR and Everyday Data Access

C8: A model to understand the five different origins of held personal data

Through the analysis of privacy policies and GDPR legislation conducting during the design and interview phases of Case Study Two, I produced a model of the five different types of data organisations can hold about individuals [Table 5.X]:

- **Volunteered Data**
- **Observed Data**
- **Derived Data**
- **Acquired Data**
- **Metadata**

This model has been used as both during design and ideation sessions at BBC R&D as well as being used and cited within Sitra/Hestia.ai's digipower study, both as a model for explaining data holding to participants and as a frame for data analysis (Bowyer, Pidoux, *et al.*, 2022; Pidoux *et al.*, 2022).

C9: A rich understanding of the lived experience of accessing data using GDPR rights and of motivations for GDPR data access

Case Study Two fills a critical research gap in understanding the human experience of using GDPR to access one's personal data: confirming past findings that compliance is poor and returned data often incomplete (Ausloos and Dewitte, 2018), but going much deeper [5.4] in uncovering specific attitudes such as **resignation about data sacrifice, disappointment in GDPR handling** by service providers, and **a lack of answers to questions**. Specific motivations for GDPR data access (and hence more widely for HDR) are enumerated, which provides a valuable starting set of requirements for future research and innovation (see [Table 5.X] and the supplemental materials of (Bowyer, Holt, *et al.*, 2022)).

C10: Evidence for the impact of knowledge about data handling practices on provider trust and perceived individual power

A particularly novel and surprising discovery from Case Study Two was that the use of GDPR rights and privacy policy analyses to scrutinise data-holding service providers often resulted in a **decrease in trust** in those same data holders. At the same time, GDPR use on the whole failed to provide a net increase in perceived individual power; it was **not empowering** people and hence not meeting its own goals. Further analysis of these patterns also showed that **data handling practices are critical to trust and consumer loyalty** [5.3.4; [5.5.2]].

C11: Guidance for policymakers, data holders and individuals on how to improve HDR

Synthesis and analysis of participant experiences in Case Study Two enabled the production of specific guidance [5.5] for parties involved in data relationships:

1. Policymakers and DPOs should do better at **enforcing GDPR rights**, and regulate to **improve response quality** and legislate to mandate data holders to **support data subjects in understanding data**.

2. Data-holding service providers should **improve transparency** over data and data handling process, and could seize the opportunities of **more inclusive and collaborative models of individual data access** to improve trust, empower users and reduce their own liability.
3. Individuals should recognise the critical role of held personal data in modern life, embrace opportunities to **access and exploit their own data** and **use data access rights to hold service providers to account**.

C12: A proto-methodology for educating individuals about held data, data access and the data ecosystem

While it was not designed as a methodological contribution nor formally evaluated as such within the scope of this thesis, the guided-data-retrieval-and-interview approach of Case Study Two [5.2] has proven to be **highly valuable and replicable** as means to connect people with their held data and conduct research at that intersection point. Indeed my creation of this methodology was the primary reason why I was approached and employed as lead researcher of Hestia.ai/Sitra's digipower investigation [ADD REF], which **adopted Case Study Two's methodology**, with some adaptation and broadening of scope, for an important EU study auditing and understanding the power of data holders in the data economy [Bowyer, Pidoux, *et al.* (2022); Pidoux *et al.* (2022); ADD REF TO FINAL REPORT].

1.3 Publications arising from and connected to this research

1.3.1 Pilot Study

My Doctoral Training programme at Open Lab began with a Masters in Research in Digital Civics. For my MRes project³, I conducted a pilot study, interviewing and exploring issues around data with families who had experience of social care services. During the first months of this PhD I conducted new analysis of the data collected, resulting in the synthesis into a full first-author paper published at CHI 2018:

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

This study is given a special status in this thesis; while it is not officially to be examined, it plays a critical role as a pilot study for Case Study One and its findings and insights are built upon in Chapters 4, 6 and 7. As such, the paper is included in full in Appendix 1.[ADD APPENDIX REFERENCE]

1.3.2 Primary Case studies

1.3.2.1 Publications from Case Study One The work exploring shared data interaction in Early Help carried out in Case Study One has been initially

³MRes result awarded: Distinction.

published as an Extended Abstract at CHI 2019:

- “Human-data interaction in the context of care: Co-designing family civic data interfaces and practices” (Bowyer *et al.*, 2019)

This work was also presented at the conference in the form of a poster, which is shown in Figure X.

A 15,000 word+ detailed first-author journal paper has been drafted to supplement the extended abstract and will be submitted for publication in due course.

1.3.2.2 Publication from Case Study Two The work exploring the human experience of GDPR data access carried out in Case Study Two has been published as a full first-author paper at CHI 2022:

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

I carried out all field research myself. Data analysis and paper writing was shared between myself and Jack Holt.

1.3.3 Workshop papers & presentations

During the PhD, I gave a number of presentations and published three workshop papers which included material from, or directly contributing to, this thesis and helped shape the ways in which I express the arguments within:

- “Designing For Human Autonomy: The next challenge that civic HCI must address” - a short talk I presented to my peers in January 2017 laying out the landscape of reduced agency and possible avenues for improving humans’ relationships to their data.
- “Free Data Interfaces: Taking Human-Data Interaction to the Next Level” - a CHI 2018 workshop paper formalising past pre-PhD design thinking and outlining a vision for unconstrained and useful data interaction interfaces
- “A Grand Vision for Post-Capitalist HCI: Digital Life Assistants” - a CHI 2018 workshop paper where I imagined a form of digital computer assistant that is far more helpful and human-data-centric than the digital voice assistants of today.
- “Personal Data Use: A Human-centric Perspective” - in early 2020 just prior to the pandemic, I was invited to give lectures on my research to undergraduate students at both Northumbria University and Newcastle University.
- “My Thesis in 3 Minutes: Understanding and Designing Human Data Relations” - in April 2021, I presented my thesis in the 3 minute thesis competition, and was co-winner of the people’s choice prize.
- “Human-Data Interaction has two purposes: Personal Data Control and Life Information Exploration” - A workshop paper I presented at CHI

Human-Data Interaction in the Context of Care:

Co-designing Family Civic Data Interfaces and Practices

Alex Bowyer¹, Rob Wilson², Stuart Wheater³, Matthew Snape¹, Kyle Montague¹
 Open Lab, Newcastle University; University of Northumbria; Arjuna Technologies Ltd.

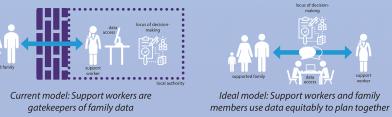
Context

Data Use in 'Early Help'

- Support workers routinely view client families' data.
=> "We need to check the family's details."
- Families have **no access** to the data stored about them.
=> "We want to see what they have on us."
- What if families & support workers viewed data together?
=> "Let's explore and use this data."

Research Question

Could shared HDI shift the locus of decision-making towards the family?



Methods

3 participatory workshops:

- families & children
- support workers
- parents with support workers

Methods used in the workshops included:

- Sentence Ranking
- Card Sorting
- Ideation Decks
- Poster Creation
- Storyboarding Cards

Findings

Identified Interaction Practices

- make no assumptions
- let people question, contribute, initiate
- data openness and transparency
- clear, **meaningful**, visual summaries
- involve people in data-centric work
- safe and respectful data handling
- processes must always support **change**
- personal data interfaces
- strive to understand the family **completely**
- building (and keeping) TRUST** is key

Discussion

Design & Practice Implications

- There is both **a need** and **a desire** for Human-Data Interaction.
- Current practices disadvantage families; HDI could **empower** them.
- Data involvement and openness would **build empathy & trust**.
- Shared data interaction** could help the support relationship.
- Involving the family can help with issues of **consent** and **accuracy**.

Open Lab
at Newcastle University

Paper: <http://bit.ly/hdi-in-care-paper> Poster: <http://bit.ly/hdi-in-care-poster> Email: a.bowyer2@newcastle.ac.uk Twitter: [@alex_bfree](https://twitter.com/alex_bfree)

EPSRC
Engineering and Physical Sciences Research Council
CDT in Digital Civics (EP/D006760/1)

Figure 1: Figure X: Poster Presentation of Case Study One

2021, where I first outlined my model of the two motivating factors for interacting with personal data.

1.3.4 Publications from other work

During the same timeframe as this PhD, I have also contributed to a number of publications tangential to my primary research agenda:

- As a researcher and developer on the Connected Health Cities SILVER project [ADD REF TO 3.x], I contributed to work published through Newcastle University's internal report to CHC (not publicly available) and the overall impact report (p129-130), and more directly published demonstration videos of a health data interface prototype developed by myself and Stuart Wheater.
- As a researcher and developer on DERC's Healthy Eating project, I developed interface prototypes (no longer online) and was co-author to two research publications at BCS 2021 (Goffe *et al.*, 2021) and in *Interacting with Computers* (Goffe *et al.*, 2022).
- As a research intern on BBC R&D's Cornmarket project [ADD REF], I published an internal research report [ADD REF] into personal data store design, as well as a 'stimulus presentation' to launch an internal hack week and a BBC blog article about the work (which was not officially published) [ADD REF].
- As project leader, data access coach and researcher at Hestia.ai, I was a lead author on a research report auditing the data economy (Bowyer, Pidoux, *et al.*, 2022), and co-author on a research report on power mechanisms in the data economy (Pidoux *et al.*, 2022).

1.4 The structure of this thesis

The overall structure of this thesis is illustrated in Figure X. This introduction is followed by a literature review [Chapter 2] and a methodology chapter [Chapter 3]. Both research questions RQ1 and RQ2 are examined in both Case Studies, and these studies are documented as self-contained pieces of research in [Chapter 4] and [Chapter 5] respectively. In [Chapter 6] the findings and insights from the Case Studies are synthesised into common findings as to what people want from data and from data holders, which concludes the investigation of the two research questions. [Chapter 7] opens up the future of the HDR landscape and sets the stage for future research and innovation, building on both the research conclusions in Chapter 6 as well as my practical experiences from other related research and development activities conducted outside of this PhD research but during the same timeframe.

Chapter 2 is a literature review divided into three key sections. The first [2.1] examines the difference between data and information, outlines the central role data has taken in our society, why people need effective access to their data and how laws have been introduced to try and deliver this. The second [2.2] serves

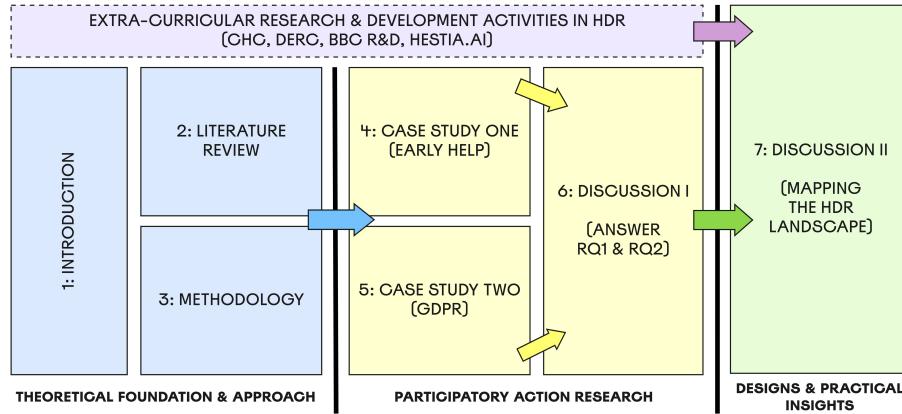


Figure 2: Figure X: The Structure of this Thesis

as history of personal data interaction, from Personal Information Management to the emergence of complex digital lives involving relationships with many data-holding providers. Finally [2.3] charts a path from HCI and Human-Data Interaction foundations through to the embracing of sociotechnical thinking around data and the current bleeding edge of human-centred innovation.

Chapter 3 describes the methodology used in this research, explaining first the constructivist ontology and pragmatist epistemology behind the approach [3.1]. Then the choice of participatory action research and co-design from a Digital Civics standpoint is explained [3.2]. The two research questions (RQs) are explained in more detail [3.3] and the contexts for the Case Studies are introduced from a personal ‘what did I do?’ perspective [3.4]. Finally the specific methods and techniques adopted in the research are explained and illustrated, including sensitisation, workshop activities, recruitment strategies and modelling [3.5].

Chapter 4 is a detailed, self-contained account of Case Study One. This begins [4.1] with a detailed introduction to the UK’s Early Help social care context, including its history of data-centrism and the inherent contradiction between that and the empowerment goals of Early Help which make it an ideal setting to explore my research questions. In 4.2, prior findings on family and staff perspectives are introduced, motivating the ‘*Shared Data Interaction*’ vision and structure of the workshops, whose participants’ shared values are introduced. The thematic findings are detailed extensively [4.3] including participant quotes, and then contextualised in the discussion [4.4] in terms of the value of involving people with their data, the link between HDI and effective data access, and the implications of shifting the locus of decision-making. The research is summarised in [4.5].

Chapter 5 is a thorough and self-contained write-up of Case Study Two. The

first section [5.1] provides additional context on the need for data access, prior GDPR research and the human-centric approach to this study, whose design and configuration is detailed in 5.2. Findings are shared: firstly [5.3] the largely quantitative outcomes of the participants' GDPR requests, interview responses and trust/power scores, then the thematic findings 5.4, illustrated with participant quotes. The discussion [5.5] contextualises and synthesises the findings into human-centric, GDPR-improving guidelines for policymakers, data holders and individuals. The research is summarised in [5.6].

Chapter 6 synthesises the two case studies, and answers RQ and RQ2, bringing the thesis's central empirical research to a close based on findings and discursive insights from both studies backed by literature references. The answer to RQ1 is provided [6.1] as people's three wants from direct data relations (for it to be visible, understandable and useable), and the answer to RQ2 is provided [6.2] as people's three wants from indirect data relations (process transparency, individual oversight, and involvement). The chapter concludes [6.3] by outlining this thesis's perspective power and positioning the pursuit of these six '*data wants*' as *empowerment* relative to that perspective and thus as '*better HDR*'.

Chapter 7 is a deliberately broad, shallow and open-ended chapter framed around synthesised research and designerly insights. This moves beyond a traditional thesis structure, with the goal of providing real-world experience and ideas on *how such human-centric empowerment might be achieved*, which I believe is the best way to share the HDR insights I have gained along this six year journey that go beyond traditional empirical findings. This is also augments this thesis to become a *valuable and actionable reference material* for future researchers and innovators. First [7.1], the additional extra-curricular industrial and academic research contexts that I worked in alongside this PhD are introduced, along with relevant additional background around *Theories of Change* (ToC) and consideration of those who pursue better HDR as a *recursive public*. In [7.2], insights from the thesis's findings and from external experiences are synthesised to frame the pursuit of this thesis's '*data wants*' as a defined field of future research, called *Human Data Relations* (HDR). The landscape of HDR is mapped out in two parts. First [7.3], the six data wants are mapped into four '*landscape objectives*', which are then used as a frame to outline the identified *obstacles* to pursuit of those objectives, followed by useful *insights* that could be embraced to tackle those obstacles. Then [7.4], using a ToC framing, my identified *opportunities for progress* are introduced, divided into *four different trajectories of change* that could be executed in practice to pursue better HDR. Finally [7.5], the thesis concludes with a summary of its findings and contributions.

2 Literature Review

"We are the sources of surveillance capitalism's crucial surplus: the objects of a technologically advanced and increasingly inescapable raw-

material-extraction operation.” - Shoshana Zuboff (Harvard professor, social psychologist and author)

2.1 Understanding Data and its Role in Today’s World

2.1.1 What is Data?

Data is an oft-used word that carries multiple meanings. In everyday speech, it might refer to mobile phone bandwidth, a filled application form or a collection of 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

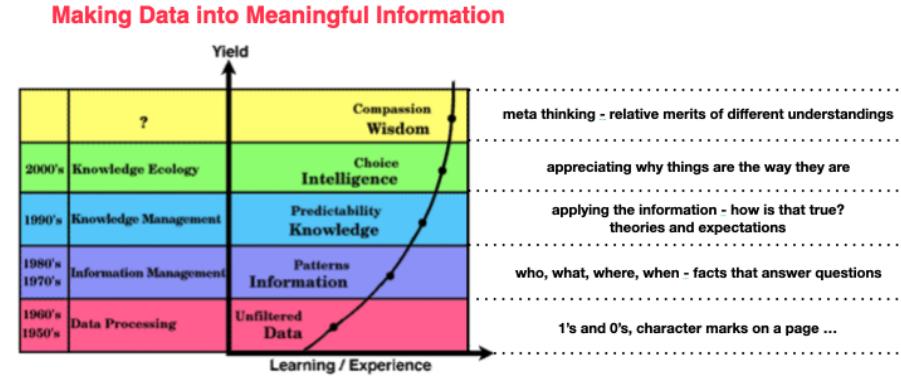


Figure 3: REDRAW Figure 1: Making Data into Meaningful Information

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 Personal Data & the Rise of Data-centrism

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

The personal computer revolution ('The personal computer revolution', no date) of the late 1970s and 1980s put computers in every office and eventually every home too, 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.

In 2017, the average American Internet user had 150 online user accounts with different providers (Caruthers, 2018). Data for the UK shows the number of service and supply relationships each individual has to manage increasing from around 45 in 1997 to around 250 in 2020 (Henderson and Group, 2020). As the amount of personal data relating to each of us has increased, the need for individuals to be able to manage this has grown, but 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 Legislating to Protect Personal Data: the Story so far

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

access and challenge personal data is [...] the most important privacy protection safeguard" 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 (ICO)) with which organisations were required to register their usage of personal data), Ireland's *Data Protection Act (1988)* (which introduced the concept of a 'duty of care' for data collectors - that they are expected to avoid causing damage or distress to data subjects), the EU's *Data Protection Directive (1995)* and the UK's *Data Protection Act (1998)*. However, such laws were generally found to be ineffective - in 2002 Simon Davies, director of Privacy International, said that the UK's DPA was "*almost useless in limiting the growth of surveillance*" (Millar, 2002), and research commissioned by the ICO in 2008 stated that the European Data Protection Directive was "out of date, bureaucratic and excessively prescriptive" (McCullagh, 2009).

In 2018, the EU's *General Data Protection Regulation (GDPR)* came into force, carrying with it significant designed-to-hurt fines for non-compliance (Kelly, 2020; Leprince-Ringuet, 2021). This legislation has finally given individuals the practical means to exercise their data rights ('The GDPR: Does it Benefit Consumers in Any Practical Way?', 2020) against a backdrop of massive personal data use across society that had previously rendered data access requests impractical (Cormack, 2016). The GDPR – which gives individuals key rights including rights to timely data access, explanation, erasure and correction (Information Commissioner's Office, 2018) – can be seen as the first serious attempt to rebalance 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 *Act on the Protection of Personal Information (2017)*, India's *Personal Data Protection Bill (2019)* and the *2020 California Consumer Protection Act (2020)* (CCPA). In the USA, there has been no national privacy law yet, but the GDPR's influence is being felt in court rulings (Hoofnagle, Sloot and Borgesius, 2019).

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

Agency for Fundamental Rights, 2020) and at the time of writing in 2022, 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 use in this thesis, specifically (Information Commissioner's Office, 2014; The European Parliament and the Council of the European Union, 2016a):

- *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 Personal Data Access

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

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

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 Personal 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 A History of Personal Data Interaction

2.2.1 Computers as General-Purpose Information Tools

In the immediate aftermath of the second World War, Dr. Vannevar Bush wrote a landmark article for *The Atlantic Monthly* in which he envisioned a new scientific agenda for America and the world - 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 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 principled that storage should be transparent, archiving and compatibility should be automatic, and concise overviews of groups of related information should be available (Freeman and Gelernter, 1996). Central to this system is the idea that personal data can most easily be navigated when viewed as a *timeline*, 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, 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 its users.

2.2.3 Self Informatics

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 monitoring our own behaviours, and being too busy to introspect. He proposes that "*Computers can help: They can store large amounts of data, analyse the data for patterns, visualise the data, and provide feedback at opportune times*" (Li, 2009). Just as QSM has gained traction with enthusiasts in the general public, so personal informatics has grown as an area of research, development and study in academic circles. While QSM and lifelogging focus slightly more on capturing data about oneself and personal informatics focuses slightly more on the mechanisms of integrating and reviewing self-tracking data, there is so much overlap that all three can be considered the same field, 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.

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, p. p72) 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,

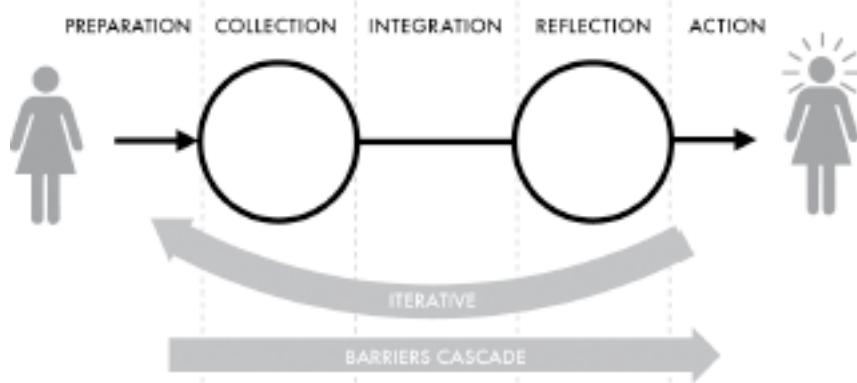


Figure 4: Figure 2: Li *et al.*'s Stage-based Model of Personal Informatics Systems (Li, Forlizzi and Dey, 2010)

- *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, Forlizzi and Dey, 2010) but can improve self-control (O'Donoghue and Rabin, 2001), increase self-awareness (Aslam *et al.*, 2016) and enable positive behaviours such as saving energy (Seligman and Darley, 1976).

Reflection can be facilitated in SI systems by enabling the tracking of subjective factors such as mood or intent as well as objective factors such as health or activity, 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, Forlizzi and Dey, 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 we see that the increased automation around self-tracking and data collection was judged to have given people more free time and energy for reflection and managing their goals (Feng and Agosto, 2019).

2.2.4 Overwhelmingly Complex Digital Lives

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), Rightmove (estate agents), Steam (computer games), Just Eat (takeaways), and many more (Levine, 2011; Carter, 2015). As a result, we now produce rich data trails simply by going about our daily lives, and this has become "*the driving force for value creation*" online (Symons *et al.*, 2017). More recently as we start the 2020s, the trend has accelerated, with the COVID-19 pandemic 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* (Abiteboul, André and Kaplan, 2015), not only because it is held by organisations without giving us effective access, but also by various practical means such as format incompatibilities, device restrictions, paywalls, and a lack of data portability. We need to free our data, as I expand upon (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 as *AllOfMe.com* (‘AllofMe Company Profile’, 2007; ‘AllofMe.com Teaser Clip’, 2008) in 2008 and *myTimeline* a decade later (‘myTimeline’, 2018); however none of these products have reached public availability. To date the closest market-successful tool that people have for general purpose information handling is *Facebook*, given that it can store personal information, handle asynchronous and instant messaging, news, photo sharing, some retail functionality, brand interaction & support, calendaring and event management, and group discussions. However, it is a closed system with no capability for customisation; none of its content is available outside the network and external content cannot be linked or interacted with except by import; as such it cannot be considered a PIM system. Its own *Timeline* feature, promoted at launch in 2011 as “*the story of your life*” and “*a new way to express who you are*” (Siegler, 2011) has been retired, along with many other tools designed to make information easier to manage such as personal news feeds and friend lists (Perez, 2018), 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 at time of writing 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 of the provider, 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 a matter of arranging your own bookshelves, but rather a multi-party negotiation over representation, ownership, access and consent. Data is a shared resource with multiple users, and only a few researchers have begun to look at people's interactions with data in this context (for example, activity streams (Hart-Davidson, Zachry and Spinuzzi, 2012), social sensemaking (Puussaar, Clear and Wright, 2017), and decentralised file storage (Zichichi, Ferretti and D'Angelo, 2020)). There has been negligible research into the role data plays within human relationships.

This is the second research gap that 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 An Alternative to Data-centricism: Human-centricity

2.3.1 Human-Computer Interaction (HCI) Foundations

Up until the 1980s, the only reasons to consider the relationship between a human and the computer they were using were ergonomics, comfort and efficiency. People

were shielded from the complexities of the machines they were using—the machine did the work (usually a single task) and the human was just the *operator*. In the 1990s, the “*first wave*” of what is now known as *Human-Computer Interaction* (HCI) recognised humans as *actors*, often operating in groups, who had tasks to perform either using or assisted by technology (Bannon, 1995). People were now *users* of technology. Design thinking shifted from machine-centric to *user-centric design* (UCD), motivated by the goal of helping the user to perform their tasks better. In the personal computer revolution of the 1990s, people began to work in complex and varied multi-user situations, and observation and *understanding of a user’s working environment* 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; individuals would be free to use computers as they saw fit. This created a “*chaos of multiplicity for HCI in terms of use technologies, use situations, methods and concepts*” (Bødker, 2015); designers would now need to “*embrace people’s whole lives*” (Bødker, 2006). The blueprint for how this could be achieved 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 (HDI)

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

decades for *a new relationship with our stored knowledge* (Bush, 1945; Lansdale, 1988; Rogers, 2006; 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; BBC R&D, 2017; 'HDI Network Plus, University of Glasgow', 2018; 'HDI Lab, Heerlen', 2020), as does this thesis. In order to understand what HDI might mean in practice we can look to Gregory Abowd's 2012 paper which aims to update Weiser's vision. In it, Abowd emphasises the importance of programming for *environments*, building a complete experience for the individual that considers not just the 2D screen they are using, but the entire surroundings and context of their situation. He imagines a hybrid, conjoined experience between people, devices, sensors and the cloud where data storage and processing need not be constrained to the input and output devices we use (Abowd, 2012) and crucially, that the individual within this "*everyday computing*" experience is harnessing technology for their own ends, not just being aided to complete a predetermined task (Abowd and Mynatt, 2000) – in essence they are able to program their own environment.

2.3.3 Human-Centred Design: A Sociotechnical Challenge

Abowd's vision is a helpful reference point to remind us how far from true human-data interaction we are today. As described above, data is trapped, and very few computing interactions today are designed as a situated experience. Some TV streaming services show a good example of an interaction whose design has taken into account context: instead of typing in long email addresses and passwords, difficult on a TV remote, you can visit a short link from a smartphone or PC where you are already authenticated. But even though there are pockets of research around contextual experiences (for example the work around second screening (T. Jones, 2011; Zúñiga, Garcia-Perdomo and McGregor, 2015)), in general most design work today still focuses on a single interaction surface. In order for technology to disappear into the background so that we might live in a calm, engaged manner, as outlined by Weiser and expanded upon by Yvonne Rogers (Rogers, 2006), a more humane interface is needed (Raskin, 2000), one which designs for the whole person. Judging the success of a user interaction can no longer be done by assessing task-completion efficiency (Abowd and Mynatt, 2000) but should consider the holistic needs of the individual at that moment in time.

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*" (Simon, 1971; Croll, 2009; Cogran and Kinsley, 2012; Brynjolfsson and Oh, 2012). Social media technologies like infinite scrolling and smartphone notifications had created "*a culture of perpetual distraction*" (Timely, 2020) which "*hijacks people's minds*" (Harris, 2016). 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 Pursuit of Human-centred Personal Data Ecosystems (My-Data and beyond)

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 sociotechnical designs for personal data interaction. One of the earliest was Doc Searls, who launched a project called *ProjectVRM* with colleagues at Harvard University around 2008. He envisioned a model he called

Vendor Relationship Management (VRM) 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.

Drawing together the work of Searls and various other innovators in the VRM space, David Siegel catalogued an emerging vision of a personal data interface that could realise VRM thinking, which he called a *Personal Data Locker* (Siegel, 2009, 2010). (Equivalent terms *Personal Data Store*, *Personal Data Vault (PDV)* and *Personal Data Services* are also common). He described '*Pull-centric Computing*', where information is 'pulled' at your request rather than being pushed upon you. The WEF's *Rethinking Personal Data* project (mentioned earlier) describes the potential for individuals to have their own *personal data ecosystem (PDE)* of "*commercial entities, acting as trusted intermediaries, exchanging assets on behalf of the individual, following a clear set of principles and legally binding contracts*". They describe the PDV as 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). In my view, this 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* (PIMS⁴) 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

⁴The usage of the abbreviation PIMS here is not to be confused with its earlier use to refer to '*Personal Information Management Systems*' in traditional PIM terminology.

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* (CitizenMe, 2021) is another company with a similar positioning. Other operators such as *UBDI* (‘Whose data is it anyway?’, 2019) and *datacy* (‘datacy - About Us’, no date) are positioned under a different business model which aims to help individuals take control of their personal data for profit. *Open Humans* has a 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 personal data from data-holding organisations (Jelly, 2021). Many of these operators are building for a human-centric world that does not yet fully exist, and Geneva-based firm *Hestia.AI / HestiaLabs* also operates in this space, with a different focus: helping people to understand the existing data ecosystems that make use of their data (Dehaye, 2018), to build towards MyData ideals from the status quo.

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 identifying 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 synthesise an understanding of 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, Forlizzi and Dey, 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, building on the theoretical foundations of effective data access, information management and human-centric data interaction, this thesis progresses PDE / MyData thinking, using methods detailed in the next chapter, in pursuit of my primary research question, which is:

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

3 Methodology

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

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

3.1 Forming a Research Paradigm: Ontology & Epistemology

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

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

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

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

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

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

[TODO shrink this section]

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

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

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

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

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

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

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

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

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

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

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

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

[TODO diminish or place number 3 outside of PhD]

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

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

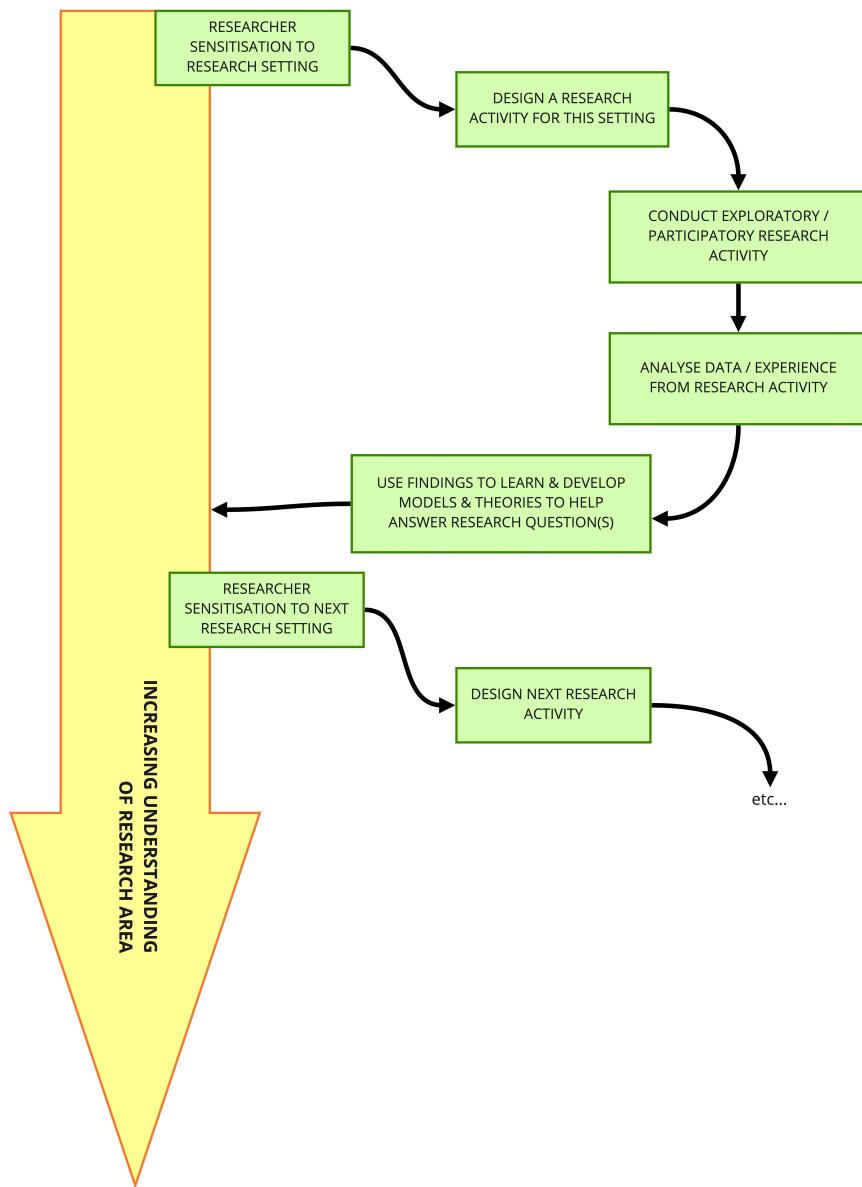


Figure 5: Figure 3: My action research approach

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

3.3 Research Objectives

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

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

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

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-centricity? If your data is used in ways you do not understand or consent to, how does this affect your outlook on the world? This is the second strand of research I will be exploring: to gain an understanding of the data world beyond the individual, so that we can design not just better individual relationships to one's data, but improve people's relationships with organisations that hold and use data. (Note: for the purposes of this study, we only pay attention to service relationships, not social or interpersonal relationships). In this thesis and its title I use the term "human data relations" to encompass both of these aspects - human-data relations (the individual's relationship to their data, as imagined by HDI), but also human data relations, i.e. human relationships that involve data.

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

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

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

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

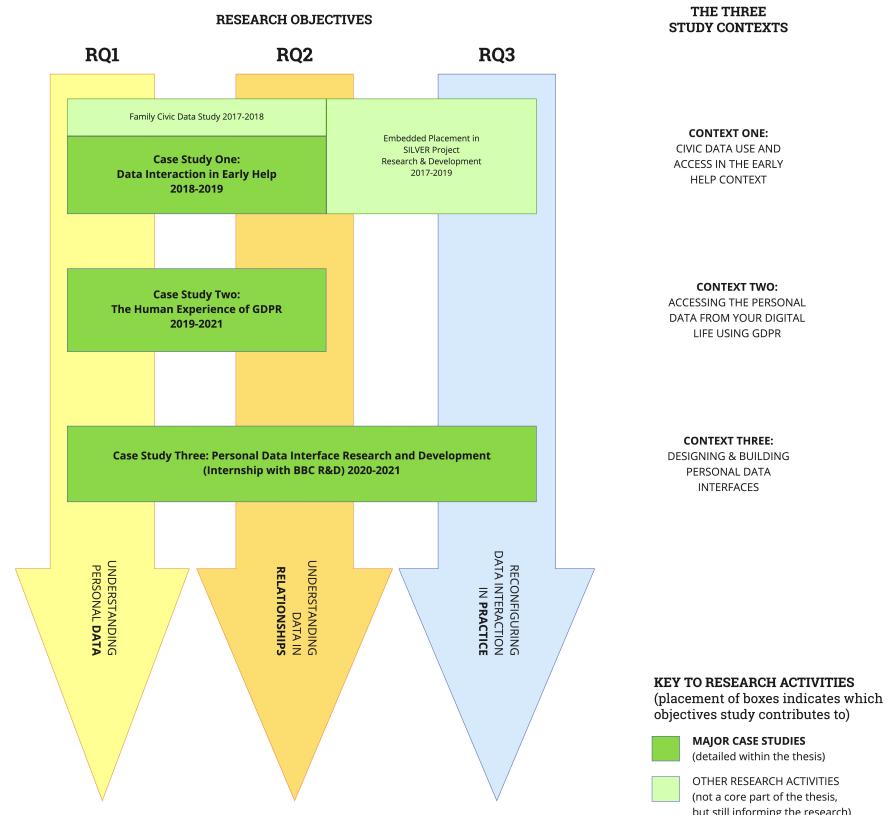
[TODO remove/deformalise this - instead talk about how this will be explored in the discussion, looking at external embedded research and the work of others]

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

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

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

3.4 Overview of Research Contexts and Activities



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

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

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

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

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

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

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) [TODO deformalise RQ3] but also crucially it enabled me to become a participant in my own research, enabling a deeper understanding of this research context.

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

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 develop a detailed understanding of what actual stored personal data looks like (which informs RQ1), they have given me an awareness of the evolving response to GDPR from data controlling organisations (which informs RQ2), and has allowed me to test the limits of what is and is not possible with GDPR (which informs RQ3) [TODO deformalise RQ3]. A summary of observations and findings are presented in [INSERT APPENDIX REFERENCE HERE - OR DON'T!].

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: The Practical Pursuit of Better Human Data Relations

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

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

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

cycle of this thesis (3.2).

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

3.4.3.1 Health Data Interface Development: CHC's SILVER project

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

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

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

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

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

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

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

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

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

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

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

3.5.1 Sensitisation of Researcher and Participants

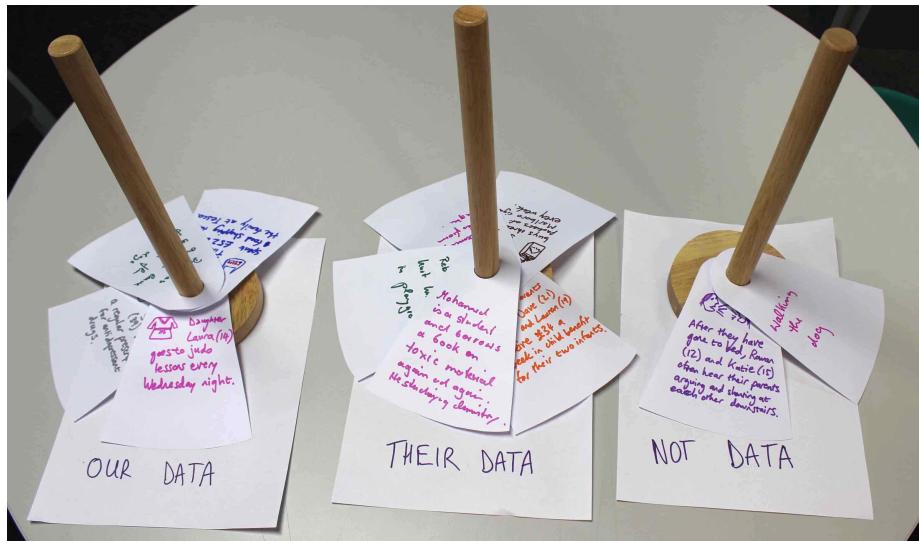


Figure 6: Figure 5: “Family Facts” – What is Data?

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

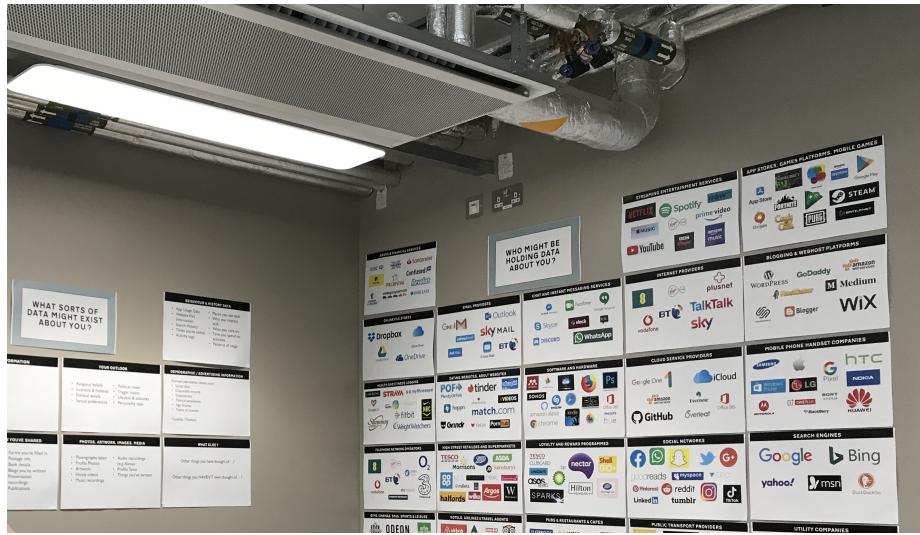


Figure 7: Figure 6: Walls of Data – Sensitising participants to the world of commercially-held data and GDPR

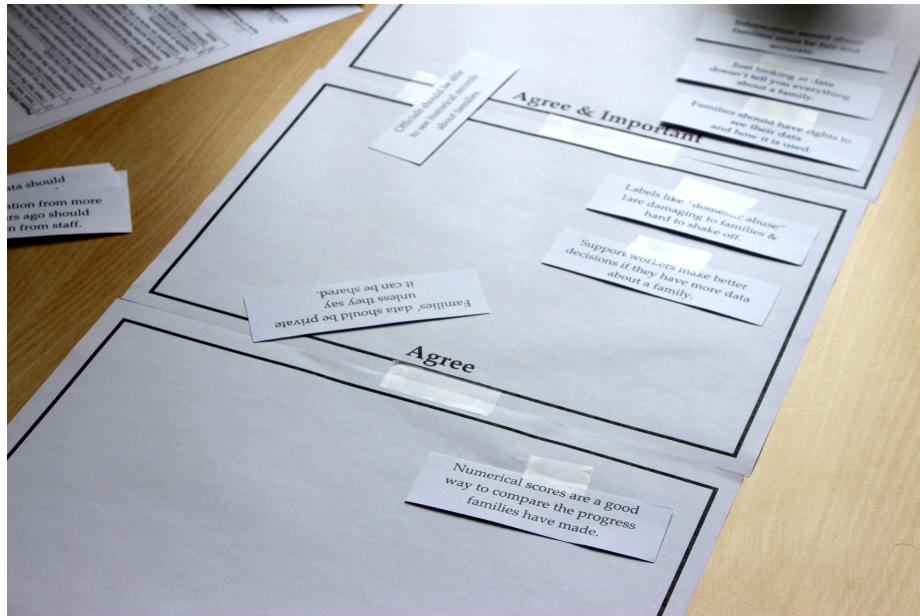


Figure 8: Figure 7: Sentence Ranking – Bringing staff and families to a shared problem space

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

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

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

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



Figure 9: Figure 8: Family Civic Data Cards – things to think with, that can also be used in card sorting tasks



Figure 10: Figure 9: Interviewing Families in the Home – Card sorting with a family in their living room

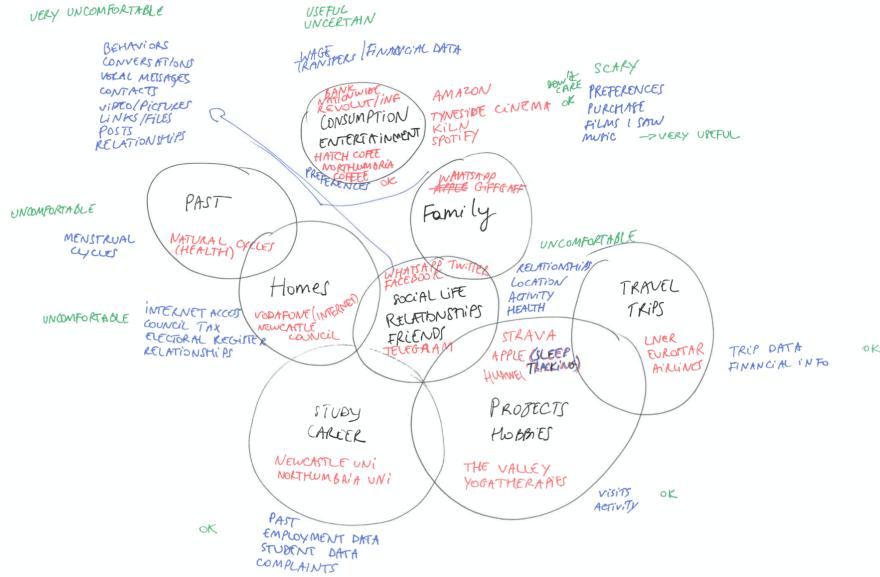


Figure 11: Figure 10: Sketching Dialogue – An example life sketch created by a participant and annotated during discussion

3.5.2 Discussion and Exploration with Stimuli

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

3.5.3 Participatory Co-Design of Possible Solutions



Figure 12: Figure 11: Ideation Grids – Combining random design ingredients to generate new ideas

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

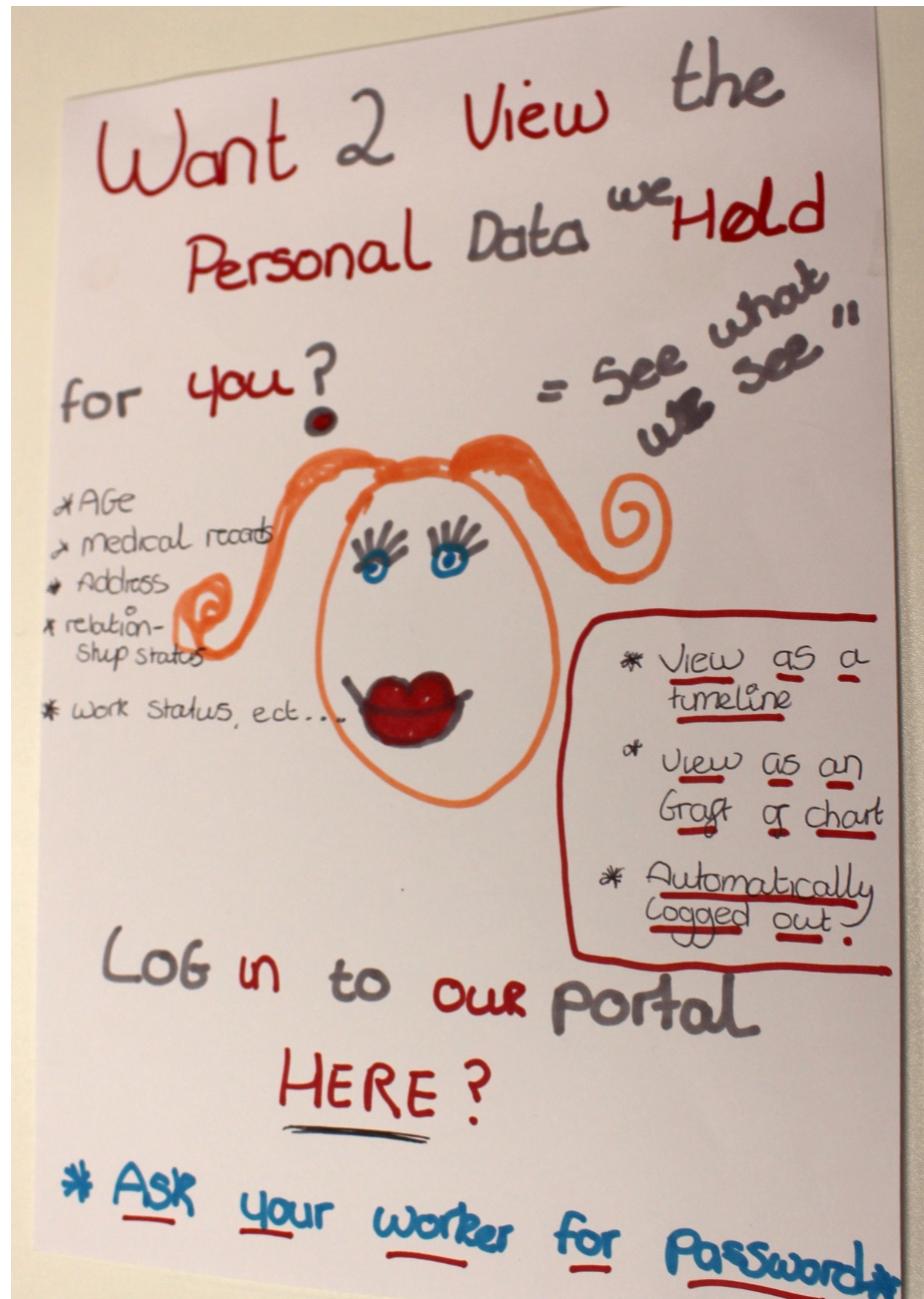


Figure 13: Figure 12: Group poster design – A participant-designed poster to advertise features of imagined data interface products



Figure 14: Figure 13: Storyboarding cards – A mutually constructed narrative created through discussion from a palette of possible parent and staff actions

them into a previously unconsidered solution space.

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

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

3.5.4 Practical Data Experiments, Interface Design and Prototyping

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

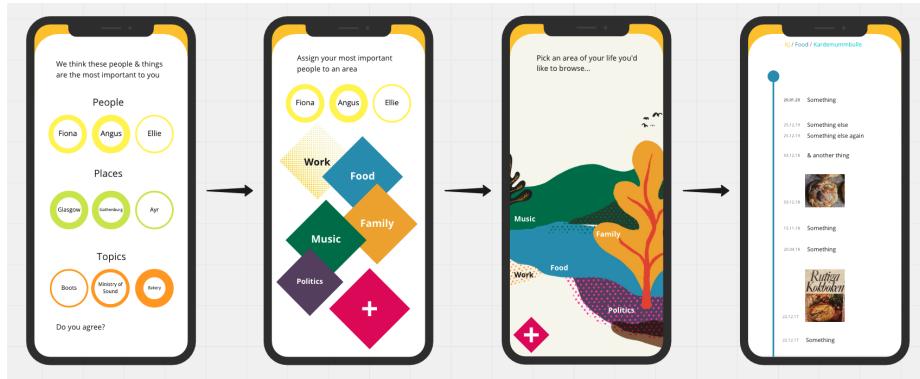


Figure 15: Figure 14: Visual Design Mockup for Life Partitioning in a PDV – A visual design mockup collaboratively created with BBC Research colleague Jasmine Cox

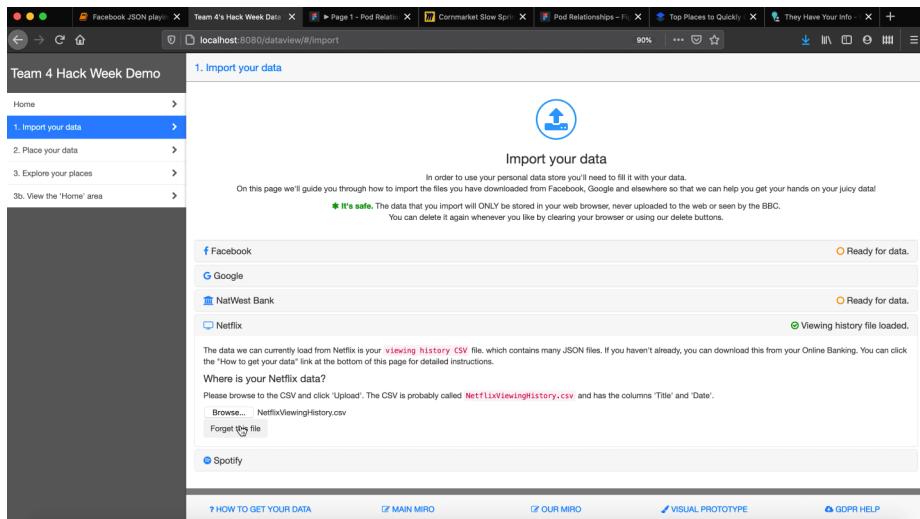


Figure 16: Figure 15: Prototype interface for GDPR Data Viewing – A working prototype that I developed during a hack week at BBC R&D

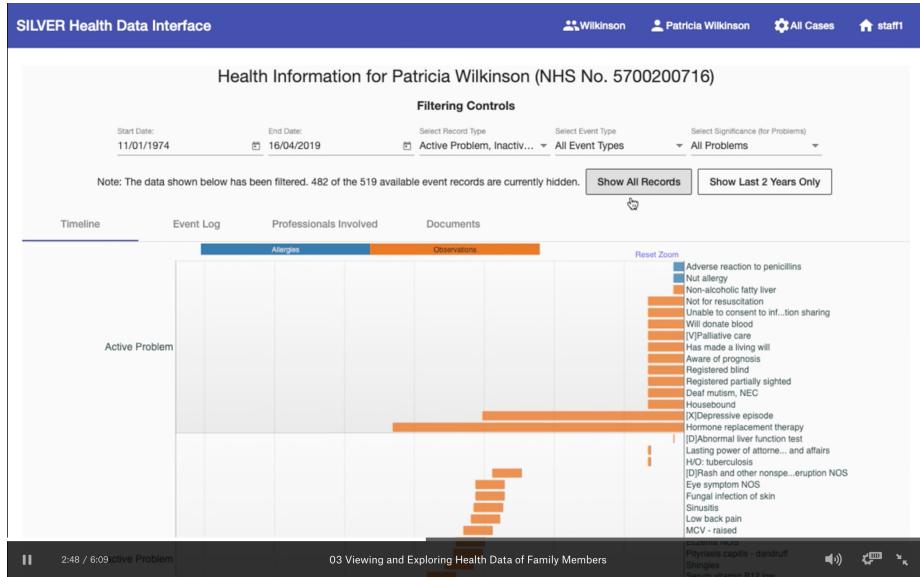


Figure 17: Figure 16: SILVER Health Data Viewing Interface – A working health data viewing interface for Early Help support workers that I developed as part of the SILVER project

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

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

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

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

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

3.5.5 Analysis, Modelling and Learning

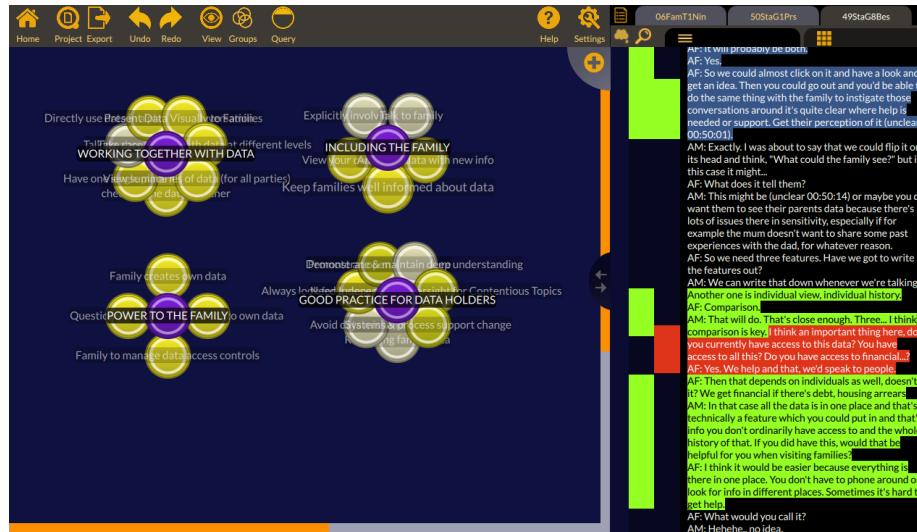


Figure 18: Figure 17: Thematic Analysis – A screenshot of thematic coding of qualitative data using Quirkos for Case Study One

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



Figure 19: Figure 18: Quantitative Analysis – A screenshot of spreadsheet-based quantitative analysis of interview data from Case Study Two

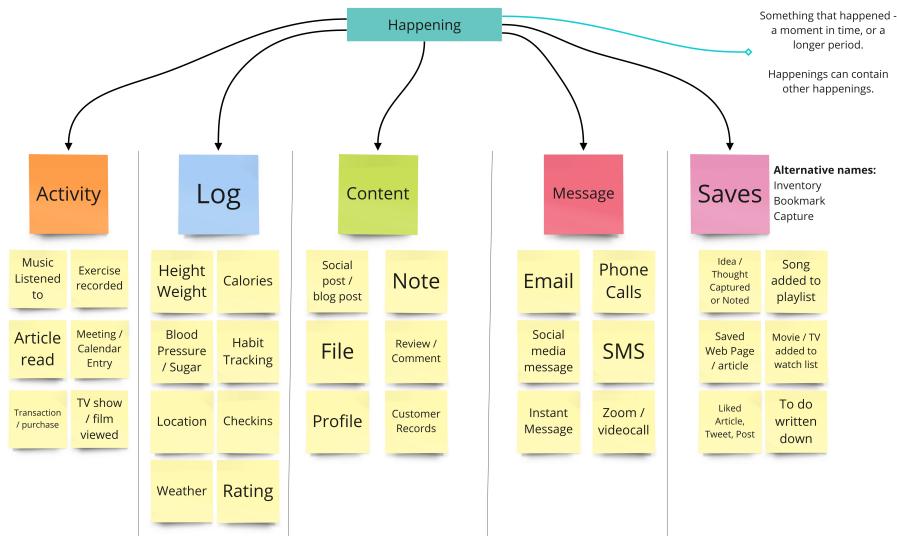


Figure 20: Figure 19: A Model for Personal Data – Developing a common model for personal data imported into a PDV a part of BBC Cornmarket R&D work

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

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

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

3.5.6 Recruitment

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

Research Activity	Engagement	Stage or Phase	Duration	Number of Participants	Recruitment Method
Pilot (prior to PhD): 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 supporters & related staff	Selected by Local Authority Care Services

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

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

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

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

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



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

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

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

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

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



Open Lab
at Newcastle University

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

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

Figure 21: Figure 20: Recruitment Poster – Poster used to recruit participants for Understanding Family Civic Data study

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

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

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

In all cases⁶ 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.

⁵One 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.

⁶(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.)

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

[TODO CUT THIS SECTION AND COVER IT IN INTRO]

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

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

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

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

"If you can see the invisible, then you can see the possible and provide the opportunities for trust, commitment and ways of empowering others to manage their past, present and future." - Ken Wyatt (Australian politician, teacher and equal rights campaigner)

In this chapter, I describe the first major case study of this PhD, in which I ran four 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, a targeted social care provision offered by local authorities to '*at risk*' families across the UK. 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, 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 (4.4.3). 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

[TODO add section about Pilot study, or refer back to one in Chapter 3]

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 [ADD 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, Baines and Wilson (2013)], later expanding to include their families (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* [Cornford, Baines and Wilson (2013); REF Crossley] 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*⁷ – 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 (Cornford, Baines and Wilson, 2013; Barbosa Neves and Casimiro, 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 (Corra and Willer, 2002). 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 (Field, 2010; OFSTED, 2015; Bate and Bellis, 2018; Department for Education, 2018).

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

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

⁷The 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.

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’ [ADD 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 (OFSTED, 2015). 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 (Morozov, 2013) 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);REF White and Wastell;REF Crossley]. 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 (Cornford, Baines and Wilson, 2013) 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, Baines and Wilson, 2013; Lowe and Wilson, 2015). 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, Baines and Wilson, 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: 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.
Education	Children	Parentage, adoption, fostering, childcare.
	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	Salary, savings, credit cards, spending, debt
Civil	Family Finances	Job history, periods of unemployment, performance at work, NI, PAYE, pensions.
	Employment	Council house provision, eligibility criteria.
Crime	Housing data	
	Legal documents	Birth/marriage/death certificates, citizenship/immigration status, work permits.
Medical	Criminal records	Arrests, cautions, offenders' registers, prison time, speeding tickets, spent convictions.
	Court orders	Restraining orders, lawsuits, custody, ASBOs.
Leisure ⁸	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.
Leisure ⁸	Medical conditions	Diagnoses, diseases, allergies, blood type.
	Mental health	PTSD, breakdowns, depression, sectioning.
	Addictions	Substance abuse, gambling, rehab, crime.
Leisure ⁸	Library Usage	Books/CDs borrowed, computer access.

⁸Some 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.

Category	Type of data	Examples/Details
	Sports & Health	Gym usage, class attendance.
	Shopping	Loyalty cards, store & online purchases.
	Habits	
	Transport	Buses used, ANPR tracking, walking patterns.
	Data	

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 colours 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 homes⁹ 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.

⁹The 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.

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 (Morozov, 2013) framing, the SILVER project gave priority to the support worker perspective as its requirements 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’ nor to pursue a purely data-centric solution; I wanted to explore whether it might be possible to satisfy the needs of both parties and to maintain focus on human-centricity and the need for a balanced relationship.

4.2.4 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 emerged in part from the participants in the first phase (see below) of the research engagement, and became a main focus for the second phase, so that we would not only be exploring RQ1 and especially RQ2 in the context of current practice, but also be asking participants to imagine a different set of practices that might potentially serve their needs better. In doing so, we would be able assess whether the imagined 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 be helpful as it would put participants in a speculative, co-design mindset that would elicit deeper insights about how civic data *should* be used, not just expressing opinions on how it was used currently.

4.2.5 The Workshops

Table 4: 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 (in total)	- 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 SECTION HERE]

PENDIX]. 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 (discussed further in 4.3.3.1), 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.

[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] [TODO: use a different word than ‘agreement’ within coloured boxes to avoid confusion; explain ‘neutral’]

Having completed the sentence ranking sensitisation activity, participants went on carry out the other co-design activities as detailed in Table 4. Findings from the analysis of these activities’ transcripts is 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 the coded texts were 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

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>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 & NOT IMPORTANT	NEUTRAL	NEUTRAL & IMPORTANT
	<p>MODERATE AGREEMENT "Any personal information from more than 5 years ago should be hidden from staff."</p>	

Figure 22: Figure 21: Participants' Shared Values Deduced from Sentence Rankings Data

is further detailed in sections 4.3.2 to 4.3.4, including participant quotes¹⁰.

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”)¹¹. 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.

¹⁰The 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.

¹¹As judged at the time of the workshops - summer 2018.

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

Subtheme	Description & Quote	Status
Understandable Information Summaries	To maximise understanding, simple summaries of the information within families' data should be available to both families and support workers. Visualisations should be used to ease comprehension, and information should be contextualised at different levels (individual, family, community). “ <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] “ <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]	Emergent
Interact With Data Together	Support workers should work to actively counter the knowledge imbalance by informing families what their data says. They should make use of specific datapoints as talking points to aid planning conversations. “ <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]	Emergent / Imagined
Direct and unified data access	Individuals should be able to directly access their civic data through a personal interface; this should be a single, common place where all of an individual or family’s data is brought together to give a complete and consistent overview to all parties with a need to know. “[<i>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]	Imagined

Subtheme	Description & Quote	Status
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. “[<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: 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. Worker A: “<i>You should never make a judgement on data... that data could be wrong.</i>” Worker B: “<i>It takes individuality, working with that person as well, doesn’t it?</i>” [SQ11]</p>	Current / Emergent
Checking 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. “[<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

Subtheme	Description & Quote	Status
Changing Lives Means Changing Data and Changing Consent	Recognising that families' lives are in constant flux, routine reviews of data should occur, and they should be invited to regularly review their choices regarding data collection, keeping and sharing. All systems and processes should treat data as fluid and flexible, not static unchanging facts. Feeds of recent changes should be available to both parties. “[<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]	Imagined
Individual Agency & Family-sourced Data	Individuals should be able to create or contribute their own data to tell their own story and annotate particular datapoints with their own explanations. Worker A: “ <i>If you read information [...] about me, you wouldn't expect to meet the person you meet.</i> ” Worker B: “ <i>That's it. It's the same for everybody.</i> ” 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> ” Worker B: “[<i>I'd like it if you could] give your bit of personal data, your own story.</i> ” 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]	Imagined
Granular Access Controls	Families should be given controls to manage access to their data and configure and change preferences at a fine-grained level. “[<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, FQ16]	Imagined

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

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

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

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

visually impaired [SQ45] to aid understanding. Visualisations also need verbal explanations [CQ11].

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*' [FQ17, S5]. Bringing together data from multiple sources would allow patterns to be spotted by correlating data from different sources, which workers perceived would help their preparation: "*[This imagined interface] would provide individual histories but you could also pull them all together so you can prepare, so for instance if mum was having some significant issues with mental health, you might be able to correlate the [child's] school attendance alongside that and find out why that's happening.*" [SQ8]

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 [SQ77], data must be current and complete [SQ13], but this state can only be achieved with the family’s cooperation. Looking at data will never provide support workers with a complete understanding. Yet, workers often ‘*tend to just trust that everything that has been put down is right*’ [CQ1], allowing the data perspective to dominate. Such assumptions should be avoided [SQ10]; processes must recognise maintaining human face-to-face dialogue as a priority. Data should only provide supplementary insight: “*You should never make a judgement on data... that data could be wrong. It takes individuality, working with that person as well, doesn't it?*” [SQ11]. All participants presented with the sentence “*Public sector officials can make good decisions just by looking at a family's data*” [S18] disagreed with it.

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

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

Many participants concluded that only ‘relevant’ information should be available, to those who ‘need to know’, but the wide range of opinions 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, SQ76]. 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 Checking Data Together The idea of families and support workers reviewing data **together** arose from many of our participants in workshops A and B, and this led us to explore this concept of ‘shared data interaction’ in more depth through the storyboarding exercise in workshop C (see 4.2.4 above). Families perceived value in having not just data representations (as in 4.3.2.2 above) but a data interface present within their care meeting, so they that they could see actual data and have it explained to them. One practice embodying the concept of transparency that is emerging in some care services is the use of ‘2-in-1’ devices (laptop/tablet hybrids) within the care engagement so that workers can visibly record data in front of families and then ask them to ‘*approve*’ the accuracy on screen [FQ12, SQ67]. Participants believed this would help to build trust between the support workers and families; if a family begins to feel powerless, they may disengage [SQ35], but even minor involvement, such as this emergent practice of signing off approval of data records [FQ12] or an imagined process of checking & correcting data records together (see next section) could make families feel more empowered which could make the support relationship more productive.

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

data accuracy. This does not mean free editing of records (as, for example, as discovered in the earlier study (Bowyer *et al.* (2018)) fears and/or self-interest could lead to families misrepresenting themselves in data) but rather taking a role in reviewing, annotating, explaining, or requesting changes, through direct data-centred collaboration between involving 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]

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

4.3.3.3 Changing Lives Means Changing Data and Changing Consent

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

Worker A: “*We would get a report through to say...* ” Worker B: “*They’ve recorded something.*” Worker A: “*Yes. Then I suppose we would follow it up [...] face to face.*” [SQ65]

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

4.3.3.4 Individual Agency & Family-sourced Data

The idea of families reviewing data has significance not just for how it can help within the support interaction, but because it can give families an independent role in their data ecosystem. Both families and support workers imagined the family being able to interact with their civic data on their own, something that is currently not possible. This is a vital step for empowerment: if something goes wrong, families must be able to discover this and must be able to do something about it. Without a cycle of feedback involving individuals as stakeholders having the ability to

review and correct data, data will quickly become inaccurate (Pollock, 2011). Thinking about data interaction at home unlocked additional thinking, such as families helping to fill gaps in data [SQ57] or contribute new data that may not otherwise be recorded [SQ58]. Giving families the ability to contribute new data would empower them to '*tell their own story*' [FQ10]. Many participants saw this as-yet-unavailable capability as expected common sense:

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

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

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

With new ways for self-expression, families could add context for support workers [FQ9, SQ55], unlocking new support topics [SQ56]. The overriding sense from both groups was that families having the ability to annotate or explain their data would allow them to hold authorities to account, and empower them to tell their story and '*show the real me*', as illustrated in [FQ10, shown in Table 7 above].

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

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

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

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

Looking at Theme 2 as a whole, 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.

4.3.4 Theme 3: Earning Trust through Transparency

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

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

“I think that [families] have got a right to know what is held about them and what is said about them.” [Worker, SQ50]

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

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

The current approach, which relies on the support workers mentioning data that they consider relevant, can reassure families when they are kept thoroughly and

regularly informed, but can result in expectations being broken by accidental sharing of information if its sensitivity is overlooked:

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

Data must be handled respectfully, with attention to family and individual privacy. A lack of transparency and trust can lead to an atmosphere of suspicion [SQ17] where families have ‘*a totally overwhelming feeling of people checking up on them*’ [SQ71] and apply extreme scrutiny to what they are told: “*You can get families who [no longer] believe what’s being said about them.*” [Worker, SQ73]. Fearful of consequences [SQ72], families may withhold information:

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

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

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

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

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

4.3.4.2 Always Seek and Demonstrate Greater Understanding In order to earn, build and maintain trust, support workers must always be seeking to form a completer and more up-to-date picture of the family, in line with the finding in 4.3.3.1 that individuals are more than what is stored in their

records, and this requires human interaction to uncover. Demonstrating a deep understanding of the family, and that a family's lived reality has greater priority than what a database says can be a critical to trust-building: “*You don't want to reduce them to this number in a database. You want to understand their actual experiences and support them in getting better.*” [Worker, SQ74]. It is important that families understand workers' good intentions when accessing data about them [FQ15]. However, if workers had to show all available data to families this could make it challenging to maintain good relations, “*because literally [the data we have] is like everything, isn't it? So I don't know how I would feel...*” [Worker, SQ21]. In addition to avoiding breaches of expectations (see Theme 2 above), a transparent approach ensures that the privacy of families is respected, because data is not used in decisions without the chance for explanation:

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

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

4.3.4.3 Pro-actively Challenge Data-centric Norms Exploring this need for reassuring dialogue in more depth, we see that to avoid damaging negative spirals of emotion, deliberate openness is needed from support workers (and the entire care system) [SQ18] as to what information is held, and how it will be used and shared, in order to alleviate fears of data being used ‘*against*’ families that can arise without that transparency – giving them instead confidence that their interests are being protected, thus putting them at ease [SQ20]. To our understanding, data handling processes is only done once in very loose terms during initial engagement with a family for the purposes of collecting informed consent and rarely revisited. We found that workers could easily imagine explaining data practices in greater detail than they currently do [SQ41] and clearly there is a need for proactive action by workers to counter the inherent knowledge imbalance of data being collected into systems that they are gatekeepers for.

Workers however lack control over the quality, coverage and timeliness of the family data and see this as a systemic issue they could not adequately address. From my experience with early help teams through the SILVER project (see 3.4.1.1) it became clear that while support workers can see more data than most, they have far from the complete picture; in fact, there is no one organisation or individual with visibility of the entire family-information ecosystem, suggesting that greater openness with data would benefit not just the family, but other civic actors involved in the family's lives and in their care. Some participants

suggested that openness about data handling needs to accompany data access, so for example if browsing information together (as described in 4.3.2.2 and 4.3.3.2), it would be important to explain where the information has come from and why the support worker has it, rather than just reporting its content:

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

As mentioned in 4.3.3.3, there is a need to replace the current practice of treating consent as a one-off formality at the start of the support process with something better. In our earlier study we identified this as a need for *dynamic consent* (Bowyer *et al.* (2018);Kaye *et al.* (2015);Williams *et al.* (2015)). A common heuristic expressed by families here and in the earlier study is that data should only be seen by those that ‘*need to know*’, but this is very hard to achieve: first, because without transparency of data handling, a family cannot verify whether this is happening, so has to rely only on feelings and supposition to inform their trust. Second, the need for fair judgement over who should access families data is objectively important given that some support workers expressed a belief that their right to access families’ data should overrule families’ consent:

Worker A: “*I think to enable us to work with families, we need to have as much information to give them the best possible service. So, I think we should be able to [access their information] regardless of what families say.*” Researcher: “*Regardless of what they say?*” Worker A: “*I do, yes.*” Researcher: “*Does everyone feel the same way then, that they don’t get a say?*” Worker B: “*Yes, because you need as much information as what you can.*” [SQ22]

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

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

4.4 Discussion

Through the workshops described in 4.2.5, I have successfully advanced my understanding of the human experience of data (RQ1) and the role of data

within service relationships (RQ2), specifically for the Early Help context. In the section, the findings described above in 4.3 (and the preliminary findings in 4.2) will now be contextualised in respect of existing literature from Chapter 2 and beyond, drawing conclusions as the value of involving people with their data (4.4.1), the need for human interaction to make data access effective (4.4.2), and the possible impacts of a shared data interaction approach in terms of shifting the locus of decision-making closer to the supported family (4.4.3).

4.4.1 The Value of Involving People With Their Data

The above analysis of attitudes to data usage in the UK early help context reveals that data about supported individuals and their families is already an integral part of current care practice, providing great value in 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, just as in the commercial sector (see 2.1.2 on dataism), families' civic data is considered as a resource to be utilised. This mindset carries an implicit assumption that data is an objective source of truth, which our participants tell us it can never be. 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. The present data-centric approach across civic systems mean that stored data can often serve as a proxy for families' involvement, and without any involvement of the family in checking data accuracy, 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 (Bowyer *et al.*, 2018).

Our findings that trust is critical to an effective support relationship are consistent with literature which states that trust in the independence and integrity of the data-collecting and data-holding institutions is essential (Dijck, 2014). Trust currently rests upon feelings and impressions rather than the true accountability families would get by seeing what data is held and how it is used. This trust is often absent or reduced due to Early Help services not involving families with their data. Families must trust not only the system, but the support worker themselves; our findings suggest 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*' (Cornford, Baines and Wilson, 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, which would be very powerful in building trust. 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 (established as requirements from families in the preliminary study (Bowyer *et al.*, 2018)). 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, Kim and Church, 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. 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 personal data interfaces would unlock 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 - a '*point of severance*' (Luger and Rodden, 2013). In effect, the ongoing access to and direct use of data by families would serve as a practical implementation of a '*dynamic consent*' model (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 onboard with the way their data is being used, since their ongoing awareness combined with the 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 simpler processing and reducing liability. Families will be happier with the use of their data if they can see it, notice issues and speak up when they feel something is amiss. 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; data becomes a resource that both parties make use of to inform their conversation, rather than the support worker's sole responsibility. 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 (Cornford, Baines and Wilson, 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." (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 taking a position of authority, passing judgement and delivering 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 Theme 2); this is also a practical instance of the HDI concept of *agency* (Mortier et al., 2014), and in shifting the power balance toward the family it can also be seen as an antidote to current data-centrism in the system and society at large (see 2.1.2). 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 meaningfully effect change based on what they learn from it. This involves the translation of raw data into meaningful information (see 2.1.1) – through summaries, visualisations and explanations – a need that we have identified even though the creation of information representations such would be challenging as it is not clear who would have the access, skills and mandate to do this. In the designs and desires of our participants we see confirmation that, as described in one of the central tenets of HDI, the information available to the individuals must be *legible* (see 2.3.2 and

(Mortier *et al.*, 2014)) but also that their access must be effective (see 2.1.4 and (Gurstein, 2011)). This includes providing suitable opportunities for access — for example via personal data interfaces and not just within the support meetings — as well as addressing technology, literacy, mental or physical handicaps. Our participants' ideas around audio interfaces are a good example of the extra steps that would be needed to provide effective access for all. Supporting the range of all possible needs means that to be effective, information access **must** be supported by a human relationship — one where someone can both explain the data as well as answer questions about it (see 4.3.2.4). 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; the storage of and access to data necessitates an ongoing conversation between data holder and data subject. The system needs to have a human face or point of contact that the individual may put their trust in and to whom they can address their questions; as others have noted, simply giving access to raw data would be inadequate and limiting (Cornford, Baines and Wilson, 2013).

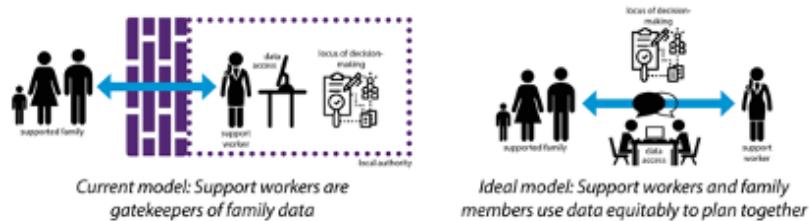
By focussing on the human aspect of the proposed use of data within the support relationship, we can see that as well as improving accuracy, consent and trust, shared data interaction could bring practical benefits by facilitating a better interpersonal interaction. By physically bringing data into the 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 (Star, 1989, 2010; Bowker *et al.*, 2015), just as my Data Cards did within my own research. 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, even if only to query or challenge something at first. Many of our participants talked about how looking at data would provide a discussion stimulus or serve as a conversation starter. This initial use could lead on to using that data, as it changes from meeting to meeting, as a metric against which to measure progress, something which could bring a feeling of reward and accomplishment to the family and contribute to their future success. 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 research (see 3.5.2). In particular in workshop C, 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. This interchange within a research setting, gives some insight into how the dynamics of shared data interaction might work if implemented in practice. Having access to the data within the context of the support relationship is a key enabler of *agency* (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 in line with our findings that regular reviews of consent and data need to take place and that the ability to raise a question or start a conversation at any time is needed, we can consider that the availability of these capabilities on an ongoing basis would satisfy a second HDI requirement, *negotiability*. 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 protection regimes (described in 2.1.3) 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.*’ [S4, see also 4.3.3.1]. 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*’ (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.*”
(Lecluijze *et al.*, 2015)

4.4.3 The Implications of Shifting the Locus of Decision-making Through Shared Data Interaction

In pursuit of RQ2, the four Case Study One workshops and the preliminary research have explored the role of data within the Early Help support relationship (see 2.2.5), looking separately at family and staff perspectives before bringing both parties together to discuss how both parties' goals might be served by a model of *shared data interaction*. In workshop C, we explored the mechanics of shared data interaction at a interpersonal, sociotechnical level (see 2.3.3), mapping out a possible narrative in terms of human-human and human-data interactions. I present here a model for understanding why this could be important for rebalancing power between the supported family and the state, based upon a concept I have developed called *shifting the locus of decision-making* (LDM). This concept is distinct from *locus of control* (Spector, 1982) which normally refers to personal willpower, and *locus of power*, which refers to the concentration of power within an organizational hierarchy. LDM refers to **the place where decisions are made**, and it may or may not coincide with existing authority structures. A pattern can be stipulated, in which decisions are typically made, germinated or championed **close to where data is accessed**. In an effect that has been expected since as early as 1970 (Klatzky, 1970) the increasing use of data in services across private and public sectors (a phenomenon detailed in 2.1.2) has concentrated the LDM with data holders, who collect service users' data to serve their own purposes.



[TODO: Replace with higher resolution version]

The current and imagined approaches are shown in Figure 22 above. In the current model (left), all access to data by families is through the support worker as gatekeeper, who decides the scope, content and nature of their access – here the LDM is effectively locked away from the family's participation. The use of data by families is limited because any data must flow through the support worker as gatekeeper. In a more equitable model (right), both support worker and family member are positioned as allies looking at the data together. This model changes the nature of the support relationship, as 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) would now take place in a different context – the two-party context of the support meeting itself. The removal of the gatekeeper role redistributes the power to interpret, select and judge data much more equitably between the two parties; families would no

longer be prevented from participating in data-based decision-making. I theorise that shifting the data access from the domain of the support worker to the shared domain of the meeting between the two parties, would therefore move the LDM closer to the middle of the relationship, where it will rest at the heart of the support relationship, creating a more balanced relationship and increasing families' agency and power. Within the findings above we see evidence that both families and staff would value a shared data interaction approach, with multiple participants independently suggesting potential benefits that could be gained by techniques such as reviewing data and consent together (4.3.2.2, 4.3.3.2, and 4.3.3.3). While participants perceive shared data interaction as an improvement, such an approach has not been tested in practice, so it is important to consider what the benefits and implications of such a shift might be:

The potential benefits 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 (Bowyer *et al.*, 2018). 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 only indirectly possible, if at all. By shifting the locus of decision-making, families could 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 (see 2.2.3 and (Abiteboul, André and Kaplan, 2015)), thus '*encouraging the family to take full accountability for their own responsibilities*' as one support worker put it [SQ75]. In their 2016 paper, Crabtree and Mortier also recognise the importance of exposing individuals to actual data if accountability is to be achieved (Crabtree and Mortier, 2016). The perceived benefit of individuals directly using data-based interfaces for health and wellbeing are already accepted, with 93% of doctors believing that apps can improve health outcomes (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 LDM 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 provided the data subject remains engaged, informed and understands the data and processes that exist, 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 4.1.2) is not well-suited to the creation of such unified data interfaces due to its fragmented nature (Copeland, 2015). Identity management in this context is already very challenging to negotiate (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 (Honeyman, Dunn and Mckenna, 2016) and in other countries such as Poland where it is deemed critical (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 (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 (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 or misleading workers for personal gain). While very unlikely to be a mainstream issue, this is a fringe possibility that any process or system must still consider 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 [Copeland (2015); ADD REF Local Government Association]. The state has increasingly adopted a data-centric approach to citizen interaction in part because it cannot manage to provide human relationships with every individual citizen. But now this approach has become ingrained into government approaches to citizen relations — *'it is no longer a technological necessity but it has become a political intention'* (Bridle, 2016). What we have identified is that there is a need to reverse this trend, not just in practice but in political ambition, if people's interests are to be best served, and if a welfare state is to be truly *enabling* (Miettinen, 2013). By taking a more innovative approach to digital policy, it is possible that governments could be more effective in helping to involve those citizens that have become disadvantaged by the current system – a more human-centred approach could help to combat the digital divide (Kalvet, 2005; Steyaert and Gould, 2009).

My model that shifts the LDM is theoretical; it does not yet provide an implementable solution that could be rolled out at scale, rather it should be thought of as useful mental model to stimulate further discussion about how care providers could or should change their processes and systems. The value of this contribution is that it shines a light on the positive and negative impacts of current data-handling and data-use procedures upon relationship effectiveness, and identifies imagined practices that could be preferable and more efficient than current practice. The 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 in pursuit of the primary goal of Early Help; to empower families to help themselves as effectively as possible.

4.5 Summation

Through four participatory co-design workshops with supported families and support workers in North-East England, I 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 fragile, 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 even just to see themselves in data as the state sees them, opportunities are missed to truly empower families to be better represented and to better themselves.

Through these explorations of shared data interaction and personal data interaction, I have shown there is both a need and a desire for a new approach. 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 over families' civic data 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 could 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.

Further benefits of a shared data interaction approach have also been uncovered; 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 currently 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, much like the reflection capabilities we see in the self-informatics space. The shift from support workers reporting what the data says to '*looking at data together*' would help 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 a combination of data and dialogue. Individual family members would 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 data that is as accurate as possible, the quality of decision-making would naturally improve and conversations would be likely to be more effective and efficient as they would 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, treats the family with more respect 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, would allow 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 to open up new opportunities to individuals for using their data for their own ends in ways that were not previously possible. It is through the adoption of such measures that we could begin to facilitate the emergence of a human-centred personal data ecosystem (as described in 2.3.4) in a civic context.

In exploring the usage of civic 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 Human-Data Interaction itself needs to be developed as a concept. As a sub-field of Human-Computer Interaction, HDI is largely considered in the traditional context of interacting with data through an interface, but this work, which has, guided by our participants, focused less upon layout and screen interaction and more upon the wider sociotechnical context of the support relationship, suggests that HDI can be more effective when the word 'interaction' is considered in an interpersonal sense, and these insights begin to address the research gap identified in 2.3.5, to define the research agenda for human centricity in practice. Informed in part by this idea I have explored further in a workshop paper how the HDI field needs to advance to consider the sociotechnical level as well as the interface level, which is outlined in (Bowyer, 2021).

Capabilities – or their absence – matter more than the on-screen technicalities of the data interaction. Data interfaces are limited by their operating context as to how much they can offer, but considering data interaction as a sociotechnical process, including the wider human-facing relationship between the individual and the representative of the state as well as the data interface itself, 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 the sentence ranking exercises I have been able to gather a snapshot overview of what this sample of support workers and supported families think about data, and where they agree and disagree (see Figure 21). The detailed analysis of 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 best practices, seen in the subthemes of sections 4.3.2, 4.3.3 and 4.3.4 and expressed in our CHI 2019 publication as 38 specific practices for Early Help services (Bowyer *et al.*, 2019), many of which are currently imagined or only just emerging. Participants believe these best practices would improve families engagement and the support they receive. 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 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 this work serves as a reminder that as we move into the data-driven age it is important that **data should stay close to the people it is about**, rather than to those that use the data to provide services, and that service practice and processes should remain human-centric rather than data-centric.

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

5 Case Study Two: The Human Experience of GDPR

“The Crystal Wind is the storm, and the storm is data, and the data is life. You have been slaves, denied the storm, denied the freedom of your data. That is now ended; the whirlwind is upon you Whether you like it or not.” – from ‘*The Long Run: A Tale of the Continuing Time*’ by Daniel Keys Moran (computer programmer

and science fiction writer)

In this chapter, I will describe the second major case study of this PhD, in which I took 11¹² participants through an longitudinal in-depth one-on-one process of three interviews with coaching and support in between, with the total engagement per participant lasting approximately 4 hours over a three month period. The purpose of the research was gain a deeper understanding of people's attitudes to the kinds of personal data held by companies in people's everyday lives and what they want from that data (in pursuit of RQ1) and specifically to examine the human experience of existing in a data-centric world (see 2.1), with each individual having a number of relationships with service providers that involve the use and holding of personal data; in line with RQ2 the goal is to better understand the role of that data in those relationships. In particular, having gained an initial understand of attitudes, hopes and expectations, a further objective was to examine how those expectations might change during the journey of digital life mapping, data request making, receiving and examining of data, and scrutiny of responses, collectively forming a holistic understanding of "the human experience of accessing your data with GDPR."

In section 5.1, I will expand on chapter 2 to explain the context of using GDPR in research as a means to retrieve personal data. In 5.2, I will explain the stages of the interview process (including details of how participants were sensitised) as well as the preparatory and intermediate steps I undertook as researcher. In section 5.3, I will explain the model of personal data types developed for this study, and will present quantitative and summary data from the interviews, explaining how participants' GDPR access requests progressed, highlighting participants' shared hopes and goals, and examining in particular how their perceptions of power and trust were affected by the experience. In section 5.4, I will describe the three themes uncovered through thematic analysis: that organisations provided participants with insufficient transparency to meet participants' hopes and their legal obligations (5.4.1), that people struggle to find meaning and value in their data when they do manage to access it (5.4.2), and that providers' data practices (in particular their GDPR request handling) can be harmful to their users' trust, but that greater openness can have an opposite, positive impact (5.4.3). I will discuss the implications of these findings with reference to prior literature, from the perspective of policymakers (5.5.1), data-holding companies (5.5.2), and individuals (5.5.3). Finally in 5.6, I will summarise these insights in terms of how they can advance our understanding of the research questions and their wider significance.

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

5.1 Context: Accessing Your Personal Data Using Your GDPR Rights

5.1.1 The Current Need for Data Access

[TODO add visual on The Types of Personal Data]

As established in 2.1.2 and 2.2.4, people live digital lives, inevitably involving the use of myriad digital services that collect personal data, which is subsequently mined for value and exploited at scale, creating an imbalance of power between data holders and data subjects, and a exclusionary landscape around data use which is difficult for individuals to navigate: having acquired data about individuals, this becomes a focus for service providers' decision-making and customer relations become less important. This everyday context is the chosen research setting for this case study.

Section 2.1.4 established how unaware many people are of this imbalance around data, that there is a want¹³ for effective access to data to restore individual agency. As described in section 2.1.3, policymakers have been attempting since the 1970s to introduce legislation to tilt the balance of power back towards individuals, most recently and most notably the European Union's General Data Protection Regulation, which legally endows at least 513 million individuals¹⁴ with new rights to timely data access, explanation, erasure and correction (Information Commissioner's Office, 2018).

Data protection and misuse issues have grown in the public awareness since the Snowden revelations in 2013 (Gellman, 2013), and have become even more important following the Cambridge Analytica scandal in 2018 ('Facebook–Cambridge Analytica Data Scandal', 2014; Chang, 2018), which may have resulted in manipulation of voting outcomes through personal data use, and the COVID-19 pandemic in 2021 (O'Donnell, 2020; Hamon *et al.*, 2021). Since the GDPR's launch in May 2018, it has undoubtedly resulted in new data access offerings; many large consumer companies have developed 'privacy hubs' or improved privacy policies where individuals¹⁵ can learn how their personal data is handled or access data download portals to easily download copies of it ('Privacy - Apple (UK)', no date; 'Privacy & Terms – Google', no date; 'Privacy', no date; 'Facebook - Data Policy', no date). Almost all data controllers and processors have

¹³In this study and throughout this thesis, my usage of the word 'want' in the context of data capabilities deliberately includes both meanings of the word: the need or desire of the individual, but also that which they **lack** (see opening of Chapter 6).

¹⁴In this study and throughout this thesis, my usage of the word 'want' in the context of data capabilities deliberately includes both meanings of the word: the need or desire of the individual, but also that which they **lack** (see opening of Chapter 6).

¹⁵At the time of writing (summer 2022) the GDPR legally applies in both the European Union and the United Kingdom (which adopted a copy of EU legislation when it left the EU in 2020), which have a total population of 513 million individuals [37]. GDPR rights are also conferred to any individual who is a customer of businesses with registered offices in EU or UK countries, meaning that these rights are in effect globally available for non-EU, non-UK users of many multi-national digital service providers.

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

However, it is not known how effective these offerings and processes are for service users, and how individuals feel about them in light of this backdrop of public concern. No service providers make data access statistics publicly available, but anecdotal reports from industry insiders suggest GDPR access rights and data download dashboards are not well-known and hardly used. This presents an opportunity to take individuals who have not previously used these capabilities on a journey of discovery that might enable us to assess the impact of these processes over time and whether—by compelling data holders to create such offerings and respond to access requests—GDPR succeeds in its goals to ‘enhance the data protection rights of individuals’ (Council of the European Union, 2015) and to give people ‘control over their personal data’ (The European Parliament and the Council of the European Union, 2016b).

5.1.2 Current GDPR Research and its Limitations

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

The GDPR process itself has also been examined from many perspectives by researchers: to understand data holder’s compliance with legislation (Ausloos and Dewitte, 2018; Arfelt, Basin and Debois, 2019); to evaluate data portability (Wong and Henderson, 2018) and ‘privacy by design’ (Waldman, 2020); to compare its effectiveness in public/private sector contexts (Quinn, 2021) or in improving explainability (Hamon *et al.*, 2021), fairness (Kasirzadeh and Clifford, 2021), consent (Human and Cech, 2021), transparency (Spagnuelo, Ferreira and Lenzini, 2019) and the reduction of data breach risks (Gonscherowski and Bieker, 2018). Potential negative impacts have also been considered; the GDPR could be seen as a threat to privacy (Bufalieri *et al.*, 2020) or as an impediment to health research (Clarke *et al.*, 2019).

Clearly the GDPR has spurred a broad variety of research, spanning legal, social and technology domains. Yet, there is scant research into the individual human

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

At a more fundamental level, there is a need to understand the *experience* people have when using the GDPR; companies’ GDPR processes have been designed to comply with litigation but often with insufficient design thinking (Cormack, 2021). GDPR-handling processes and data access systems have been motivated by a need to comply rather than by focusing on individual needs or desires (Abowd and Mynatt, 2000; McCarthy and Wright, 2004; Wright and McCarthy, 2008) (for more details on experience-centred design refer to section 3.2). It is highly likely that many user needs or desires have been overlooked. Such experiential understanding could inform the design of improvements to companies’ GDPR mechanisms (be they interface interactions or response-handling procedures), as well as identifying specific needs that might be best met through improvements to policy, including to the GDPR itself.

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

Given the fact that data-centric services now span all aspects of our lives, and the amount of personal data about individuals has grown, it has become critical to think about the way people interact with data as a ‘whole life’ problem. This is one of the reasons this study focuses on the layman rather than a particular demographic, and ‘everyday services’ rather than a particular domain. Data has transcended the machine and now encodes facts about our lives, it exists across devices and across providers (Weiser, 1991; Mydex CIC, 2010; Abowd, 2012). This means that personal information management has become a sociotechnical problem (see section 2.3.3), that can no longer be solved as a filing-and-retrieval problem as per traditional PIM approaches (see 2.2.2), but only when considered as multi-party negotiation over representation, ownership, access and consent. It is important to evaluate the GDPR in this context. Up to now, individuals

have not had the means to participate in or initiate such negotiations. On paper, it would seem that GDPR rights do convey this capability, but it is not known whether in practice, service providers' responses to GDPR can actually deliver data subjects the ability to take part in negotiations around data in a fully-informed way. While some research on relationships around data and data as a shared resource is now emerging (see 2.2.5), the relationship with data-holding service providers has not been examined in this way.

A roadmap for best practice in this space can be found in the emergence of the 'personal data ecosystem' concept (see 2.3.4). Researchers have identified that a human-centric approach to personal data is needed, placing individuals at the centre, as controllers and overseers of their own personal data (Mydex CIC, 2010; Symons *et al.*, 2017). This is an emergent space of much activity and research ('Human Data Interaction Project at the Data to AI Lab, MIT', 2015; BBC R&D, 2017; MyData, 2017; Symons *et al.*, 2017; MyData.org, 2018; 'HDI Network Plus, University of Glasgow', 2018; 'HDI Lab, Heerlen', 2020) and provides a strong framing for us to evaluate the human experience of—and interaction with—the GDPR; given people's diminished agency and control over their data (Woolgar, 2014; Crabtree and Mortier, 2016), do the GDPR's access rights, as implemented by service providers, provide the effective access (Gurstein, 2011) people need? Does the GDPR help people to achieve legibility, agency and negotiability, the three tenets of Human-Data Interaction (see section 2.3.2 and (Mortier *et al.*, 2014))?

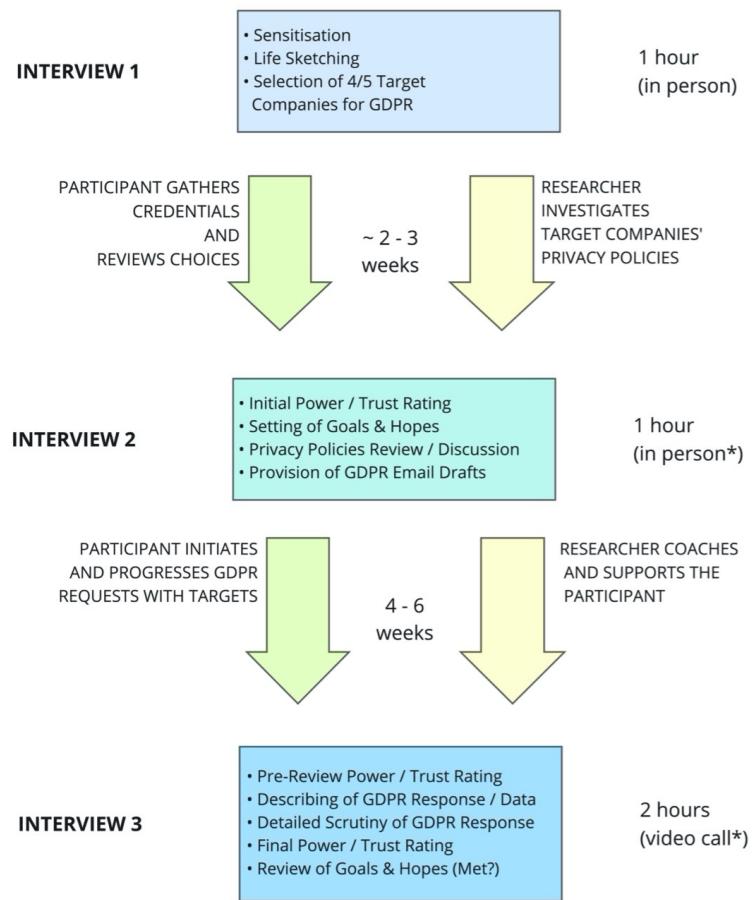
This case study aims to explore the research gap identified in 5.1.2 above, from this perspective of greater human-centric need in a sociotechnical multi-party data use context. It will do so by scrutinizing the experience of using one's GDPR rights to discover how well the process meets individuals'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

5.2 Study Design and Configuration

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

Each participant's journey progressed at its own pace (see Figure 23) with participants invited to three separate 1-on-1 interviews between December 2019 and April 2020. The scope and purpose of each interview was as follows:

1. **Interview 1: Sensitisation, Life Exploration and Company Selection** [1 hour, in person]. Participants were sensitised to the research



* Due to COVID-19, two Interview 2's and all Interview 3's were conducted via Zoom

Figure 23: Figure 23: A Journey Map of Each Participant's Study Progression

context using an interactive tour of a poster display on the topics of GDPR rights, potential data-holding organisations, potential types of data and potential uses for GDPR-obtained data. Baseline data was collected on participants' hopes and motivations, their current understanding of personal data, data access, data control, and power as it relates to data. Using a sketch interviewing (Hwang, 2021) technique, participants mapped out their 'data lives' (e.g. Figure 24), annotating key organisations that they have relationships with, types of data those companies might hold, and feelings about such data use and storage by each holder. Each participant selected 3-5 candidate companies to target with GDPR requests.

2. **Interview 2: Privacy Policy Reviewing, Goal Setting and GDPR Request Initiation** [1 hour, in person]. To stimulate reflective thinking and measure impacts, participants were asked to discuss and score their initial feelings of trust and power with each company. Participants then viewed key sections of privacy policies on a screen with the researcher, to identify each company's statements on collection and use of personal data. Participants then initiated an email GDPR request for each company, which had been prepared using a tried-and-tested template generated by [personaldata.io](https://Wiki.personaldata.io) (Wiki.personaldata.io, no date). Interview 2 took place in person, except for P10 & P11 whose interviews took place over Zoom due to the COVID-19 pandemic.
3. **Interview 3: Detailed GDPR Response Review** [2 hours, online video call] Having allowed sufficient time for GDPR requests to conclude (there is a legal duty to reply within 30 days), a deep dive into the specifics of each GDPR experience took place. Participants' personal data was not collected by the researcher, only described verbally by the participant; screen sharing was used to show excerpts to the researcher where the participant wished to do so. Participants were asked a structured set of questions about the completeness and value of any data returned, as well as to provide new evaluations of trust and power, whether their hopes had been met, and any general feelings about the experience. Answers were recorded in a screen-shared spreadsheet, which was also used to structure the discussion (for a sample see [INSERT REF TO APPENDIX]).

Interviews were audio and video recorded, then auto-transcribed using Google Recorder/Zoom, producing a 370,000-word corpus. Transcripts were split up and recombined across participants into six topic areas - digital life, company-specific discussions, general discussions, power, trust, and hopes/goals. These transcript topics were then analysed through reductive coding cycles to produce thematic findings (see 5.4). Quantitative data from interview spreadsheets was also summarised and analysed (see 5.5). Sketches, recordings, screenshots and field notes were referenced throughout thematic analysis to aid interpretation of the transcripts.

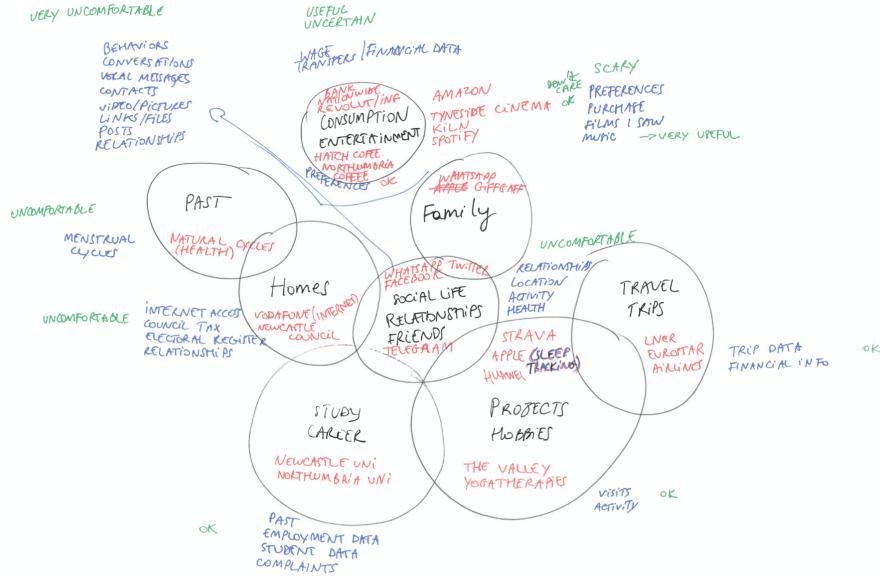


Figure 24: Figure 24: An Example Life Sketch from Interview 1, with Data Handling Companies in Red, Data Types in Blue, and Feelings in Green

5.3 GDPR Request Outcomes

5.3.1 Interview 1: GDPR Target Selection

Initially eight participants chose 5 target companies and three chose 4 to request data from. One participant (P9) withdrew from the study due to COVID-19 after Interview 1. Five participants withdrew a chosen company upon further consideration. Reasons for withdrawing chosen targets included having one's personal data mixed with other household members (Netflix), the account being in someone else's name (Morrisons), not wishing to impact active customer support matters (LNER), and inability to contact the provider by email (ifun.tv, see below). One participant selected Newcastle University, which was vetoed by the research team to avoid conflicts of interest. Hence, 41 out of a possible 52 GDPR subject access requests were made (to 28 distinct data holders) as shown in Table 8:

Table: Table 8. Types of Data Holding Organisation Targeted for GDPR Requests by Study Participants^a

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

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

[TODO replace table with text]

To ensure fairness and consistency, the aim was that all GDPR requests be sent by e-mail by the participant to the data-holder's identified Data Protection Officer, requesting both a subject access request (Information Commissioner's Office, 2021a) and a data portability request (Information Commissioner's Office, 2021b) be initiated, and specifically enumerating and asking for all those datapoints that the company stated in its privacy policy, as well as others which the GDPR entitles individuals to obtain. To identify these datapoints, company privacy policies were analysed and the necessary information was compiled in personaldata.io's semantic wiki ('List of target companies for GDPR requests', no date). This has a feature to generate bespoke GDPR request emails, which were adapted then provided to participants. [INSERT APPENDIX REF]. Facebook, Apple, Huawei and Philips Hue do not offer a contact e-mail address, so the email text (shortened where length restrictions applied) was pasted into a contact form. In one case, entertainment website ifun.tv, the only available means of contact was via WeChat, resulting in the participant (a Chinese citizen) choosing not to contact ifun.tv due to fear of Chinese government surveillance. Through analysis of companies' privacy policies and with reference to GDPR rights, a taxonomy of the types of personal data that could be returned was constructed, using terms from those privacy policies and GDPR legislation: there are five types of personal data, as shown in Table 9.

Table: Table 9. Types of Personal Data Potentially Accessible from Data Holders via GDPR Rights

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

[TODO replace table with text]

5.3.2 Interview 2: Privacy Policy Review and Goal Setting

Participants reviewed and discussed privacy policies for their chosen target companies and were asked to define hopes and expectations for each GDPR request (see Table 12). 74% of goals expresses related to participants wanting to have greater insight and control into their personal data ecosystems; most commonly a desire to see the breadth and depth of data collection by companies, to understand what was being inferred and how personal data was used, and to use such information to better assess trustworthiness of those companies. Such goals were often motivated by curiosity or suspicion, or a desire to shed light on specific incidents or answer specific questions. In some cases participants wanted not just to learn and acquire knowledge but to take control of or delete held data. In contrast, 26% of goals related to gaining personal benefit from one's obtained data: motivators included the desire to reflect on past data to gain self-insight, as well as goals relating creativity, fun, and nostalgia.

At the conclusion of interview 2, participants were provided with the emails and instructions to start their GDPR requests, which progressed as illustrated in Figure 25. Eight requests resulted in no data being obtained, due to either data holder non-responsiveness, inability to access the right account or satisfy ID requirements, or confirmation being received that there was no data to supply. 32 requests (80%) resulted in at least some data being returned; 10 of these directed the participant to use a publically-available download dashboard such as Google Takeout, and the rest resulted in data being made individually available. Of these, one was mailed as printouts, another was mailed on CD-R, and the rest were delivered by e-mail (sometimes involving a secured online website to download). While 22 companies supplied bespoke data packages, 4 did not

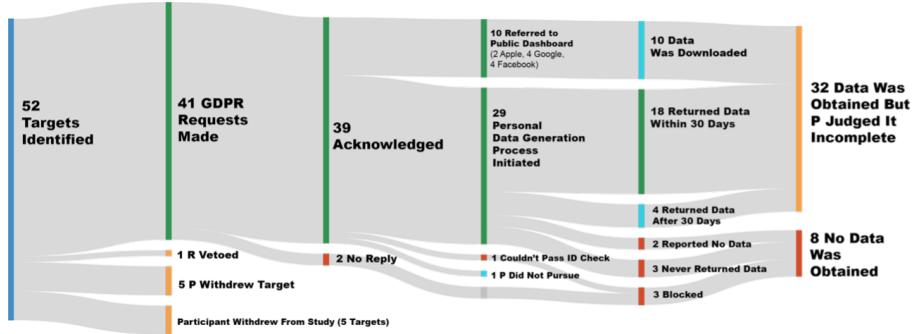


Figure 25: Figure 25: A Sankey diagram giving an overview of the GDPR requests undertaken by participants (P)

return it within the 30 days the legislation specifies (note: requests took place within the context of a global pandemic so response rates may not be typical). Following discussion, participants judged that all 32 requests receiving data had failed to return all requested data (across all five of the categories in Table 9).

5.3.3 Interview 3: Reviewing the GDPR Response

Once each participant's GDPR requests had reached a conclusion point (as described above), they were invited to discuss the GDPR response in detail. Participants were asked to describe (and optionally show) the data they had received, then to evaluate the data holder's response for each data type, according to multiple metrics designed to assess the perceived quality of the GDPR request handling and the subjective value of any returned data. All questions were posed from the perspective of (a) the data that providers said they collect and process in their privacy policies, and (b) the rights that the GDPR specifies, to ensure discovery of missing data or unfulfilled rights would be considered objectively. Participant responses were considered quantitatively (see Table 10) and qualitatively (see section 5.4).

Table: Table 10. Presence and quality assessments of GDPR responses by data type (as percentagesa)

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

^a Percentages represent the proportion of “Yes” answers to each question, per data subtype, from all those where a judgement was given.

^b Participants were asked whether this category of data from each provider would be valuable if they were to receive it.

^c dk = don’t know (percentage of cases where participants felt unable to assess data accuracy or completeness).

^d p = partially (percentage of cases where data was judged partially understandable).

[TODO replace table with text]

Table 10 shows quality assessments for each data type, with rows descending by subjective value. Notably, the kinds of data participants value most (derived, acquired and metadata) were less frequently returned, especially metadata (returned in 4% of cases). Where data was returned in these categories, it suffered from poor data quality, often judged as incomplete, inaccurate, unusable and not useful (although acquired data was largely understandable). At 53%, even the most returned category, volunteered data, was lacking. Where it was returned, accuracy (92%), meaningfulness (72%) and understandability (72%) were high. Observed data was least valued and also rarely returned or complete (yet judged to be of moderate quality). Looking below organisation level at the perspective of individual data categories, data was only judged to be complete in 22% of cases. In 62% of cases, personal data specified in privacy policies to be collected was not returned, despite the legal obligation.

The above quality and coverage datapoints also allowed insights about which service providers were strongest or weakest in each category, and overall, to be drawn. This was done by tallying the “Yes” responses for each category and overall, then dividing by the number of times that provider was selected, to avoid inflating scores for popular companies. The outcome of this analysis is shown in Table 11. The companies that fared worst overall were those that did not return any data at all in response to a GDPR request (Sainsbury’s, Freeprints, Tyne Tunnels, LinkedIn, Huawei, Bumble, LNER). It should be noted that Sainsbury’s and Huawei *did* respond, claiming to hold no data for the requesting participant, though participants found this implausible, which indicates either a problem with compliance, explanation or trust. The other named companies here did not respond at all, despite at least two follow-up emails being sent to them, and despite in some cases having initially acknowledged and promised to satisfy the request.

Companies producing responses with good coverage and good quality included Niantic, Nectar and Sunderland AFC as well as to a lesser extent Natural Cycles, Revolut, Spotify, Tesco and Amazon. Facebook and Google fared well for the breadth of data returned (due in part to their download dashboards), though the quality of Google’s data was found lacking across multiple categories. Last.fm

(owned by CBS) fared poorly overall due to poor category coverage, despite the limited data that it did return being of high quality.

Table: Table 11. Best and Worst Data Holders in Different Categories, According to Participants' Judgements^a

Category / Metric	Best Companies	Worst Companies
Availability of Data / Breadth of Data Returned	Nectar, Niantic, Sunderland AFC, Natural Cycles, <i>Facebook</i> , <i>Google</i> , <i>Spotify</i> , <i>Revolut</i>	Sainsbury's, Freeprints, Tyne Tunnels, LinkedIn, Huawei, Bumble, LNER, Nexus, <i>Three</i> , <i>Philips Hue (Signify)</i> , <i>Check My File</i>
Completeness of Returned Data	Niantic, <i>Nectar</i> , <i>Sunderland AFC</i>	Sainsbury's, Tyne Tunnels, Freeprints, Nexus, LinkedIn, Huawei, Revolut, Bumble, LNER, last.fm (CBS), <i>Google</i> , <i>Tesco</i>
Accuracy of Returned Data	Sunderland AFC, Niantic, <i>Tesco</i> , <i>Nectar</i> , <i>Amazon</i> , <i>Natural Cycles</i>	Direct Line, last.fm (CBS), <i>Google</i>
Understandability of Returned Data	Nectar, Spotify, Sunderland AFC, Niantic, Apple, <i>last.fm (CBS)</i>	AirBNB, Virgin Media, <i>Google</i> , <i>Instagram</i> , <i>Tesco</i>
Meaningfulness of Returned Data	Niantic, <i>Spotify</i> , <i>Sunderland AFC</i> , <i>Natural Cycles</i> , <i>last.fm (CBS)</i>	AirBNB, Credit Karma, Philips Hue (Signify), Direct Line
Usability of Returned Data	Amazon, last.fm (CBS), <i>Facebook</i>	AirBNB, Credit Karma, Virgin Media, Sunderland AFC, Huawei, <i>Three</i> , <i>Google</i>
Usefulness of Returned Data	Amazon, Facebook, Virgin Media, Spotify, Revolut, Niantic, last.fm (CBS)	AirBNB, Credit Karma, Nectar, Direct Line, <i>Three</i> , <i>Google</i>
OVERALL^b	Niantic, <i>Sunderland AFC</i>, <i>Facebook</i>, <i>Spotify</i>	Sainsbury's, Freeprints, Tyne Tunnels, LinkedIn, Huawei, Bumble, LNER, last.fm (CBS), <i>Philips Hue (Signify)</i>, <i>Nexus</i>

^a Companies were ranked according to total number of all responses in that category for this company that were "Yes".

^b Company names in normal text are best/worst; names in italics are second best/second worst.

[TODO replace table with text]

At the conclusion of the final interview, participants were reminded of the specific hopes and anticipated data uses they had expressed at the start of their journey and asked about how well each goal had been met. These answers were recorded and combined to produce percentage values showing in how many cases goals were fully met, partially met, or not met at all, as shown in Table 12.

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

Table: Table 12. Participants' hopes, imagined data uses and goals for GDPR, as well as resultant outcomes

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

[TODO replace table with text]

5.3.4 Perceived Power and Trust

Repeating scoring questions were used to examine how participants' feelings towards the data holders changed throughout the process: Participants were asked to assess trust from 0 (total distrust) to 10 (total trust), and to assess their perceived power on a scale of -5 (total provider power) through 0 (balanced power) to +5 (total individual power). Explanations and reasoning for initial ratings and for any changes were uncovered through questioning. By repeating the same question at different times, longitudinal comparisons could be made. Many participants' attitudes did change as a result of the experience (as summarized in Figure 26), for both perceived power (45% of cases) and trust (66% of cases). For those with changed attitudes, the change was often negative: in 63% of cases where participants perceived a change in individual power, that change

was a loss in individual power, and in the majority (52%) of cases, participants felt more distrustful of GDPR targeted companies after completing the process (constituting 79% of cases where a change in trust was perceived). However, it is important to note that in some cases GDPR had a positive impact; in 17% of cases participants felt their perceived power had increased, and in 14% of cases participants felt more trusting of providers after GDPR.



Figure 26: Distribution of Net Changes in Participant's Perceived Power and Trust Scores over the Study's Duration

Looking deeper into these datapoints, changes in attitude could be attributed both to the impact of reviewing the privacy policy as well as to the experience of the GDPR process and the discursive review of GDPR responses. Figures 27 and 28 show snapshots of power and trust ratings at different points in the process which illustrate these impacts. Looking to explain these changes qualitatively, it was found that privacy policies often contradicted participants' expectations, resulting in discomfort. In two cases (Philips Hue and last.fm) privacy policy review revealed that the service relationship was with a completely different company than the participant thought, which was disturbing to them. LinkedIn's privacy policy was noteworthy as being exceptionally clear, reassuring and trust-enhancing to the participant, largely due to its 'easy read' text sidebars but also good use of examples. However it does appear that simplifying privacy policies can go too far: Google's privacy hub (which includes video explainers) was considered easy to understand but necessarily broad (given their breadth of services) and thus over-simplified, raising uncertainty about generalisations made, and in some cases increasing distrust.

Considering the process as a whole, participants' attitudes were impacted particularly by the "*hassle*" (P11) they experienced in getting through the data

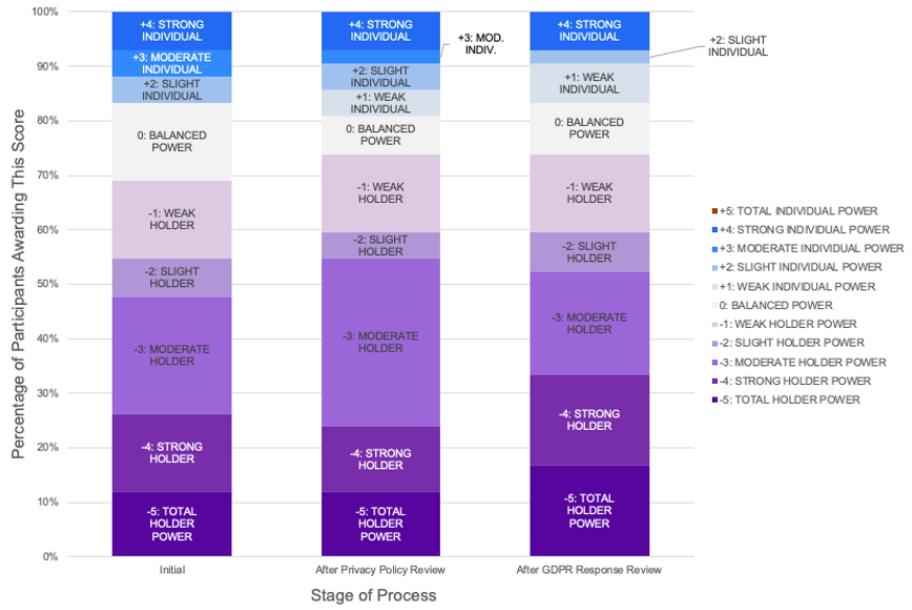


Figure 27: Figure 27: Perceived Power Balance Between Individual and Data Holder at Different Stages of the GDPR/Study Process

access process, and from the realization that what seemed at first glance to be a thorough response, when scrutinised more closely in Interview 3 and viewed through the lens of the privacy policy promises and one's GDPR rights to the five categories of data, was in fact quite poor.

5.4 Thematic Findings

As described in 5.2, the topic-focused transcripts were carefully examined to identify themes and findings, a process involving over 200 person-hours of iterative data analysis (Huberman and Miles, 2002) of the interview transcripts. In this section the findings from that analysis are presented and summarises, with the three main themes being introduced in 5.4.1, then each theme is further detailed in sections 5.4.2 to 5.4.4, including participant quotes.

5.4.1 Themes & Subthemes

The findings are presented through three main topics: Insufficient Transparency, Confusing and Unusable Data, and Fragile Relationships.

Theme 1 (see 5.4.2 and summary in Table 12 below) describes the **Insufficient Transparency** that participants experienced in privacy policies and through the GDPR process; Organisations appear evasive over data when responding to GDPR, leaving people "in the dark" even after making GDPR requests.

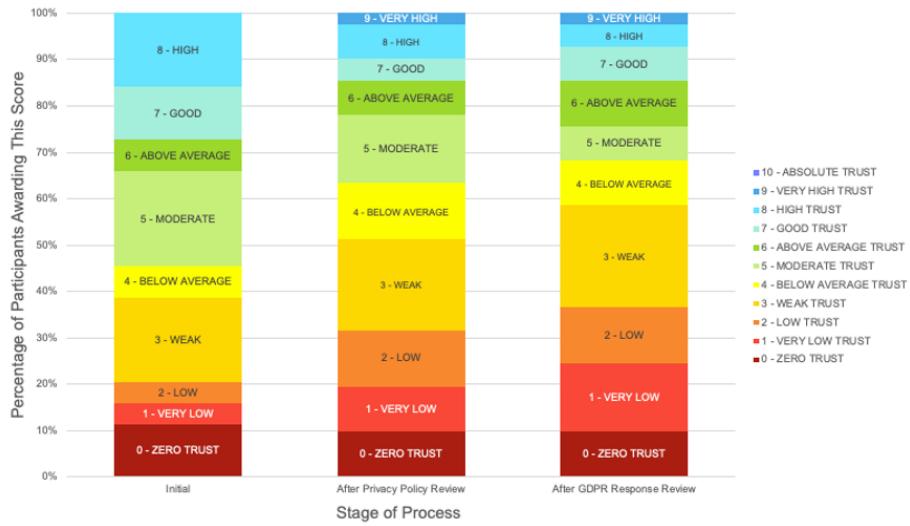


Figure 28: Participants' Perceived Trust in Provider at Different Stages of the GDPR/Study Process

Table 8: Table 12. **Theme 1 - Insufficient Transparency.**
Subthemes & Participant Quotes.

Subtheme	Description	Quote
A Desire for Awareness and Understanding	Participants want to see, know and understand the data held about them. There was particular interest to see data collected or inferred about them without their involvement, and to understand how data is used and shared and how that might affect them.	<i>“[Companies have more power] because they’re making decisions about things and you don’t know how they’re making those decisions.”</i> (P5)
Non-Compliance Without Consequence	Many providers failed to provide data on time or at all. In 100% of cases, returned data was incomplete, and many viewed this as non-compliance. Data holders’ freedom to disobey legislation was attributed to a lack of enforcement and seen as an exertion of power.	<i>“I am surprised at Google’s unwillingness to provide me with all of the data . . . they haven’t provided me with all of my data. And that’s not legal.”</i> (P7)

Subtheme	Description	Quote
Inadequate Data Responses	Participants judged data holders to be unhelpful, GDPR procedures to be painful and ineffective, and returned data to be lacking in coverage and in quality. Their questions remained unanswered; after GDPR they were still “ <i>in the dark</i> ” (P4). There was widespread disappointment and a view that GDPR did not confer any power to the individual.	<i>“It’s kind of disappointing because I would have hoped that this process would have levelled the user power versus the organisation power in a way that holds them accountable and [it doesn’t] seem to be doing that.”</i> (P1)

Theme 2 (see 5.4.3 and summary in Table 13 below) explains how participants received **Confusing & Unusable Data** from data holders; When presented with their data, people struggle to understand it and relate it to their lives and are not able to make use of it.

Table 9: Table 13. **Theme 2 - Confusing & Unusable Data.**
Subthemes & Participant Quotes.

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

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

Finally Theme 3 (see 5.4.4 and summary in Table 14 below) examines the **Fragile Relationships** that individuals have with data holders; Companies' data practices, and in particular their privacy policies and GDPR response handling, can be impactful to customer relationships, carrying a risk of damaging trust but also the potential to improve relations. These three themes are

Table 10: Table 14. **Theme 3 - Fragile Relationships.** Sub-themes & Participant Quotes.

Subtheme	Description	Quote
Power and Enforced Trust Through Data Holding	<p>Participants feel that the sacrifice of (or the giving of permission to collect) personal data is a necessary cost in order to get the valued benefits of the services they want to use, something they are pressured to do and have no choice about. Such sacrifice is seen as the giving up of power, as participants lack access and control to that data. In the face of providers making decisions based on data and processes that they could not observe, participants felt powerless. This amassing of data was sometimes seen as surveillance, and some saw great potential for misuse and abuse of it.</p>	<p><i>"For me to have power over my data, I think is a fair and normal thing. But for a company to have power over [my] data means that it's basically a proxy to have power over me."</i> (P8)</p>
Perceptions of Data Holders	<p>Participants entered the study with varying perceptions of providers' integrity, influenced by reputation, business model and size. Participants' various observations reveal a strong link between their perceptions of providers' data handling practices and the trust they hold in those same providers.</p>	<p><i>"When I like the company already, I'm more willing to give them my data."</i> (P2)</p>

Subtheme	Description	Quote
Changed Perspectives Through Scrutiny	In general, the more that participants found out about data-centric practices through the process of scrutinizing privacy policies and making data access requests, the more they distrusted providers. Failure to explain or provide complete data was harmful to trust. Conversely, where providers were more transparent or participants did obtain interesting data insights, trust was increased.	<i>"If someone's not completely open with you, then you're like, 'What are you hiding?', which means you trust them less."</i> (P4)

5.4.2 Theme 1: Many Companies are Evasive and People are “Still in the Dark”

5.4.2.1 A Desire for Awareness and Understanding As Table 12 shows, in the vast majority (62%) of cases, participants wanted to see, know and understand what data was held about them and how it was used. For example, P11 wanted to know what data was collected by train company LNER when he bought tickets, so that he might judge whether it was appropriate:

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

Beyond the data that participants had directly volunteered (see Table 9), most data was currently unknown to participants. In particular they wanted to gain awareness of what data might have been collected without their knowledge.

“The bit that concerns me is where I don’t know what data is being taken by companies. If I’m registering for a library or something, I know [what] data I’m giving to them, but what I don’t know is all the other stuff that they’re recording”—P9

Participants were equally unaware of what holders might infer from the data they had collected. P4 wondered if Philips could use data from his smart home lighting to deduce his sleep and TV-watching routines. P7 had received targeted advertisements relating to pregnancy that she felt weird about because she did not understand why she had been targeted in this way. P5 raised concern about how data inferences could affect decision-making, surmising that the data holder had greater power than him because *“they’re making decisions about things and you don’t know how they’re making those decisions”*. Sharing of personal data is also insufficiently visible to participants; two participants (P3,P4) targeted

GDPR requests to credit-check websites (Credit Karma, CheckMyFile) - P4 wanted to get “*a picture of what other companies can currently expose*”.

5.4.2.2 Non-Compliance without Consequence As detailed in 5.3.2, few requests resulted in a timely provision of requested data (44% or 68% depending whether referral to a download portal is excluded or included in the count). Many data holders responded late or not at all; such actions are objectively a breach of legislation. However, participants were broadly unsatisfied even when they did receive a GDPR response. In 100% of cases where data was obtained, it was considered incomplete, and this was usually seen as further failure to comply. Participants had reviewed their GDPR rights in Interview 1 (though, as expected (Rughiniş *et al.*, 2021), most were already aware), and so several participants saw this apparent non-compliance relative to their understanding of their rights as a poor quality of response, for example:

“I feel more concerned now, [...] what they’ve given me seemed reasonable. But then comparing against what we asked them for, what I’m legally [entitled to], it’s a fraction.”—P5

For some participants, sceptical from the start, such poor responses were consistent with their expectations; P6 found the incompleteness of Facebook’s response “*alarmingly unsurprising*”. Others had expected compliance:

“I am surprised at Google’s unwillingness to provide me with all of the data... they haven’t provided me with all of my data. And that’s not legal.”—P7

Many participants, reflecting on a feeling of having less power than they had initially thought, felt that the prevalence of non-compliance showed that too much power relative to the authorities, that a lack of pressure is being applied by regulators and that “*there needs to be more enforcement*” (P11). P6 revised his view of Facebook’s power versus his own because he felt that after review he now could clearly see “*which [data] they are prepared to share and which they aren’t*”. P11 also framed the selectivity of responses as an exertion of power:

“It seems like there’s a lot of derived data about things like purchases and stuff [that I would expect] that just isn’t there. So they’re free to not give me the data. That, to me, suggests [that despite GDPR] they retain an awful lot of power.”—P11

5.4.2.3 Inadequate Data Responses While in some 22% of cases participants did meet their goals through GDPR (see Table 12), when it came to the desire for greater awareness and understanding discussed in 5.4.2.1, this want was largely unmet. Only volunteered data such as basic personal information or user-generated content was usually returned complete; this was often viewed as mundane and uninteresting, and the focus on these data types in returns was viewed as evasiveness. Facebook, P6 observed, “*give you that kind of descriptive boring data which is mainly all publicly available anyway*” and had omitted “*the stuff that I would consider valuable to them*”.

In general, the data responses did not provide the answers participants sought. Many reported “*still*” not knowing what they wanted to find out. P4 said he remained “*in the dark*” (P4). P7 stated that “*even though I did the process correctly, I still didn’t get that much back*”. Concerns held by participants from the outset remained unaddressed, as in P11’s case:

“I still am quite concerned about how much data organisations have, particularly how they link that other data and how data is bought and sold, and I haven’t really got any answers on that.”

It was not just the data returned, but the process itself, that participants were dissatisfied with; requesting and achieving data access was time-consuming and difficult. “*Jumping through hoops*” was a phrase used independently by four different participants (P4, P5, P7 & P11) to describe the experience. Some found data holders obstructive and unhelpful:

“I feel like they give you a response that [makes it so] you cannot proceed intentionally”—P10

Participants recognised that they had received help and coaching, and that the processes were so tedious that without that, they may not have persisted. P1 suggested that without the provided template, it would be “*a lot harder to get meaningful data out*”, and P7 attributed her sole successful request to the guidance she had received in progressing it. P5, having experienced problems with expiring links, delayed responses and missed emails, had been surprised at “*how difficult it was just to get my data, and the fact that I had to ask them about six different times*”.

Not all requests were this painful, some were handled smoothly. As P11 put it, “*Some companies make it dead easy to get, but then the data is not massively useful. [...] Other companies make it a pain in the neck to get it.*” Overall the view of GDPR data access was one of disappointment. Participants found GDPR ineffective: P10 said “*Frankly, [GDPR] doesn’t have as much influence as I expected*” and P1 commented that:

“It’s kind of disappointing, because I would have hoped that this process would have levelled the user power versus the organisation power in a way that holds them accountable and [it doesn’t] seem to be doing that.”

5.4.3 Theme 2: People Struggle to Understand, Use and Control Their Data

5.4.3.1 The Search for Personal Value in Data Prior to receiving data, participants had anticipated discovering insights about their own lives by browsing and reflecting on their personal data, consistent with personal informatics literature (Li, Forlizzi and Dey, 2010). However, there was a comprehension gap between the useful information they imagined and the actual data returned; data was typically delivered as a bundle of technical files, which were hard to understand and often delivered without explanation. Some felt (in line with concepts

of *effective access* described in 2.1.4) that they lacked the necessary skills or tools to make the data understandable or usable “*for a non-techie person*” (P11). When the researcher guided P7 to jsonlint.com, an online formatter, she found her JSON-formatted data more understandable. P2 made the point that data holders must be using tools themselves to make sense of people’s data: “*They’re not just looking at a JSON file, so I would like to have the same visualisation [as them].*”

There was a sense that by sending people individual data files, data had been removed from the environment in which it has meaning, and that the returned data excluded necessary context for interpretation. This was often manifested in the form of internal codes and abbreviations that individuals could not understand. P4 stated of his experience looking at smart-lightbulb data from Philips Hue, that there was “*just so much of it that it’s impossible to know [what it all means]... You’d have to spend a few hours going through this and being like, ‘OK, what does that line mean, and that symbol, and that code?’*”. This lack of context also materialised as a failure to explain decision-making processes: P5 reflected, when looking at driving scores from a car insurer that uses a mobile app to monitor her driving, “*I could see the data; it was the score that was weird for me. Like, it doesn’t tell you how it’s calculated.*” P1 noticed that although some companies did make some effort to explain the returned data, this varied substantially across providers. He said that “*it would be nice if these companies had a standardised model of how this information is presented to people, so it [could] be easily understood*”.

One of the greatest obstacles to understanding that participants faced was being faced with a large volume of information and no means to quickly digest or navigate it; either very large files, or complex hierarchies of nested directories containing many separate files. It is clear that there is a need for *summaries* so that participants can quickly get a handle on what is - or is not - present. Returned data “*could be valuable if you knew what the hell [was] in there*” (P4). P1 described one of his data responses as “*almost too much [...] for a normal person to be able to process and understand.*” He said that it “*could do with a document detailing, like, ‘this is what is in here’*”, and described the disparity across responses as being “*either like death by thirst or death by drowning [...] It would be better to drown, but still not ideal*”. Ultimately is is clear that in general, returned data was not presented in a way that is optimised for understanding.

Another question that our findings were able to shed some light upon, in service of RQ1, was to consider what precisely makes data valuable to individuals. This is especially important given participants did identify the potential to gain personal benefits from their data (as seen in the second set of goals in Table 12). An idea that came up again and again was that data is most valuable when it *spans a period of time* and can be related to events in the individual’s life over that period. This could potentially provide new insights to participants.

P2 for example hoped to see, or be able to construct, breakdowns and charts that would help him examine his food shopping habits. Through the GDPR

process, P10 accessed details of her spending on micro-transactions in the mobile game Pokémon Go that had not been available to her through the app. P11 wanted to derive insights about his train travel by examining the geography, cost, journey length and patterns of his past journeys through data he hoped to receive from LNER. Long-duration data offers the potential ability to identify trends and changes in one's own behaviour over time.

It was these historical parts of their data that participants found most meaningful, offering as it does a means of remembering, with data potentially serving as a “*window into your past*” (P11). P5 saw value in perusing music-listening data “*just because it’s cool to look back on stuff that you’ve done and you don’t necessarily distinctly remember it*”. Generally the longer period the data covered, the more valuable it was deemed to be:

“I would actually be interested in last.fm, partly because the data goes back to 2008 . . . Spotify only goes back about four or five years and not everything I listen to is on Spotify.”—P11

P6 saw the data accumulated by service providers as potentially forming part of a valuable background context to understanding life events in his past: “*I would like to [...] build a picture, not just like, ‘I remember going to Reykjavik’, but if there’s other data around that time [I could] sort of paint a biography of myself*” and described some of his data as “*a kind of personal history that has been quantified and sort of datafied*”.

5.4.3.2 Unusable Data Formats This personal value that captured data has the potential to offer shows that it is all the more important that participants be able to understand and make use of their data. Our participants found that the format in which data was returned often meant that it was not only difficult to understand, but difficult to use as well. Using data meant different things to different participants, with imagined uses including budgeting, record-keeping/archiving, or using the data for creative or fun purposes. Some participants (e.g. P5) saw value in potentially combining data from multiple sources, though this did not turn out to be practical. Participants did not know what data to expect, and generally imagined returned data being more useful than it turned out to be:

“I think . . . you could do some interesting mashups, but I don’t really know what with until I’ve got the data. It depends on the data; I’m sure there could be some cool uses of it.”—P4

Once data was received, participants struggled to interpret and understand it to a sufficient extent to be able to identify the useful data or meaningful information they had hoped for. Returned data formats and response structure were extremely varied. Some reported that there was not sufficient machine-readable data to make use of the data. For example, P4 received a Microsoft Word document full of pasted screenshots from an internal portal as part of his response from his ISP Virgin Media, and said that its usefulness “*depends on*

what you want to get out of it, really. If you want to view the data they have about you, it's quite usable. If you want to do something automated[analytical], then it's not. P11 found a similar returned screenshot from an internal system to be “completely non-understandable”. In other cases, the opposite problem occurred, with data being too technical for the participant to use. P10 said of JSON data: “*For normal people who don't understand programming, I feel it's just, there's no use at all.*” P7 felt she lacked the technical proficiency to make use of the returned data:

“They have provided it in formats where I can see that, if I were a developer, I could do things with it, [...] but if I was not that sort of person, it might be quite difficult to understand”—P7

In P5’s case, she saw the potential to use the data but felt that what was missing was additional explanation or guidance on how to interpret it:

“They did give me the data, but not how it fitted together. It's like being given the bricks to a house, and then they're like ‘Here's your house’. It doesn't really mean anything when it's just bricks, if you don't know how to put it together.”—P5

P11 highlighted a problem with his Tesco shopping data that was not just a matter of formatting or skill, but the granularity or focus of the data itself:

“As a technical person, having a CSV of data is quite useful, potentially, but actually what can I do with that if it's Tesco's internal systems data?”

While on the face of it the findings of 5.4.3.1 and 5.4.3.2, and the conflicting demands for both more technical and less technical data might seem contradictory, what we can infer is that participants collectively need *both* usable technical data and easy-to-read information summaries - and that those summaries should cover both the relatable life information encoded within the data and the information about the data, what it means and how to use it; this idea is explored further in (Bowyer, 2021).

5.4.3.3 The Liability of Data You Can't Delete or Control Having recognised that potential value of data relating to their lives, before or during this research, several participants were concerned about personal data being held. P10 for example said with reference to dating site Bumble: “*Since I found my partner [and therefore no longer need a dating site] I deleted my account and I've been wondering, ‘Are they still keeping my data at the back?’*” and with reference to both Instagram and Bumble, expressed a desire to have her data deleted and expected GDPR to play a role in the enforcement or verification of that deletion, something she could not otherwise be sure of. P8 considered the holding of sensitive data to be a liability that she was only willing to tolerate while she was actively using a service, and this was part of her motivation for targeting Natural Cycles:

“I now use a different one, but I used, for about a year [their] app to track my menstrual cycle. [It was my] main contraception method, so that's things that

this company probably has. Now that I'm not using it any more, I don't know if they delete the things or not

Many participants expressed a desire that data be held only for a short time, and questioned the default practice of data being kept beyond the period where it was needed to deliver a service:

"The thing that concerns me is that I haven't used Tesco online for at least four or five years, so why are they hanging on to my IP address from five years ago?"—P11

He went on to spell out the liability he saw in such apparently mundane data being held, the liability coming from the duration of the data: “*10 years of worth of shopping records... how much would that be worth to a health insurance company, and would [Tesco] succumb to the temptation to sell that on?*” P10, a Chinese citizen, identified long-term sources of personal data as an enabler for future privacy violations, saying that “*in China, [there is a trend] that as soon as someone becomes famous, people begin digging [through] all their past experiences*”.

Most participants described the ability to delete or enforce the deletion of their data as having control over it, and given the current practical lack of such a capability felt that they had insufficient control over data holding. One of the first steps participants identified in gaining control of their data was simply an ability to see it, for accountability, so that they might check the accuracy, security and breadth of collected data and flag any unforeseen concerns. They felt that a deeper understanding might lead to an increased sense of individual safety and data control and facilitate them to make changes in data habits or choice of service provider:

"I want to understand how much they're keeping. And what they're doing with it. I'm hoping that by knowing that, I might change my behaviour about all the data I accidentally create."—P7

In this participant's case, this hope was unsatisfied, and upon looking back at her experience she remarked:

"I guess that's one of my criticisms of GDPR in general - that although I can understand what data a company holds about me, there's no obligation for them to tell me what they are doing with it.. And sometimes I think my willingness to give a company data might be quite intrinsically linked with what they're gonna do with it."—P7

In fact, that legal right does exist through GDPR, but as we can see it was not delivered in practice. What participants want is to feel aware and in control of their data; this must begin with better data legibility and explanations of data use, accompanied by clear pathways to enable data correction or deletion.

5.4.4 Theme 3: Poor GDPR Handling Can Damage the Forced & Fragile Trust Relationship

5.4.4.1 Data Holders Enforce an Uneasy Trust The lack of visibility and control over personal data discussed in the previous section, combined with a sense of being in the dark (see 5.1) about data practices, caused participants discomfort before, during and after the GDPR process. This stemmed from a sense of finding themselves facing uncertain risks they feel powerless to change. Many participants, discussing their relationships with providers, expressed a range of emotions ranging from curiosity to anxiety and distrust:

“I’m curious... I wonder what they’ve got on me. [...] If it’s anything other than the barest minimum that is necessary for them to do their job [...] then I get creeped out by that.”—P11

Participants felt most uneasy about the amount of “*intimate*” (P1,P2) data that providers collect. P1 was uncomfortable about Facebook having information about his social circles. P2 said he felt “*quite vulnerable*” that his Google search terms “*say pretty much everything you have done... the most intimate things you were thinking about*”. P11 singled out ISPs as having the potential to track everything their customers look at online, noting that “*I don’t think you’ve got much choice about that.*”

Participants also felt that some data holders held so much data that it had begun to resemble surveillance, such as in the case of P1, who used “*an absurd amount of [Google’s] services*” and reflected that “*if I’m driving somewhere, I’ve got Google Maps open, so they know exactly where I’m going, they know how fast I’m going, they know what I’m listening to while I’m driving...*”. Participants saw the potential for abuse, fearing this kind of deeply personal knowledge could be “*used against*” them (P2). P11 felt that Apple had enough data to “*screw me over*”, and P5 considered that her car insurer Direct Line uses data to “*Judge*” her, noting that “*it’s not like I can contest the data and say ‘Actually, no, I disagree.’*” In a more extreme illustration, P10 shared her fears that data collected by WeChat and Weibu (Chinese services similar to Facebook Messenger and Twitter respectively) would be at risk of abuse against citizens by the Chinese government. In some cases participants were able to identify concrete instances where providers had exploited the personal knowledge they held: in P6’s view, Facebook use their knowledge of their users’ friendships and relationships to “*hook your attention*” and prevent users deactivating accounts in a “*disingenuous*” manner.

Whether or not data is used nefariously against individuals, thinking about the potential for this caused participants to associate the mass collection of personal data as an acquisition of power over them: “[*Companies that] know a lot about everyone will inherently be able to have power either through persuasion or manipulation*” (P1). P7 saw the **holding of data** as the source of holders’ power: “*when I think about other people having my data [...] the control isn’t sitting with me.*” Others identified the ability of data holders to **deny or limit**

access to data as their key source of power:

“If you’re not getting what you perceive to be yours back in completion [sic] then you’re not in control of your own data and you have fairly little power over it.”—P1

The view of data holders having more power in the service relationship (mirroring reports described in 2.1.2) was reflected in participants’ evaluations of power balance: in 69% of relationships participants felt that the data holder had more power than them (rising to 74% after GDPR), whereas in only 17% of cases (unchanged by GDPR) did participants feel they themselves had more power.

Several participants equated power over a person’s data with power over the individual. When asked to define power in the context of data, P8’s unprompted comments aligned with my Case Study One findings describing power over data as a proxy for individual participation or inclusion in decision-making [Bowyer et al. (2018); 4.2.2; 4.4.1]: *“For me to have power over my data, I think is a fair and normal thing. But for a company to have power over [my] data means that it’s basically a proxy to have power over me.”*

A key dynamic to understand the value exchange within these relationships is that individuals sacrifice their data in exchange for value - that value being the capabilities offered by the services. All 11 participants expressed the idea that the sacrifice of data is something that they have grown to tolerate in exchange for some benefit. P6 tolerates data collection by travel agents because “*they might help me pick a better deal next year.*” P11 said he was happy for Tesco to collect data in order to “*profile me to try to sell me more cheese, fine, whatever,*” though expressed caution that he doesn’t “*know what else they’re doing with it,*” and more generally was “*deeply concerned*” about unseen data trading. The benefit can be convenience too; P10 had logged into Pokémon Go with her Facebook account, knowing that implied data collection by Facebook, “*because it’s much easier*”.

Participants often felt this sacrifice was something they had no choice about, but they did not like it. Unease over the trade-off being made surfaced most often in the context of recommendations; generally, participants valued data-derived suggestions provided they were “*relevant*” (P1, P8) and not too “*intrusive*” (P1, P6). It is clear that data sacrifice is only tolerable up to certain limits: P10 said of Niantic found the sacrifice acceptable provided that “*they don’t sell where I live or my daily routine*”; however while Niantic’s privacy policy promises data is not sold, it does appear that some level of personal location information *is* accessible in some form to third-party advertisers (Varghes, 2019). P8 said that relevant music recommendations were “*very useful*” but found Amazon shopping recommendations “*very scary*” because “*I don’t want to see that I’m predictable*” and felt that “*if someone out there knows [what I want] before you [it’s] like taking agency away from me.*”

Permission to collect and use data is knowingly provided by individuals to data holders, but the mechanisms to do so are considered inadequate: P2 felt that

permission giving options are “*not granular enough*”, and in P11’s view “*it’s not a negotiation at all, it’s all or nothing.*” Worse, some participants feel that permission is coerced from them: P10 observed that Niantic “*pressure you into*” giving continuous access to your location data by tying it to the availability of in-game benefits such that “*you don’t want to lose out*”.

Such lack of choice or coercion led to feelings among participants of resignation about data collection, seeing it as a Hobson’s choice:

“I feel like it’s inevitable that if you want to access their services at all, in any normal kind of way, that you automatically have to give them your data.”—P7

Ultimately, participants felt that their data was “*revealing*” (P2,P3,P11) a lot of information about them, and so their only real option to maintain their privacy was to prevent data collection in the first place by not using that service at all (P1,P2,P3,P7,P10,P11), and living with the subsequent lack of service capability.

5.4.4.2 Perceptions of Data Holders The discussion of attitudes towards data holders through this study allows some insights to be drawn as to how data holding service organisations are perceived, particularly with regard to data handling:

Factors such as reputation, size and business model were often a major contributor to participants’ impressions of companies. For example, P2 described feeling “*more at ease*” with Apple, due to their hardware-oriented business model, than with Google, who “*make money through data*”; in general, where there was a lack of clarity around how a company makes money, or that model was clearly based exploiting sacrificed personal data, there was a greater suspicion, while trust was higher in those companies that offered a paid service:

“One of the main things was there [are] no ads. [Natural Cycles is] a paid service, so there’s no, like, ‘you don’t have to pay but we use your data to make money’.”—P8

Returning to Apple, P2 also noted that they “*position themselves as a defender of privacy rights*” and along with P11 (another participant who had targetted Apple) held a more favourable view as a result. P10, on the other hand, had been influenced by a documentary she had seen, becoming suspicious of Apple’s control over her hardware.

While attitudes to Apple were generally positive, Facebook—which has, and continues to be, the subject of much negative media attention over its apparently cavalier attitudes towards personal data—was held in much lower regard. P6 said Facebook had “*in every shape or form, shown themselves not to be trusted*”, an opinion formed from “*high profile news stories where they have done unscrupulous things and are very willing to just hand over data*”. P9 reported feeling “*slightly dubious*” about Amazon as a result of “[press coverage] about their ethics that may or may not be true, and just the size of them... and just the level[amount] of data, as well”. Clearly expectations around data handling are a strong influence

on attitudes toward service providers, though sometimes other factors play a role, such as with P8, who was comforted not just by Natural Cycles' payment model, but the values they project: “*This is woman-empowerment-orientated [sic], so in that sense I think I do put my trust there as well.*”

As well as these more impression-based influences, it also became clear that participant’s direct experience of interacting with a company affects their feelings toward that provider. P1 found that “*in the same way that Amazon is quite janky [unreliable and awkward to use], Google feels fairly polished and so I trust them more.*”. As well as customer/user experience, a perception of receiving a valuable service creates trust: P4 said of Google that “*the amount I trust them is in line with the utility I get from them*”. In the context of data sacrifice, high levels of trust do have an effect on customer behaviour:

“*When I like the company already, I’m more willing to give them my data*”—P2

5.4.4.3 Changed Perspectives Through Scrutiny and Transparency

Longitudinal examination of participant’s feelings of perceived individual trust and power across their GDPR experience allows the impact of the experience to be analysed. As illustrated in Figure 5 and 6, trust in data holders tended to diminish through the data request process. Some of this distrust arose from examination of privacy policies, for example in P5’s case who commented that Spotify “*shouldn’t need to know that much about me, they should just give me music*”. In most cases the most noticeable decline in trust occurred between Interview 2 and 3 (when the participant experienced the GDPR process) or within Interview 3 (where the returned data was examined), showing that both the quality and coverage of the data return, as well as the execution of the data request process, often have a detrimental effect on trust. Individuals’ perceived power, however, did not undergo a corresponding change:

“*They’ve not given me everything back that I thought they’d be collecting, which makes me trust them less. So power-wise, I don’t think [anything]’s changed, but trust, I think it has.*”—P1

The absence or sparsity of derived and acquired data and of metadata around sharing and handling (see 5.3) noticeably damaged trust. P1 directly attributed his reduced trust scores to what he saw as non-compliance (5.4.2.2) through failure to return all data categories. P5 lowered her evaluation of trust in Spotify further still upon completion of Interview 3 “*because they didn’t say anything about what they’re doing with my data or where it’s going*”. P8’s trust score Natural Cycles was similarly reduced “*because I think it’s hard to get any sensitive data, and it’s not really made clear what they’re using it for*”.

It is also clear that poor GDPR handling in itself can damage trust, independently of the data returned. P2 reduced his trust score for AirBNB “*because of the way they’ve handled [the data request], and the way they’ve made it hard for me to read the data*”. Similarly, P7 downgraded her score for LinkedIn “*because I*

feel like they have my data and [they've] not bothered to find my data, and that makes me feel like I shouldn't trust them quite as much".

Participants want greater transparency than the current processes are currently providing, and the failure to do so is a direct cause of distrust:

"If someone's not completely open with you, then you're like, well 'What are you hiding?', which means you trust them less."–P4

Despite the impacts on trust, both using GDPR access rights and the wider process of scrutiny and discussion surrounding that process within this study had a positive impact on participants' awareness, offering "*insights into how big companies are actually handling these requests*" (P7) and how to practically use one's data rights, showing that transparency (even in the hampered form of current GDPR handling) has an education benefit. Participants had initially expressed wishes to gain insight into data handling practices in order to increase accountability and inform their decision-making on provider loyalty and privacy settings. GDPR offered the potential to compare data expectations with reality—for example P11 was initially "*curious to find out if [Apple's] marketing claims match their reality around privacy*". While such broad goals were generally unmet, several participants found the process thought-provoking and reported feeling more aware about what data they were enabling their providers to gather. P4 felt the process "*got me thinking about, like what other things could I try, and what other sources of personal data are there*". P8 reflected that "*it's a skill and a kind of knowledge that I think everyone should [have]. I don't think it [should be] normal that I felt so clueless*". Some commented on the value of understanding GDPR itself through the experience:

"[I] think the exercise was useful in that I understand what a GDPR request can do and what it cannot do. And there's a lot it cannot do. And I think it might seem that it gives you a lot of power, but really, it doesn't."–P2

While considering the negative impacts of the GDPR experience on trust some realised the potential trust-engendering impact that a more transparent response could have brought:

"I think the lack of transparency in a lot of these processes has not helped, you know, if Tesco had [...] plain English processes for getting the data and you've got the data in a plain English way, that would do a lot to bolster trust."–P11

In a small number of cases, this was witness in practice, with a good GDPR response actually increasing participants' trust in certain providers. For example, P5 reflected that she may have been "*a little harsh*" in her initial judgement of Instagram and said she "*actually really liked what they sent... in comparison to the three others, I was genuinely, I opened Instagram's one and I was like 'this is really cool.'*". P10 was very impressed with the response from Niantic and after GDPR she trusted them very highly "*because they replied really fast, the data provided is very detailed, and their attitude towards this whole issue is very positive,*" concluding that they are "*a really nice company*" and even

indicating an increased willingness to spend money on their product. P6 trusted Sunderland AFC because “*they were really kind of upfront and … I got the data from them first, [...] no messing about, the format they gave me just made sense*”.

In these comments, we can see an indication that, although the data requests often did not live up to the hopes of the participants, positively engaging with the process was influential and did affect participants’ outlook. In particular, close attention was paid to the willingness of companies to be transparent and forthcoming, with GDPR representing an opportunity to test organisations on their data practices and assess their integrity and competence as holders of their data.

5.5 Discussion

This study examined the GDPR’s effectiveness in improving individuals’ access and control over their personal data. The participants’ experiences support the existence of a power imbalance over data (see section 2.1) and suggest GDPR largely fails to empower individuals: both objectively (to the extent possible by this limited sample), in that most companies do not comply fully (either by returning insufficient and inadequate data, or by failing to return data on time or at all), and subjectively, in that returned data was often difficult to understand, impractical for use, and raised new questions and concerns. The findings also indicate that swift, transparent, and easy-to-use GDPR procedures can positively impact an individual’s perception of an organisation. In light of these findings, this discussion offers insights on how the personal data landscape might be redesigned through policy (5.5.1) and business practice (5.5.2), and how individual action can have important impact too (5.5.3) – all in pursuit of the human-centric empowerment goals described in 5.1 as well as 2.2 and 2.3):

5.5.1 Implications for Policymakers: Compliance, Quality and On-going Access

Despite significant and obvious GDPR-motivated investment by service providers in dashboards, processes and bespoke data package production, the findings (while limited by the small number of participants) indicate that inadequate compliance with the GDPR is common. The findings are consistent with literature too: the participants’ issues with completeness and compliance echo those first reported within the GDPR’s first year (Ausloos and Dewitte, 2018), suggesting completeness and compliance have not improved over this period. However in this study, the focus was on the effectiveness and experience of engaging with GDPR procedures from the individual’s perspective. Participants’ experiences were overwhelmingly of disappointment and frustration, with their hopes rarely met. They found that data holders often did not engage meaningfully with the process, and that the responses typically excluded or obscured data that could have provided them with the insights into their data privacy and the organisation’s data practices that they sought. Evaluations of perceived power

compared to data holders largely remained the same or worsened after accessing data through GDPR, and participants were not confident in the capabilities of the legislation to shift the balance of power. The process was perceived by some as a “*box-ticking exercise*” that was both frustrating and time-consuming and did not ultimately help them. Even though in 7% of cases participants did feel empowered by the GDPR, *all* participants receiving data were in practice left with the prospect of additional time-consuming and sometimes technically-skilled work to take advantage of or interpret their returned data. This suggests that to improve the situation, policymakers need to make changes towards:

- 1) Better Compliance Through Enforcement of Complaints.** At present, enforcement of the GDPR is uneven; each country has its own DPA (for example in the UK, this is the Information Commissioner’s Office or ICO) and complaints are rarely pursued for individual cases. Instead, cases are processed by specific DPAs in a form similar to a class action lawsuit. This means that individuals have little impact when they do raise a complaint, and many GDPR complaints “become lost or resulted in lengthy delay” (Burgess, 2021), or may even be erroneously dropped (Lomas, 2020). Until individuals have a clear and effective means to issue complaints (Baker, 2018) that result in enforcement action (or a clear threat of it), it is likely that individuals will continue to have little recourse other than to repeat the request and hope similarly dissatisfied individuals will act on their behalf. Data holders must be held to account when they do not deliver the full set of data that they report possessing, or when they fail to do so within the legally obligated time frame.
- 2) Policies to Enforce Better Quality Responses.** Many participants received data in frustrating formats, including screenshots, printouts or files that were too technical or littered with acronyms. Data was provided in formats too technical to understand, or not technical enough to be usable (see 5.4.3.1), showing a demand for both human-readable information summaries and machine-readable data files, where most providers typically provide only one or the other. Policymakers could provide suggested data formats or even propose new standards; this would help data portability, improve effectiveness (Gurstein, 2011) and legibility (Mortier *et al.*, 2014), can reduce costs through common tooling and catalyse the building of tools to interpret and understand data. Such standards are emerging (Morgan, 2020) as they are a technological necessity for data unification, but lack adoption. We note that the European Data Protection Board has published new guidelines [ADD REFERENCE] that could help somewhat to improve GDPR responses, though these do not offer new standards, and will not be as effective as legally-mandated policy changes.
- 3) Policies to Enforce Data Access as Ongoing Support, not One-Time Delivery.** A radical redesign of policy is needed to give people the practical outcomes they desire and, according to the GDPR itself, deserve. Data access needs to be seen as more than the one-off delivery of data files. People need understanding of their data and of its handling. Not only that, they need a timely, up-to-date and ongoing view of the changing picture of how they are seen in data,

and the occasional ‘snapshots’ with a 30-day delay that GDPR delivers can never deliver this. Giving people an ongoing awareness and understanding of their data is the measure by which compliance should be assessed. The explanations GDPR mandates are not forthcoming; of the 119 hopes expressed by participants (see Table 5), 70 (59%) related to acquiring greater understanding of data practices. 38 (54%) of these were unmet, and a further 15 (21%) were only partially met. By mandating data holders to support individuals with not just the delivery of data, but assistance to understand (and potentially make use of) that data, policies could become more impactful, not least because such understanding is critical to inform judgements around consent, loyalty and compliance.

5.5.2 Implications for Data Holders: Earn Trust by Opening Up Data and Enabling Users

While this study, and the GDPR itself, might seem adversarial to data holders given the goal to reduce their power by imposing new procedures, the findings emphasise the role of personal data in consumer relations. Data holders are likely aware of the paramount role of personal data in decision-making, but may not be aware of individuals’ perceptions about this. The findings suggest that failure to satisfy users who are concerned about the collection and usage of their personal data risks harms to consumer trust and confidence, at least for those users, and perhaps for others they might influence. In turn, however, this presents opportunities to use the mechanisms of the GDPR for customer loyalty and building better relations.

In 52% of cases, following the process of examining privacy policies and engaging in GDPR data requests resulted in a decrease in reported trust in the data holder. While such impacts may for now be minimal, as only a small proportion of users read privacy policies (Steinfeld, 2016) and—one can assume—an even smaller number conduct GDPR requests, this is likely to change as issues around data privacy and trust continue to take centre stage in global geopolitics (Véliz, 2020; Zuckerman, 2021). Furthermore, the growing number of businesses focused on “getting your data” or “taking control” (‘Whose data is it anyway?’, 2019; Dehaye, 2021; CitizenMe, 2021; Gener8, 2021; ‘datacy - About Us’, no date; ‘Ethi’, no date; ‘Digi.me’, no date; ‘Exist.io’, no date) suggest demand for data access is growing. From the findings, there are three positive takeaways for data holders:

- 1) Data transparency is an opportunity to increase customer loyalty and trust.** GDPR’s basic rights provide a starting point for delivering practical data transparency that will allow organisations to demonstrate that they are deserving of trust. By responding clearly and engaging openly and helpfully with GDPR data requests, organisations can demonstrate consistency between their privacy policy and their actions and demystify to their users the role that data holds in their business model. Research has shown that explanations can “*ease humans’ interactions with technology [...], help individuals understand a system’s function, justify system results, and increase their trust*” (Glavic *et al.*,

2021). This was borne out in our results: in 14% of cases, participants felt more trusting of the service brand as a result of their GDPR experience (sometimes even displacing prior apprehensiveness or distrust), citing reasons such as speedy, hassle-free responses, clear and understandable data, providers being upfront and open with data, and staff who exhibited a positive attitude to the request.

2) Data transparency is an opportunity for improved and re-imagined customer relations around data. Beyond the opportunity to improve trust, the mechanisms of data transparency suggested by the GDPR provide individuals with new capabilities for data curation and involvement. By offering individuals the ability to engage in empowering data interactions, data holders have the opportunity to improve engagement with their organisation and their services. If organisations view personal data as a shared resource to be curated and co-owned by the individuals that contributed it, there may be correspondingly shared benefits: for the individual, a sense of agency, influence and negotiability (Mortier *et al.*, 2014); and for the service provider, an incentive for individuals to generate and share more data, an increased likelihood of individuals correcting inaccurate data, and more reliable and human-centric forms of ongoing consent closer to dynamic consent (Kaye *et al.*, 2015) than today's ineffective models of informed consent (Luger and Rodden, 2013).

3) New customer demands indicate untapped business opportunities. As the 500-member-strong MyData Global organization (MyData.org, 2018) shows, there is growing demand for personal data empowerment. People's personal data is splintered and trapped (Abiteboul, André and Kaplan, 2015; Bowyer, 2018), and they cannot correlate data from different sources in order to reflect upon it, gain insights, and set goals (Li, Forlizzi and Dey, 2010). Due to commercial motivations, service providers generally deliver capabilities within a closed silo, not at the level of one's wider environment (Abowd, 2012). To be better empowered the individual could be the point of integration, the centre of their own Personal Data Ecosystem (PDE) (MyData, 2017). Life-level capabilities (Bowyer, 2021) and the opportunities that well-designed and well-regulated GDPR-type regulations promise in this regard have not yet been exploited. Thorough, complete and timely data access in standard formats, as mentioned above, will be critical to enabling this vision. Growing companies such as CitizenMe (CitizenMe, 2021), Digi.Me ('Digi.me', no date), Mydex (Mydex CIC, 2010), ethi ('Ethi', no date), HestiaLabs (Dehaye, 2021), udaptor (Udaptor, 2021) and exist.io ('Exist.io', no date) as well as larger organisations like BBC R&D (BBC R&D, 2017) and Microsoft (Microsoft, 2021) are already starting to innovate in this space.

5.5.3 Implications for Individuals: Becoming Aware of the Value and Power of Data, and Demanding More

While participants experienced disappointment and frustration in their GDPR journeys, all participants gained new understandings; if not always of their data itself, at least of their target companies' approach to data access requests. This

new knowledge was sufficient to re-affirm or challenge existing attitudes or inform judgements—P1, for example, left Facebook after the study. Even an attempt to access data can be educational, and even a cursory look at a provider’s ‘What data do we collect’ privacy policy section can provide pause for thought.

Today, individuals remain largely in the dark about the collection, use and sharing of their data through a combination of perceived complexity and effort combined with a lack of clear benefits. Table 12, alongside the increased control and insight promised by the PDE movement and platforms linked in 5.5.1 and 5.5.2 above, provide a glimpse of what the future may hold: a world where individuals take more control of their data and gain actionable self-insights. Three key messages for individuals can be inferred:

1) Your data is used to represent you and define your user experience. We hand over our data in exchange for access to services, but providers then use it (usually in aggregate) e.g. to inform product design or decide what content you see. This ‘innocent’ handover of data is in fact giving providers the means by which we are treated and – at times – controlled. Recognizing the crucial role of data (and our limited influence over it) is the first step to pursuing greater agency and control.

2) Your data contains meaningful and valuable data about your life. Data, as participants found, is dry and technical, but they all sought meaning and value within it (see 5.2.2). Within provider-held data lies potentially rich information about one’s life and past activity – some of which can even be inaccessible through any other means. This highlights both a risk (that others might gain this insight) and a potential benefit (that we could access this insight ourselves). In this context, data deletion without keeping a copy may be inadvisable. To access the value in data, individuals will need to demand data standards, better access and control mechanisms and insight tools.

3) Self-education and awareness enable accountability and informed choices. The findings highlight a lack of knowledge. Transparency is critical to judging ‘to what extent the bargain is fair’ (Larsson, 2018). It is not always delivered, but GDPR makes it your right; a right that cannot be fully refused. Through challenging poor GDPR responses and demanding better information, individuals can have impact. Providers are ultimately motivated by public demand—one of the reasons download dashboards exist. Through the public pressure of negative attention, companies can be motivated to improve data access (Dehaye, 2018). With patience, GDPR rights can be exploited to force small changes.

5.6 Summation

Through a longitudinal study of 10 participants lasting three months, this case study has qualitatively, and to a lesser extent quantitatively, evaluated the human experience of using one’s GDPR access rights and of living with data-centric service provider relationships.

The findings, while not statistically representative, suggest that people currently lack awareness of held data and its uses by service providers. By guiding participants on a journey of discovery and careful scrutiny, encouraging them to draw their own conclusions about service providers on the basis of companies' own promises, individuals' legal rights, and participants' own hopes (see Table 12), this research has shown that such a journey can be educational and enlightening with regard to increasing awareness, but also can seriously damage brand loyalty and trust in providers if comprehensive and well-explained data is not returned in a supportive and open manner (see 5.4.4).

The experience of GDPR seems to be an unsatisfactory one for individuals; participants were generally still 'in the dark'. Serious problems with compliance have been highlighted (see 5.4.2.2): Participants received data that was incomplete, impractical for use, and they failed to acquire desired explanations. By its own aim to enhance individuals' rights and control, the GDPR does not succeed. Participants continued to feel a lack of agency and choice, were largely unable to pursue goals such as data checking, correction or deletion, and their perceived sense of power within the provider relationship was largely unchanged by the experience. Nor does the GDPR allow individuals to adequately pursue their own goals related to accountability, self-reflection or creative data exploration (see 5.4.3.3). Individuals cannot be given power over their data through designing better Human-Data Interaction interfaces alone, but only through redesigned policies and business strategies that take into account the sociotechnical context (Baxter and Sommerville, 2011; Bowyer, 2021).

In order to bring the human-centric 'personal data ecosystem' concept closer to reality, action must be taken to improve both compliance and quality of GDPR responses. Considering these findings, there is cause for radical policy reform, to move away from 'data access as package delivery' and to provide individuals a more effective and ongoing two-way window into their data (see 5.5.1), providing ongoing awareness, accountability, and negotiability. Data needs to be expressed to individuals in ways they can understand, as little to no practical impact is currently being achieved by delivery of a one-time snapshot of some technical files; in fact, we have shown such responses can be harmful to customers' perceptions of the data holder in many cases.

For providers, the risk of reputational damage uncovered by this study should motivate them to engage meaningfully with data access requests; but such risk can be averted by redesigning both interfaces and processes to approach data access experiences as an opportunity to educate, and to build trust and loyalty, perhaps even through establishing progressive co-operative data stewardship relationships that truly *involve* the service user (see 5.5.2). While the GDPR experience is often disappointing and frustrating, it can provide insights that help individuals to challenge their assumptions, re-evaluate choices, and in some rare cases, feel empowered to act upon their data. Wider assertion of GDPR rights could demonstrate a desire for data holders to be transparent; without such visible demand, little may change (see 5.5.3).

Considering RQ1 (the pursuit of a deeper understanding of people's attitudes to everyday data holding and people's wants from that data), this work suggests that people struggle to develop the meaningful relationship with their data that they desire because of the difficulties faced in seeing, accessing and understanding it. They are aware that within data is the potential for value to themselves, but cannot access that value, which in turn causes feelings of resignation, concern, distrust or suspicion towards data holders. What they seek most are two things: sufficient understanding to better judge the value exchange they have signed up for with providers (see goals in top half of Table 12), and good quality insights from data that would allow them to understand themselves better, learn from the past, set personal goals, and harness personal data for their individual benefit (see goals in lower half of Table 12). This duality of needs around data interaction is expanded upon in (Bowyer, 2021).

With respect to RQ2 (the pursuit of a better understanding of the role of that data in everyday service relationships), the findings suggest that personal data, held by providers, as in Case Study One, serves as a proxy for direct user involvement, and is treated as such. Once users have sacrificed their data, or given permission for its collection, they are rarely consulted and most services exclude them from seeing how that data will travel through the organisation and be used in decision-making; this is consistent with the '*point of severance*' concept observed by Luger and Rodden (Luger and Rodden, 2013). As a result, the trust relationship between service provider and service user is extremely fragile, highly susceptible to subjective impressions of service brands, and as the findings show, discovery of poor data practices or a lack of transparency around data is sufficient to harm that relationship and in some cases even motivate individuals to change provider. As discomfort grows and scrutiny occurs, providers can expect customers to lose trust and loyalty. At the same time, this same data could play a central role in a re-invigorated relationship between a provider and a user, one based upon *earned* trust. It appears that providing easy, clear, data access and showing a willingness to respond to questions and explain data usage to users could be sufficient to allay concerns and instil strong customer loyalty. Of course, this assumes that the openness offered reveals practices the user finds agreeable, so perhaps this in some way explains why some companies that have more commercially-motivated approaches to personal data use (such as Facebook and Google) that many would find disagreeable upon examination, are apparently less willing to engage in transparency and user empowerment around data.

The general principles of earning trust through transparency, and rethinking data access as a means to involve users in decision-making, could be applied in a wide range of service endeavours that are currently very data-centric.

6 Discussion I: An Understanding of Human Data Relations

“For a participatory system to work, you need: a plausible promise, an effective tool, and an acceptable bargain.” – Clay Shirky (author and journalist focused on the social and economic effects of Internet technologies)

It will be already evident to the reader that there are significant overlaps and parallels to be drawn across the findings and discursive insights in Case Study One and Two. In this first discussion chapter, I will draw on both Chapter 4 and Chapter 5 to produce a unified summary of findings and insights in terms of the first two research subquestions RQ1 and RQ2. To recap on the research objectives expressed in 3.3, these two RQs are:

- **RQ1: “What is the human experience of personal data, and what do people want from their data?”**
- **RQ2: “What role does data play in people’s service relationships and how could relationships involving data be improved?”**

The answers to these research questions are best expressed as an understanding of individual *wants* relating to data. The word ‘want’ is used here in a broader sense than its everyday meaning, referring to the *lack* of something that would be beneficial (which may or may not be accompanied by conscious desire). By framing our accumulated understandings from the Case Studies in this way, we are exposing both the problem - the things that individuals do not have or cannot do, while also identifying the goals that any imagined solutions or improvements to the status quo would need to address. It logically follows that any solution that better delivers on individual *data wants* will lead to improved relations between individuals and their data. This is how we can conceptualise “Human Data Relations” as alluded to in the title of this thesis, and indeed this gives us a yardstick against which to understand what “better” means, which will be explored in Chapter 7. *“Human Data Relations”* is a term that I introduce here to expand upon the established theory of Human Data Interaction (Mortier *et al.*, 2013, 2014) in light of the Case Studies’ findings from a more sociotechnical, interpersonal point of view. Humans have two kinds of relationships with data: *direct* interaction (such as through an interface in an app or website) and *indirect* interaction (through interacting with services, providers or individual representatives who themselves have access to personal data about the individual). Thus, Human Data Relations is a term that can encompass both the relationship humans have with their data, but also the relationships they have in which data plays an indirect role. In this context, RQ1 and RQ2 map quite cleanly onto these two types of Human Data Relations, and in answering RQ1 we can identify what people want from direct data relations, while RQ2 helps provide an answer as to what people want from indirect data relations.

6.1 Answering RQ1: What do people want in *direct* data relations?

By comparing and grouping elements of the findings from Case Study One (see 4.3) and from Case Study Two (see 5.4), three distinct data wants are evident when considering *direct* data relations. All data about individuals needs to be:

- 1) *Visible*: People need to have knowledge of data about them and an ability to see it and effectively access it;
- 2) *Understandable*: People need to be able to interpret this data to extract meaningful information from it (and about it), including through visualisations and summaries; and
- 3) *Useable*: People need to be able to take action upon this data, including exploring it, asking questions of it, using it to serve their own goals, and gaining personal value from it.

These wants are detailed in the following sections:

6.1.1 Visible

Data matters to every individual now, in a way that previously it did not. As described in 2.1.2, the role of data in our everyday lives has changed; data has become a material used by businesses to shape our world. In 2.2.2 and 2.2.4 I outlined how this change has practically manifested itself in people's lives; where once data was viewed as *ours* for our personal reference and use, the emergence of complex multi-party data ecosystems has meant that personal data management has become a sociotechnical & societal problem, not a practical individual one. In the past when businesses were more local, more personal and less data-centric, the data that businesses held about us was minimal and much less significant to our lives than the human relationships we had with those businesses. Businesses grew and data began to be considered as a resource to be processed at scale for customer insight and marketing exploitation, and though we didn't realise it, our need to understand those processes to protect one's own interests, began to grow. In the past, you didn't need to become aware of data storage and use, because it had little effect. Now, data has become a substitute for direct communication with the individual being served, as my research in both public sector ((Bowyer *et al.*, 2018) and 4.3.3) and the private sector (5.4.3.3) has shown. In both domains, people do not have awareness, let alone access, to the extent of data that exists about them. In 5.4.4.1 we saw individuals feeling that companies forced them to hand over data in exchange for service access, and then subsequently maintaining power over them through holding that data, using it to make decisions, and denying them access to that data or even be clear about what data is held.

Data sacrifice is now required for many services, putting individuals at risk. Be it the personal financial, health and lifestyle data collected on an Early Help assessment form when a family signs up for Early Help support (4.1.2), or the contact details, payment information and preferences provided

when individuals register with commercial service providers such as insurance providers or streaming media platforms (5.4.4.1), supplying your personal data is required to access services. Consent to hold and use this data is enforced upon signup, through waiver forms or Terms & Conditions agreements. Service providers in both sectors see the acquisition of more personal data as beneficial to their operations, be it support workers wanting to gain more data about families' lives (4.2.3, 4.2.6), or commercial providers using trackers to gain more insight about users that they can exploit for advertising (Binns, 2022). In both domains, this sacrifice is seen to have an emotional effect on people, ranging from curiosity to fear and distrust ((Bowyer *et al.*, 2018) and 5.4.4). Such fears are well-founded, with mistreatment through incorrect data known in both settings (4.2.2, 5.4.4.1). While data holders almost certainly do not *intend* to cause harm, data *can* be 'used against you' (P2's quote in 5.4.4.1) (Kröger, Miceli and Müller, 2021; Strohmayer *et al.*, 2021). In providers' eyes, people are now represented through data. Despite the fact that data is never truly objective (Gitelman, 2013; Taylor *et al.*, 2015) and a recognition (at least on the public sector side) that a data record can never tell the full story (4.2.6 and (Bowyer *et al.*, 2018)), the data record becomes the object to be administered, rather than the individual (Cornford, Baines and Wilson, 2013; Zuboff, 2019), and this in itself creates risk - through mishandling or inaccuracy. Given the data record is seen as a source of truth (4.1.2, 5.5.3), it is vital that such information remains *fair*, and *accurate*; this is especially important in the commercial sector, where people are only just beginning to become aware of data misuses and data exploitations that are actually happening (Chang, 2018; McNamee, 2019; Zuckerman, 2021; 'Data Brokers: Everything You Need to Know', 2022). Ensuring fairness and accuracy of held data cannot be verified without individuals' awareness of data held about them.

Once data has been sacrificed, it enters a closed and opaque ecosystem, where the individual loses access and becomes unaware of that data's storage and use (Luger and Rodden's 'point of severance' (Luger and Rodden, 2013)). What was previously available for individuals to see becomes inaccessible and invisible. In the Early Help context, this manifested as families having a lack of awareness or direct access to data held about them and having to rely on support workers as gatekeepers to choose to inform or show them aspects of their data (4.1.1,4.5). In the commercial context, the situation is perhaps even worse, as not only is there rarely any kind of data viewing interface, there is not even a gatekeeper who might make people aware of their data and its use; and even if someone becomes motivated to gain awareness, the GDPR leaves them in the dark; in 62% of cases, the data that companies own privacy policies stated they collect, was not returned, and data that was returned was complete in only 22% of cases (5.3.2). In both contexts, no awareness is gained unless the information is actively sought. This means that the vast majority of people, busy and unaware, remain so. This is problematic because people cannot judge data accuracy or protect themselves from risk, because they may not even be aware of certain data's existence and use, or be able to access it even when they

are.

People want to see data which is hidden from them. In the SILVER project (3.4.1.1, 4.2.2), and my prior work with families (Bowyer *et al.* (2018)), and in Case Study One (4.2.6, 4.3.2, 4.4.2), families wanted to see what data was held about them (“what they’ve got on me”). They wanted to be actively kept informed and to have the ability to see if data was fair and accurate. In the commercial context, the same feelings were found; participants expressed a great desire to see and know what companies are storing about them, especially data collected or inferred about them without their involvement. This is not just a desire, but a need, given that data can have impact on their daily lives as it is used to inform decisions on how content is presented and recommended to them, and what services they are advertised, offered or can access (5.5.3). We also see from Case Study Two that awareness is not just a binary; awareness includes having an appreciation of why the existence and use of certain data is significant and what its implications might be. Article 13.2.f of the GDPR ('Article 13: Information to be provided where personal data are collected from the data subject', 2020) states that, at least in the case of automated decision-making, people are entitled to meaningful information about the significance of the processing of their data, yet such explanations were typically not given to participants of Case Study Two.

Effective access to held data is required for visibility. Having gained awareness of data held and of the significance of it, people want that to be accompanied by meaningful access to the data itself. In Case Study One (4.3.2.1, 4.4.2), we saw families and support workers recognise the need to accommodate the differences in families’ digital literacy, mental and physical handicaps, and technology skills while providing them access to the data held about them. This mirrors Gurstein’s call for ‘effective access for everyone’ (Gurstein, 2011), which was detailed in 2.1.4. Four aspects in particular are relevant here: the content and formatting of the data (which should support different levels of linguistic and computer literacy), the capabilities made available in terms of software, hardware and Internet access (sufficiently powerful, sufficiently available and affordable), and skills (ensuring that individuals are able to interpret the data). In Case Study Two, we saw several participants feeling that data was delivered in too-technical formats (5.4.3.2), or that they lacked the skill to properly interpret the data (5.4.3.1). Effective access and interpreting data goes beyond visibility of data and includes understandability, which is explored in the next section.

Visibility of, and access to, data must be timely, and ongoing. Given the ever-changing nature of data (and indeed of the lives of the people it represents), occasional or one-off access is not sufficient. In Human-Data Interaction theory, this concept is described as having *negotiability* (Mortier *et al.*, 2013, 2014): the ability to re-evaluate data and associated decisions as contexts change externally. It is also mentioned by Gurstein, who points out that time-limited access to data would not be effective (Gurstein, 2011). In the Early Help context, families wanted access to their data outside of support meetings; this implies some sort

of self-service interfaces being available, that you can use *in your own time* rather than being reliant on the support worker as gatekeeper. People wanted to see all data about them directly, through a personal interface, as reflected in their workshop designs (4.3.2.3). This echoed findings of my earlier work with families, which had identified a need for continuing rights and visibility of data over time, in order enable vigilance over keeping data accurate and meaningful as life changes (Bowyer *et al.* (2018)). Timeliness also implies that access to an up-to-date view of the data does not require special and ongoing effort by the individual, it is always available. Both support workers and supported families saw value in notification feeds about changes to data records (4.3.3.3), so that changes are discovered and can be acted upon without having to wait for the next support meeting. In the commercial context we can see that GDPR provides a form of access that is not at all timely. The 30-day delay on request processing guarantees that data will be out of date by the time it is viewed, and individuals must repeatedly make GDPR requests to maintain an up-to-date view (and in doing so, they risk the imposition of charges as GDPR states that requests should not be excessive and that fees can be levied for additional copies). This lack of timeliness in the design of GDPR data access motivates my third suggestion to policymakers in 5.5.1, that they should offer ongoing access rather than the one-off delivery of data packages.

For held data to become visible, systemic support is needed, including governance, advocacy and assistance. Offering access to data is not solvable at a purely technical level. Even a well-built data interface with 24/7 access would not provide the depth and breadth of visibility people want. As observed in Case Study Two, even those companies that provide instant data access portals such as Google and Facebook did not provide participants with all the data they desired, nor all the answers they sought (5.4.2.3), and most companies offered negligible follow-up support after data had been delivered (5.3.3). Further investigations into data access conducted as part of the #digipower investigation (3.4.3.4) confirmed Case Study Two's findings that SAR requests and data portals rarely provide insight into some of the most desired types of data including derived and acquired data and data transfers. Effective access and visibility also requires advocacy (Gurstein, 2011): people require support and training to make use of their data. Furthermore, given the insufficient breadth of returned data from companies (5.3.3, 5.4.2.2) and near-total lack of access to data on the public sector side (4.3.2.2, 4.3.2.3), it is clear that external governance (Gurstein, 2011) to ensure effective access is needed. Without the sort of pressure on data-holders that only policymakers can exert, organisations will not be compelled to provide richer responses or better information-access support (5.4.2.2, 5.5.1), and while small improvements can be achieved through individual action, people generally lack the means to effectively *demand* the increased visibility required (5.5.3). The impact of this lack of governance is most keenly felt in the PDE/MyData space (2.3.4), where emergent actors seek to encourage data-holding organisations to enable greater information access so that they might build better data access tools for individuals, but are hampered by a lack of top-down governance supporting

their requests as well as a lack of funding and investment by data-holders in data advocacy.

If data is not visible, this can lead to subjection, alienation and exclusion. Throughout both Case Studies, we have seen the negative psychological effects of people not being able to see their data. Families in both my earlier work with families (Bowyer *et al.*, 2018) and Case Study One were caused significant worries by not being able to see their data. People do not want to be treated like *subjects* (in either sense of one being subjugated (Bowyer *et al.*, 2018), or as a topic being discussed) and reducing people to a set of assertions in data causes them to become, in effect ‘objects to be administered’, which is harmful and disempowering (4.2.3, 4.3.4.2, (Cornford, Baines and Wilson, 2013)). Supported families felt helpless and resigned to being judged through data and sometimes suspicious of those holding or using that data (4.3.4.1). This led in some cases to withholding of information or distrust of support workers, harming the effectiveness of a relationship that is designed to empower (4.4.1). In that same section (and in (Bowyer *et al.*, 2018) Page 7) I outline how ongoing individual access to data has the potential to transform attitudes, remove dependence and a feeling of being a subject, and could empower families to help themselves. Such fears and worries about unseen data were echoed in Case Study Two, with participants exhibiting great unawareness of held data (Table 12, 5.3.3), and concerns over data being held out of their sight for long periods of time (5.4.3.3) as well as similar feelings of resignation or lack of choice (5.4.4.1, 5.6). Denying access to held data was seen as a key source of holding power over individuals (5.4.4.1), and visibility of data is a key part of assessing ‘to what extent the bargain’ (of data sacrifice for value as described above) ‘is fair’ (2.1.4, 5.5.3 and (Larsson, 2018)). It is an inherent consequence of representing people through data and then using that data to make decisions (2.1.2, (Cornford, Baines and Wilson, 2013; Bowyer *et al.*, 2018)) that individuals become sidelined and excluded (2.3.3 and (Crabtree and Mortier, 2016)). Without visibility of data, consent is not meaningful, and individual needs are more easily ignored or overlooked.

6.1.2 Understandable

Visibility and access to see data is not enough, people need to be able to interpret it. Data is only valuable in so much as it enables us to access the information which it encodes (2.1.1). People need to be able to make sense of it. When humans look at data, we inevitably attempt to interpret it to see what it can tell them; in Early Help, support workers try to learn more about people’s lives by examining data about them (4.2.3, 4.3.3.1). In doing so they apply their own knowledge and expectations in an attempt to extract facts. Similarly in the context of everyday digital life data, individuals search for value and meaning in that data, they reflect upon it and try to relate it to their own lives (5.4.3.1). While Early Help staff receive training on how to understand families’ data, individuals struggle to understand their data without sufficient support, as discussed above. Returned data from GDPR requests is often dry

and technical. It may contain codes, internal notations or abbreviations that a layperson cannot understand (5.4.3.1). Raw data is rarely sufficient to provide clear, unambiguous and unbiased information to the reader (Gitelman, 2013; Neff, 2013). In line with one of the three core principles of HDI, *legibility*, data should be understandable by those it concerns (Mortier *et al.*, 2014). In both Case Studies (and my prior work), individuals shared a desire to not just be aware of, but to **understand** what data was held about them and how it was used ((Bowyer *et al.*, 2018), 4.3.2.4, 5.4.2.1). People are only just beginning to understand the significance of a data-centric world that uses data to make decisions that affect their lives (2.1.2, 2.1.4, 5.5.3).

People need understandable summaries of information content and context. It was very clear from the findings of both Case Studies that all humans looking at data need *summaries* to help them digest and locate key information. In Case Study Two participants were often overwhelmed or “*drowning*” (P1) at the volume or technical complexity of the data returned from access requests, “*so much of it that’s impossible to know what it all means*” (P4) (5.4.3.1). These feelings were mirrored in Case Study One, by support workers who feared the liability of having to “*trawl through*” large volumes of data and know all the relevant and important facts about a family so that they do not make mistakes (4.3.2.1). Participants on both sides talked of needing help to see the whole picture, something that is hard to achieve from individual datapoints or sets of files. In both cases, summaries of data would help comprehension. However, the task of creating a summary is not straightforward and places power in the hands of the summary-maker, who can decide what is relevant, how the data is framed and what is omitted from the summary. People look at information for different reasons, to answer different questions, so the question of who decides what is relevant or most important within a body of data is a critical one. Different summaries would be needed for different audiences. As Mortier reminds us, effective legibility requires a recognition that individuals’ viewpoints of data can and should differ (Mortier *et al.*, 2014). There is a question about who decides what the viewer of a summary ‘needs to know’ (4.3.3.1, 4.3.4.3). This is further complicated by the fact that the data itself is not neutral (Gitelman, 2013); in the Early Help context it was clear that opinions as well as facts are recorded ((Bowyer *et al.*, 2018, p. p6) and from SILVER project), and a focus on the recording of data most helpful to the support worker. Commercial data holders record data in ways that are optimised for their existing systems and processes, as seen through the presence of internal codes, system screenshots and filenames in returned data (5.4.3.2). In Case Study Two most participants’ comments on returned data indicated that it had not been presented in a way optimised for understanding (5.4.3.1), failing to support *sensemaking*; “*Information presentation should be as clear as possible so that people can interpret their data and extract meaningful information from it.*” (Gurstein, 2011)

Rather than raw data, people need information and visualisations, arranged and optimised for understanding. Data, by itself, is not meaningful. In order to be able to answer questions and acquire knowledge, people need

information (2.1.1). Access to raw data files or database records or spreadsheets does not satisfy this, and prior research in the civic data context states that this would be inadequate and limiting (Cornford, Baines and Wilson, 2013). To comprehend the meaning of data, visualisations and explanations can help (4.4.2); as one support worker in Case Study One observed, some families might find data tables too technical, “*I think sometimes it’s easier to do it in pictures*” (4.3.2.1). Participant-designed interfaces in Case Study One included pie charts, graphs, spider diagrams and timelines, all designed to convey information more intuitively (4.3.2.1). In Case Study Two and prior GDPR requests, it was often the case that companies often returned data not in understandable forms that were less useful than the apps or websites those service providers offer. For example, run tracking apps such as Nike+ and Strava return route log information in XML-based TCX files which are meaningless without some analysis tool or visualisation. JSON files, a commonly returned data format, often use a timestamp format that is just a long number, not understandable by humans without extra work. Data was often returned in formats that were more a reflection of internal systems (e.g. screenshots, table dumps or exports) than being optimised for understanding (5.4.3.1), which some participants found useless. As P5 observed, “*It’s like being given the bricks to a house... It doesn’t really mean anything when it’s just bricks, if you don’t know how to put it together*” (5.4.3.2). It is clear that visualisations are key to accelerating understanding (and are also subject to the same challenges of selection and bias as summaries). Furthermore, visualisations of data can functional as powerful boundary objects and “*things to think with*” ((Brandt and Messeter, 2004; Bowyer *et al.*, 2018) and 3.5.2, 4.4.2) and in the care context the use of data representations as a focal point or evidence for was thought to be more productive and empowering (4.3.2.2, 4.4.1).

Information becomes most meaningful when it is recognisable and relatable and can be mapped back to life experiences. Across both case studies, there was a clear search for meaning in data, which manifested as a desire to use that data to build a fuller picture of the individual (or family)’s life. In Case Study One and in SILVER, it was evident that support workers seek a broader view into supported families’ lives by reviewing data about them (4.3.2, 4.3.3.1). Both supported families and staff saw value in seeing all the data about each individual in a common place - in other words, structuring the information around the individual’s life, rather than the information silos of different agencies (4.3.2.3). In Case Study Two, where participants were asked about the value they saw in the data companies held about them (5.3.3), it was clear that data was most valuable to participants when it was recognisable and relatable to events in their lives. This is particularly important for data that participants have never seen before such as derived, acquired or metadata - without a way to connect it to one’s own life, it is impossible to relate to (5.3.3). Echoing goals surfaces in personal informatics literature (Li, Forlizzi and Dey, 2010), participants sought insights about themselves in data, and so valued data that spanned a longer time period such that they might use it to spot patterns in relation to events

in their lives (5.4.3.1). They valued the opportunity to use data as “*a window into the past*” (P11, 5.4.3.1). Long term data was seen as a liability (5.4.3.3), and when it came to considering attitudes to providers, participants were most concerned about data collection by the larger data companies like Google and Facebook who, through their myriad apps, websites, devices and other means, had many touchpoints into their lives and thus had a broader picture of their life activity (5.4.3.3. 5.4.4.2). It is clear that to transform dry, technical data into meaningful information, it should be contextualised in relation to events in the life of the individual it describes.

In practice, ongoing human support is needed to facilitate the understanding of data. In both case studies, participants felt that they had questions to ask of data holders, especially when the data was difficult to understand but also more generally when the answer to a given question could not be found in the data. In my earlier work (Bowyer *et al.*, 2018) and in Case Study One, participants emphasised the ability to talk to someone about their data (4.3.2.4), and included such features in their designs - both to understand and receive help in understanding, but also to feedback and provide additional information, explanation or context so that they might be better understood (4.3.3.4). In some cases families needed additional support for reasons of accessibility or technical literacy. Human communication channels for data support need to be available at convenient times too; in Case Study One individuals felt constrained by needing to wait until their next meeting with a support worker, and wanted a communication channel they could use in their own time. In Case Study Two, participants were regularly frustrated by data they could not fully understand and could not ask about: sometimes literally, when internal codes were not explained or technical formats were used, and sometimes when they could not understand how the information had been derived, such as insurer’s driving scores (5.4.3.1) or Instagram’s inferred interests. The general pattern of GDPR handling by organisations was to deliver data, often handled by a back-end team with no customer face or means to ask follow-up questions. Where questions were asked they were typically hindered by delays and middlemen preventing an effective conversation (5.3.3). Considering the systemic changes toward data-centricity that the world has undergone, as described in 2.1.2 and 2.2.4, it is no surprise that human support has reduced. Across both studies we see the cost of that shift toward dealing with data instead of dealing with people; individuals get left behind, without the means to understand or ask about their data.

If data is not understandable, distrust can arise. In both Case Studies, the costs of individuals not being able to understand their data and being left in an unsatisfactory position of being unable to resolve concerns or ask questions is evident: Without understanding, comes distrust. In Case Study One, participants were concerned when they saw assertions on their records that they disputed or could not identify the source of. In Case Study Two, privacy policies that were too vague in their explanations of data (such as Google’s) caused participant distrust because they seemed broad and lacked tangible examples, and large volumes of technical data caused suspicion. It is

clear that understandable information about what is represented in data, and about the context and use of that data, can help individuals to trust the data holder. In the minority of cases in Case Study Two where the GDPR had a positive impact, the reasons given included understandable data and helpful human responses, for example P10 cited Niantic's detailed data and positive attitude, while P6 described Sunderland AFC as upfront, and said that their data "*just made sense*". P7 specifically distrusted LinkedIn because she felt they had not bothered to adequately engage with her (5.4.4.3). The importance of trust will be discussed in more detail in 6.2.1 below.

6.1.3 Useable¹⁶

People need to be able to explore and interrogate data to ask questions of it. In both case studies, many participants showed they are aware that their personal data contains insights and value they cannot access. In Case Study One and (Bowyer *et al.*, 2018) this manifested as concern over what unseen incriminating judgements might be storied in their records, and an awareness that the only data supported families can actually see is whatever subset of data their support worker chooses to share with them (4.3.4.1). In Case Study Two, this manifested as feelings of not having the technical skills necessary to explore their returned data (5.6) and that without better tools individuals were not getting the same view of data that service provider staff have (P2, 5.4.3.1). What we can see from these observations is that even visible, understandable data is not enough to meet people's needs. People need to be able to interact with their data, in order to explore it, gain insights or answer questions. There is a need for tools not only to access data, but to help people make sense of (often technically formatted) data (5.4.3.1), and explore it in different ways to answer questions - which implies the need for operations such as filtering, searching, comparing and orienteering (see 2.1.4 for others), in order to understand history, context and patterns in the data, as suggested in self informatics (SI) literature (2.2.3). Several participants in Case Study Two hoped to receive data in formats they could visualise, "*mash up*" and play with, but did not get this - not only was data not optimally formatted for such use, it lacked sufficient explanation to enable individuals to pursue such goals for themselves (5.4.3.2).

Data needs to be useable - correctly formatted and explained in a portable and standardised form. As discussed in 5.6 and 5.4.3.2, people have multiple needs around data: For understanding, people need *understandable information*, both the facts and assertions encoded within the data, but also

¹⁶The words '*usability*' and '*usable*' (spelt without an 'e') most commonly refer to a judgement of the degree to which a website or user interface is easy to use (Nielsen, 2012). Throughout this thesis, I deliberately use the alternative word spellings of '*useability*' and '*useable*' (Collins English Dictionary, no date a, no date b) respectively, to clearly distinguish from this ease-of-use concept and to denote that I am referring a different meaning: the more literal definition, i.e. "*the quality or state of being convenient and practicable for use*" (Merriam-Webster Dictionary, no date a, no date b). Any usages without an 'e' can be taken to refer to the interface ease-of-use concept.

information about the data itself - its context, history, use, and significance. But there is a distinct need for *usable data*, which is somewhat orthogonal to this. A PDF containing a screenshot from an internal system might potentially be sufficient for understanding, but is useless for exploration or visualisation-building. Similarly a technical log file might contain rich data that can be queried and visualised given the right tools, but without those tools or an informational summary, is of no immediate value (5.4.3.2). This dilemma was alluded to by P4 in Case Study Two thus: “*If you want to view the data they have about you, it’s quite usable. If you want to do something [analytical], then it’s not*” (5.4.3.2). This dichotomy of needs is discussed in 5.5.1, where the introduction of standard formats is proposed as a means to catalyse the building of data insight interfaces. As P1 stated, “*it would be nice if these companies had a standardised model of how this information is presented to people*” (5.4.3.1).

People need to be able to interact with data, which means interfaces are needed By themselves, even standardised files as described above are not practically usable. As discussed in 4.4.2 and consistent with effective access (Gurstein, 2011) participant data designs in Case Study Two remind us that simply providing data is not sufficient: to be meaningfully able to act upon data requires some form of interface not only for visualisation and interrogation as mentioned above, but also so that any physical, cognitive or accessibility needs can be met. In Case Study Two, some participants wanted not just for data access, but for tools to help them find insights from their data (5.5.3, 5.6). While several companies are starting to create interfaces for data access, most of these are still focussed upon file delivery, with the notable exception of Google Timeline and Google My Activity, which provide a glimpse how an interface to explore data could be more useful than providing a bundle of files. Both of these examples also re-iterate the value of unifying data around an individual’s life, as discussed in 4.3.2.3 and above in 6.1.2 above.

Data needs to be explorable from a temporal perspective. Another aspect of usable data that goes beyond what a data file can offer is the ability to view it over time. The importance of this temporal capability, as identified in literature cited in 2.1.4 (practical information access) and 2.2.2 (temporal PIM systems), and my prior writing (Bowyer, 2011), was evident in both Case Studies. In SILVER and in Case Study One, being able to access historical data for a full picture regularly surfaced as a desire in discussions - even though the exact bounds and mechanisms for achieving this were contentious (4.2.6, 4.3.2.1, 4.3.3.1). In Case Study Two, as Table 12 shows, 26% of participant goals related to SI-type reflection (2.2.3) on one’s past to enable self-insights, nostalgia and creative uses of data. No participant was able to achieve this, and this was in large part due to the lack of temporal data exploration capability, meaning that significant potential value (the value of a long-term dataset as described above) remained locked away and inaccessible (5.3, 5.6). Case Study Two also suggested a lack of thought to this temporal perspective from data holders, who delivered data as a one-off snapshot that was already out-of-date when delivered. Indeed the GDPR explicitly discourages excessive data requests, rendering an ongoing

view of data as proposed in 5.5.1 near impossible. Having better temporal data exploration capabilities would enable people to understand themselves and their data ecosystems better, informing both personal self-improvement goals and better decisions about personal data practices and provider choices (5.6). Other exploration perspectives that could be powerful include location-based views or person/company/relationship-based views.

People need to be able to access the valuable knowledge locked within their data. People need to see the complete picture of their data in order that they might access at least glean as much value and insight from it as data holders do (5.4.3.1). For self improvement and improving one's situation (a key goal of Early Help), access to metrics visible in data are extremely important, so that one might measure progress (4.3.2.4). While support workers do endeavour to provide this and sometime show data extracts, this can never be as empowering as having full capabilities to explore that data. "*in your own time, at your own pace*" [Bowyer *et al.* (2018); 4.3.2.4]. In commercial service provider relationships, data interfaces present data in ways that are configured to reflect the profit motives of the organisation (explored further in my work with Goffe *et al.* (Goffe *et al.*, 2021)), and so accessing a relevant view of one's own data and having the means to access the knowledge within it is similarly difficult. Participants in Case Study Two found returned data unusable (5.4.3.2), and wanted visualisations that would allow them to discover patterns and insights, and tools to explore their data (5.4.3.1, 5.5.3). If data holders can make data not just visible but useable, individuals can learn about the valuable knowledge that data holders are currently extracting from their data, and hopefully how to access that knowledge for themselves.

If holders do not make data usable, this is a barrier to individual agency and power. In Human-Data Interaction terms, people need not just legibility, but *agency* - the ability to act upon one's data (Mortier *et al.*, 2014). Personal data contains valuable and actionable information about individuals and their lives (5.5.3). The SI field has identified that there are already many practical barriers to working with one's personal data effectively, including not just access but challenges in integration, sensemaking and goal-tracking – the 'barriers cascade' (Li, Forlizzi and Dey, 2010) – which hinder the ability to use one's data for personal benefit. As observed in Case Study Two, an inability to access the value in your own data can lead to feelings of resignation, concern, suspicion or distrust (5.6). Even if one can see and understand one's data, an inability to act upon it can reinforce feelings of being passive and uninvolved; without this, any opportunity to feel engaged and motivated is lost (Bowyer *et al.*, 2018, p. p8). Being able to use data for one's own purposes is a critical ingredient of empowerment and rebalancing power (Hoffman, 2014a). Without data usability, individuals are in effect digitally impaired, leading a less functional society where innovation and growth is limited (Abiteboul, André and Kaplan, 2015, p. 2.1.4).

6.2 Answering RQ2: What do people want in *indirect* data relations?

By comparing and grouping elements of the findings from Case Study One (see 4.3) and from Case Study Two (see 5.4), especially in the context of individual relationships with care providers and digital service providers respectively, three distinct data wants are evident when considering *indirect* data relations:

- 1) *Process Transparency*: People need to know what data is being collected or held, and critically how it is being used, for accountability and safety and in order to have trust in data holders;
- 2) *Individual Oversight*: People need the ability to affect what data is held and how it is used, including reacting to changing circumstances, deleting data or withdrawing consent for certain uses; and
- 3) *Involvement in Decision-making*: People need to be invited and involved in decision-making based upon their data, so that they are not misrepresented and their needs are not overlooked. This can be aided by collaborative use of data, giving individuals a human point of contact, and consulting the person not just the record.

These wants are detailed in the following sections:

6.2.1 Process Transparency

People need a window into how their data is used; this means transparency of processes not just of data. It is well established that there is currently extensive use of personal data by service providers and other parties that is beyond an individual's view (Hoffman, 2011), forming an ecosystem of data use based upon one's data, which is currently not centred on or visible to the individual concerned (2.3.4). Decisions made based upon personal data directly affect people's lives through policy decisions (in the care context) or business/functionality decisions (in the commercial context). People need to understand the value created by the use of their data and how (if at all) they are compensated for this (Hoffman, 2011). Even with full data access, understanding and useability, individuals cannot see into this opaque world of data use; the data is just an artifact produced and shaped by unseen processes. Like an archaeologist trying to infer the customs of lost civilisations through ancient relics, observation of the data can only reveal so much. Andrew Cormack, writing before GDPR, observed that "*it is more important to know how information is processed than the actual values involved*" (Cormack, 2016). The SILVER project found that families had very little awareness or understanding how their data was used, and that consent was therefore not meaningful because consent had been given without processual understanding [4.2.2]. In Case Study One participants agreed that people need rights to see how their data is used (4.2.6). Case Study Two revealed a clear desire for awareness of how data is used, how decisions are made, and how this might affect them (5.4.2.1), with over 74% of goals in pursuing GDPR requests relating to wanting greater insight into personal data

use practices (5.3.2). 70% of participants wanted to understand what providers infer from their data and this was unmet in 73% of cases and fully met in only 7% of cases (5.3.2).

Process transparency is required to enable accountability. In Case Study Two, participants recognised that organisations had collected data about them which could be exploited, and wanted to understand the extent of that capability (5.4.3.3). Data access can provide a window into collection capability, but only process transparency can reveal the extent of data use capability. Many participants expressed a desire to assess the trustworthiness of their service providers; they had curiosity, suspicion and unanswered questions that only transparency could address (5.3.2) and sought to judge whether data use practices were “appropriate” (5.2.4.1). In the Early Help context, all data processing is hidden from individual view and no access or questioning capability except through their support worker (functioning as a selective gatekeeper) (4.1.1, 4.2.1, 4.4.1). Returning to Case Study Two, there was evidently some transparency available in the form of the ability to make a GDPR request, but many participants found GDPR responses inadequate for holding providers accountable (5.2.4.3). Nonetheless, data access request handling is itself a data process, and so in this sense, the GDPR process did offer some ability to judge the trustworthiness and integrity of providers in data handling, in part informed by the breadth and quality of data returns (Table 11) but perhaps more so by the experience of the GDPR process. Many participants formed or revise their perceptions of companies, with perceptions of providers having a lack of care and making access difficult, or of providers being helpful and open having a strong impact on participants’ attitudes toward them (5.4.4.3).

There is no accountability, processes are not transparent, and thus power remains imbalanced. Across both case studies, the lack of process transparency is clear. Early Help services have no obligation to describe or share their data use practices with supported families, and apparently only even attempt to do so at the point of initial onboarding and consent collection (Bowyer *et al.* (2018), 4.2). This equates in practice to a complete lack of accountability over data practice (4.4.1, 4.5) Meanwhile in the commercial context, some companies failed to respond at all to GDPR, which is a barrier to accountability service providers. Many routinely failed to adequately meet the transparency rights stated by GDPR, without repercussion or consequence (5.4.2.2). In both sectors, data holders’ freedom to collect and use data without adequate transparency or ability to be held to account can be seen as an exertion of power over individuals. The power imbalance (Hoffman (2011), Hoffman (2014a)) and the dominance of data holders over the individuals about whom data is held, is reinforced by a lack of transparency.

People face an incomplete picture of their data ecosystem, even after using all available means to achieve transparency. In SILVER and my prior work in the Early Help context, supported families expressed concern that data could not adequately represent the complexity of their lives (Bowyer *et al.*

(2018)), a view supported by literature (Cornford, Baines and Wilson (2013); Gitelman (2013)) and sustained by support workers & staff in Case Study One (4.2.6, 4.3.3.1). Therefore, transparency is vital in order to ensure data is fair and accurate to them, yet it is not available (4.3.4.1) so they have no means to ensure this. Similarly in Case Study Two, the most popular GDPR goals around understanding inferences made from data about people (5.3.3) remained unmet. Data was incomplete, delayed, or inaccessible (5.3.2) and the potentially most informative type of data when it comes to understanding processing – metadata, derived and acquired data – were typically absent. Apparently broad responses were discovered to be very limited when viewed through the lens of privacy policy commitments and GDPR transparency rights (5.3.4).

Trust in data holders is needed, and gaps in transparency create distrust and a risk of broken expectations, harming relations. Individuals need a functional understanding of their data and its handling, and this is crucial to trust. Good explanations (as were often found lacking in Case Study Two (5.4.3.2)) can deliver some of this needed understanding and subsequently increase trust (Glavic *et al.*, 2021), as observed in a minority of cases (14%) where a good GDPR response led to the participant's trust in their provider increasing. Conversely we found that incomplete data (or a general lack of transparency/difficulties of access (5.4.4, 5.6)) can harm trust, as in the majority of cases (52%) can harm trust (5.3.4), leading to thoughts such as ‘what are they hiding?’ (5.4.4.3). Privacy policies that contradicted expectations or lacked sufficient explanations also led to distrust (5.3.4). Trust in the independence and integrity of data holders is essential (Dijck, 2014), and this was often a concern in the Early Help context, where trust between support worker and supported family is especially critical in order for the support relationship to be effective. Earlier work found that families wanted to be confident that their data would be handled sensitively and fairly only by those with a need to know, and believed that greater visibility of data processing would allow them to trust that that was the case (4.2.2 and (Bowyer *et al.*, 2018)). When families felt alienated from their data, trust was absent (4.4.1), and a lack of transparency and accountability makes it hard for families to maintain trust in the system (4.5). These issues of transparency and trust are inherent in a data-centric operating model, and the World Economic Forum have summarised this problem thus:

“A crisis of trust is developing, stemming from the use of personal data in ways that are inconsistent with individuals’ preferences or expectations.” (Hoffman, 2014b)

Information facilitates trust; transparency therefore offers an opportunity to earn trust & improve relations. In both studies, the findings led us to conclude that increased transparency from data holders / service providers would improve trust; in Case Study One we concluded that support workers and organisations should be as open as possible about data handling and sharing (4.3.4.3), while in Case Study Two we highlighted the potential benefits of increased consumer loyalty that greater transparency might bring (5.5.2), as

well as the need for policymakers to legislate in favour of increasing individuals' understanding of data practices (5.5.1). In doing so, we are recommending a level of transparency that goes beyond current GDPR practice, and even beyond current GDPR policy; in order to redistribute power, GDPR needs to deliver meaningful transparency, not just the "*box ticking*" delivery of unhelpful files that our participants sometimes observed (5.5.1). Crabtree describes meaningful transparency by saying that it cannot be a "*one-way street*" that reduces individuals to "*being spectators*" on how their data is used; he says that it involves "*making the whole ecosystem transparent, not just the front end*" (Crabtree and Mortier, 2016). Access to good information about practices is the most effective way to earn trust (5.4.4.2), and both studies' findings suggest that a proactive attitude can do just that. (4.4.1, 4.3.4.2, 5.5.2, 5.6).

Initially, transparency may cause distrust, but only where practice is problematic; this is accountability becoming real and catalysing better data practices. It is important to note that in Case Study Two we saw the transparency of GDPR cause elevated distrust (5.3.4), however this does not mean it should be avoided. The reasons cited for distrust arising were invariably due to the discovery of practices that participants did not approve of. This is a clear illustration of the link between transparency and accountability; the transparency reveals the non-consensual or unsatisfactory practices that providers must change if they wish to maintain trust and loyalty, such as unclear data practices, data over-use or data sharing that the individual would not have consented to had they been asked. This shows that in some cases trust is fragile, where unfavourable practices are hidden and only the individual's unawareness is keeping the relationship intact (5.4.4). Data holders should not only be transparent, but should follow this up by acting upon subsequent feedback, improving practices that individuals discover and challenge (5.4.4.3). By shining a light, accountability becomes real and change for the better can occur. Ultimately, increasing transparency can help providers uncover exactly what they need to do to earn greater trust (5.6).

Without transparency of data and processes, individual action is blocked and power remains imbalanced. Across the two studies, a clear pattern emerges: transparency can increase trust, enable accountability, empower individuals, and (provided organisations respond favourably) actually tilt the power balance back toward an equitable and fair relationship where data is collected and used in clear sight of the individuals it concerns, where they might hold those organisations to account and immediately challenge any unsatisfactory practice, unauthorised processing/sharing or inaccurate data. Thanks to GDPR, individuals are now able to take direct action to educate themselves and pursue greater transparency, and utilise their rights to motivate incremental changes from data holders (5.5.3). Without transparency, data holders will continue to hold the balance of power, and individuals will lack agency and accountability.

6.2.2 Individual Oversight

Data visibility and process transparency naturally leads to a desire for individual oversight. If you see something that is ‘not right’, you are motivated to want to fix it. And therefore, people want something more than data and process transparency, the natural next step is the ability to make decisions about what happens to their data. Participants’ goals in Case Study Two included curiosity, suspicion and a desire to shed light on specific incidents (5.3.3), mirroring the desires families in (Bowyer *et al.*, 2018) exhibited to be able to know and see what data was held and used about them. In both cases, individuals wanted to have a say over what happens. Current models of informed consent have been found to be inadequate, with the initial handover of data acting as a ‘*point of severance*’ (Luger and Rodden, 2013). This was echoed in the experiences of families in the Early Help context, who gave consent at the point of initial onboarding, but lost all ability to influence what happens to their data thereafter [Bowyer *et al.* (2018); 4.2.2]. GDPR aims to adhere more to a *dynamic consent* model (Kaye *et al.*, 2015; Williams *et al.*, 2015) by giving people an ongoing set of rights, including the right to be informed about the use of your data, the right to object to certain data uses, and the right to get your data corrected or deleted (Information Commissioner’s Office, 2018). In line with the *accountability principle* (Article 29 Data Protection Working Party, 2010; Crabtree and Mortier, 2016), this in effect would allow people to act as overseers or regulators over their own data: watching how it is used, and demanding action or change to practice when they see data use that goes against their wishes.

People need agency and negotiability over held data about them, in order to ensure fairness and accuracy and reduce risk. As my earlier work in the Early Help context (Bowyer *et al.*, 2018) showed, there is a strong desire to ensure data is *fair* and *accurate*, because that data is used to inform judgements and make decisions that can directly affect the individuals concerned. Data in Case Study Two showed a clear problem with the accuracy of unseen data: while in 92% of cases volunteered data (which by definition, has been seen by the individual) was found accurate, derived and acquired data (previously unseen by the individual) was found inaccurate in 50% and 80% of cases respectively (5.3.2). Being able to ensure fair and accurate data goes beyond being able to see and understand the data, but requires also *agency* (the ability to act within a data system, such as to delete or correct data, or withdraw consent) and *negotiability* (the ability to continue to have a voice and make changes as circumstances change) (Mortier *et al.*, 2014). People need a relationship with their data (2.1.5). In both Case Studies and in earlier work, individuals perceived tangible risks both of data being held beyond their reach, but also of potentially inaccurate data being used to make decisions. Risk factors identified in the Early Help context included facilitating or encouraging crime, causing social and psychological harm, and enabling medical mismanagement or welfare support failures (Bowyer *et al.*, 2018). In Case Study Two, participants felt that held

data about them that is not visible or controllable was a liability that might lead to privacy violations, commercial exploitation, and an increased risk of data leaks (5.4.3.3). Clearly people feel that for their data to be safe, they must be able to see and verify its storage and use for themselves and enforce action when something is not right. As early as 1980, when the world was less data-centric, it was already recognised that individuals would need the ability to challenge data use, as the OECD observed in their guidelines:

“The right of individuals to access and challenge personal data is generally regarded as perhaps the most important privacy protection safeguard.” (Organisation for Economic Co-operation and Development, 1980)

Individual oversight capabilities must be supported by governance, so that individuals can effect desired changes. As Gurstein notes, a key element of effective data access is governance, that is, mandating data holders to support individuals in accessing their data and respecting their wishes over what should happen to that data (Gurstein, 2011). Individuals need to be able to give instructions, make changes and express permissions that have weight; they need to be listened to, so that they can meaningfully effect change (4.4.2). Bakardjieva, examining the use of data about others in a different context (research), identified that individuals whose data is used need the ability to influence not only the data about them, but the actual decision-making that occurs based on that data: both the data *and* the decision-making should become objects that the individual subject can manipulate (Bakardjieva and Feenberg, 2001). At the time of writing (March 2022), much of the focus on GDPR has been about access to data, perhaps because this is more tangible, and very little about GDPR’s other rights that can influence decision-making (5.1.2). This was backed up by participant experiences in Case Study Two, where desires to influence or change practices or delete data were either not actionable or ineffective (5.4.3.3). Governance over individual data rights has two elements. First, to support individuals in complaints or challenges, which are currently unevenly enforced (5.5.1). But more importantly than this, given the extensive use of data by organisations and the great potential for misuse or harm, individuals need to be able to trust that systems are in place that mandate the behaviour of data holders to be trustworthy; to compel organisations to maintain good data practices such as data security and dynamic consent in the first place (5.4.4.1). In the GDPR context bodies that can do this already exist - the Data Protection Authorities. In the public sector/care context, the picture is less clear. Participants identified a need for oversight bodies to compel good practice, identify appropriate access rules, and to provide independent oversight in contentious cases (4.3.4.3); this is particularly difficult given that no organisation can see the full picture of an individual’s civic data.

Individual oversight would bring individuals back to the centre of their personal data ecosystem as an active participant. As outlined in 2.3.4, the ideal has been established that individuals need to be at the centre of their own

personal data ecosystem, overseeing and controlling their data selves as easily as their physical selves. Currently, as seen in both contexts, data functions as a proxy for their direct involvement [Bowyer *et al.* (2018); 5.4.4.1]. Decisionmakers consult data first, as the primary source of truth (4.1.2, 2.1.2), and the individual second (if at all). For transparency to be meaningful, data flows need to open up to include individuals as part of the loop, changing them from passive spectators to active participants (Hoffman, 2014a; Crabtree and Mortier, 2016) in the processing of their data. Examples of specific oversight abilities desired in Case Study One were the ability to explain or annotate datapoints (4.3.3.4), to be able to check data together with support workers, with the record of that check becoming part of the data (4.3.3.2), or to have granular access controls over precisely which data could be seen by whom (4.3.3.5). In Case Study Two, a clear picture emerged that what participants want is the ability to make choices. They want control over the data they are forced to sacrifice to companies (5.4.4.1); to avoid the '*point of severance*' Luger describes, data sacrifice should be a loan or sublicense, not a taking-possession-of.

Given the changing and complex nature of human life, data is inadequate and consent is never complete, so longitudinal participation and oversight is needed. Too often, data is treated as a static source of truth (see above). Attempting to represent people as data in order to require less human contact is a reasonable goal from an organisational efficiency or cost-saving perspective, but any representation will never be complete or adequate (Cornford, Baines and Wilson, 2013; Bowyer *et al.*, 2018). In Case Study One, the findings showed the need for numerous efforts to augment data in order to combat its inherent inadequacy, such as support workers seeking to understand the people behind the records (4.3.3.1) and maintaining a constant attitude of seeking to understand more deeply than the data record can allow (4.3.4.2). Even if a data record can be corrected or completed, it will still be inadequate, because human lives change continuously: people move, start and end relationships and jobs, marry, divorce, have children, pursue new interests, become incapacitated, or die. The passage of time can radically change the context or relevance of data (Bowyer *et al.*, 2018). A one-time informed consent upon data collection is inadequate in this ever-changing context (4.2.2, 4.2.6, 4.3.3.3). And of course, if consent needs to be ongoing (and in order for it to be meaningful) this means that engagement with the individual concerned, and that individual having a view of their data and its use, need to be ongoing too. In order to avoid storing or using data beyond its need, ongoing data access is needed, in order to enable ongoing individual oversight (4.2.2, 4.3.2.4). Systems and processes must treat data as dynamic (4.3.3.3), as something that will become inaccurate without sustained engagement. In Case Study Two, we note that GDPR data access is currently based around viewing a one-time snapshot of your data, and does not take this need for negotiability (Mortier *et al.*, 2014) into account at all (5.5.1) (though some companies now offer download dashboards that come closer to providing ongoing access). Ongoing access, consent and participation do carry cost implications for providers, and effort implications for individuals - but these

can be improved over time: the former through automation, standards and education, and the latter through holistic approaches to personal data ecosystems; these mitigations will be explored further in Chapter 7.

There is scant individual oversight available today. Governance is lacking. If people cannot make choices about their data, they will remain powerless. Participants in both contexts faced an inability to see the full picture of how their data is processed and used. Despite families in Case Study One workshops spending time designing interfaces for seeing and correcting their data and changing permissions (3.5.3, 4.3.2.1, 4.3.2.3), no such interfaces exist. The entirety of their data access and influence is limited to what can be achieved verbally with their support workers (4.2.4, 4.4.3). Without transparency and dynamic consent mechanisms, those families lack accountability. They are excluded with no ability to oversee or participate in the life of their data. In Case Study Two, of the 41% of participant goals that concerned gaining insight into and control over the use of their data, 66% were unmet. Participants reported seeing no clear pathway on how to access rights to control their data and only 1 of the 10 cases where a participant wanted to delete their data was successful (5.3.3). At the time of writing only one company, Apple, has a privacy hub that offers clear routes to access data rights other than access. Participants also reported in some cases being unable to check the accuracy of their data, or to investigate specific incidents where they had concerns (5.3.3). The general view was one of widespread disappointment, that despite the promise of GDPR it did not confer any power to the individual to influence data use. (5.2.4.3), leading in some cases to a reluctance to submit GDPR requests in future. Access requests were also rarely seen as useful in the care context, and our understanding is they typically only occur in the case of complaints. In the GDPR context, the inability to restrict data use or delete data was seen as a lack of control, and the retaining of data against their wishes as a liability (5.4.3.3). Ultimately oversight means having choices, which is essential in the data-centric world. The case studies' findings show that, in general, participants felt they had been forced to sacrifice data to access services, and offered *no* practical choices or control over that data. Without individual oversight, there is no choice and people remain powerless.

6.2.3 Involvement

Data represents people. But people are more than can be encapsulated in records. There is a need to engage the human behind the data, as people can never be fully represented in data. Intrinsic in the move towards data-centricity has been a move away from human involvement. In the commercial sector this is due to cost-saving (call centres and web portals being cheaper than individual customer interactions) (2.1.2). In the care context it is similar but there is also a desire to create a society that functions at large without individuals requiring special handling and support (3.4.1). Both case studies' findings, consistent with literature (Abiteboul, André and Kaplan,

2015; Crabtree and Mortier, 2016), reveal myriad problems created by the exclusion (Bowyer *et al.*, 2018) of people from matters that affect them - from feelings of alienation or disengagement (5.4.4.1) to actual harms caused by erroneous or unfair judgements (4.2.2). Service providers holding data need to contextualise data as an incomplete view into the complex human world, and seek greater understanding (4.3.4.2) while looking for positives in data (4.3.4.1). It is interesting to note that the Troubled Families programme was created help find the human situations of people slipping through the cracks of the system, which highlights the inadequacy of purely data-based decision-making (3.4.1).

Consent to access and use data needs to be dynamic and meaningful, which can only happen through ongoing involvement. As established in 4.4.1 and above in 6.2.2, ongoing data consent is essential, and this is especially important where that data is used to make decisions (4.4.3). One-time consent is ineffective and meaningless (4.5). Asking individuals for consent and subsequently less involved in decision-making reinforces a hierarchical, rather than an equitable, power relationship, as Bakardjieva and Feenberg found in their work looking at how to involve virtual subjects in research (Bakardjieva and Feenberg, 2001). Without ongoing consent, the power imbalance is amplified (4.5). In the commercial context, companies view data as their asset to exploit (Hoffman, 2011; Toonders, 2014), and the simple fact of having the ability to collect or access to data about people has proved in practice, sufficient to enable a variety of practices which would be likely to be refused consent if made visible to users (Melendez and Pasternack, 2019; Evans, 2021; Claburn, 2021). Individuals feel forced into a one-sided arrangement of sacrificing data in exchange for service benefits; with no choice upfront on signup, and minimal practical choices afterwards, their only choice is Hobson's choice ('Hobson's Choice', no date). Consent has become commoditised, and from a corporate perspective the focus has become constructing a legal justification for using an individual's data rather than practically engaging with them and verifying if they approve (Woods and Böhme, 2022). This can only happen when the individual about which data is held is excluded from data handling processes.

A human channel for conversation is wanted, to enable explanations, questions, and consultation. In my earlier work (Bowyer *et al.*, 2018) and in Case Study One, all participants viewed that individuals should be able to talk to someone about their data (4.2.6, 4.3.2.4), in order to ask questions or explain datapoints. In Case Study Two, participants had questions about their data that they wanted to answer (5.3.3), yet these questions remained unanswered (5.4.2.3). Participants regularly experienced painful and ineffective processes when trying to answer their questions, found that GDPR responses, often unhelpful, provided no backchannel for followup questions or further communication (5.2.4.3). They were left '*in the dark*'. This highlights the need for a human support channel, which is not mandated by GDPR, not just to better understand the data itself (6.1.2), but to enable ongoing consent negotiation within the relationship with data holders.

Individuals should be consulted in decision-making. This improves accuracy, perspective, and fairness and reduces consent liability. A common theme in the findings of both my earlier work with families and Case Study One was the idea that data is more likely to be fair and accurate if in the individual has the opportunity to express their perspective on it. Families advocated checking data together to identify gaps (4.3.3.2), valued the prospect of making their own contributions to the data to '*tell their own story*' (4.3.3.4) and wanted a '*right to explain*' or annotate their data (Bowyer *et al.*, 2018). Support workers recognised the need to work with families to understand their situation better (4.3.2.2). These findings show it is important to give the human data subject a role in data creation or checking, and that if they are involved involved, to ensure a fairer and more complete view can be obtained than the limited view presented by the data record (4.3.3.1). Without involvement, agency (Mortier *et al.*, 2014) will always be limited. Data is not neutral (Gitelman, 2013; Neff, 2013), and this means all stakeholders should be given a role (Bowker, 2005) in order to avoid errors, harm or disempowerment (4.4.1). While involvement is negligible in the care context (being limited to the TAF and the support worker relationship, in the commercial context, there is no such involvement, and the results of this can be seen in the low accuracy of previously unseen data (5.3.3) and in the low trust ratings (5.3.4) given to many providers after seeing data returns. It seems that data and knowledge would become more accurate when it is closer the data is to the individuals concerned (4.4.3, 4.5); data created and handled far from the family would intuitively be less accurate and less likely to have been considered from their perspective (4.2.2). Organisations on both sides value data accuracy (4.2.3, 2.1.2) so greater involvement could help achieve this. Data holders would also benefit from involving individuals because responsibility for consent would become shared, resulting in a higher '*buy-in*' from individuals and a reduced liability; provided communication is effective and without barriers, involved individuals would be inclined to speak up if they see something they do not agree with (4.4.1). Of course, this call for involvement in decision-making is not absolute. Businesses need to be free to exercise their expertise. Patients are not best placed to decide what medicines they need, and over-involvement of users in product design can result in mediocre products that suffer from '*design by committee*'. In fact, the attitude conveyed toward the individual is critical (as borne out in our Case Study Two findings (5.4.4.2, 5.4.4.3); Edwards and Elwyn, in their paper on shared decision-making, argue that *feeling* involved is actually more important than actually contributing to decisions (Edwards and Elwyn, 2006).

Effective collaboration can be achieved by coming together around the data, using it as evidence (of facts or of opinions) and as a boundary object. Workshop C in Case Study One specifically explored the prospects of shared data interaction, which had emerged from SILVER and phase 1 (Workshop A/B, see 4.2.5) findings as a possible way to meet the needs of both supported families and support workers. Shared values were identified (4.2.6) and a model for shared data interaction builds upon the findings (4.4.3). Central to this

is the idea that evidence-based decision-making can be more effective (both from a rapport-building perspective but also in terms of the quality of decision made) as guidance has advised [OFSTED (2015);Department for Education (2018);4.1.2] and as seen in the findings (4.3.2.2, 4.3.3, 4.4.1). Families and staff saw potential benefits from checking data together (4.3.3.2), using specific datapoints in discussions as evidence (4.3.2.2), or to help families open up (4.4.2), on top of the simple benefits of making all evidence equally visible (4.3.2.2, 6.1.1). Throughout such shared data interactions, representations of data perform an important function as a boundary object (Star, 2010; Bowker *et al.*, 2015); it provides a common focus to discussions that is relatable to both parties, “*things to think with*” (Papert, 1980; Brandt and Messeter, 2004). This helps improve legibility by surfacing the (perfectly valid) differences in different parties’ perspectives (6.1.2, Mortier *et al.*, 2014). Working with families in 2017 I observed that gathering around data representations facilitated a less confrontational interaction than an across-the-table interview would (Bowyer *et al.*, 2018).

Being involved means being able to learn and take action at any time, including on one’s own and away from official contact or interactions with service representatives. Both Case Studies reveal that people want an ongoing window into their data and its use. Limiting data access and process transparency to specific times spent with a gatekeeper, as in the Early Help case (4.1.1), or to a process that only provides a one-time snapshot of data (5.5.1) reinforces the hierarchical power imbalance. Giving access to data and abilities to explore and ask questions “*in your own time*” (4.3.2.4) can unlock new individual benefits from data (5.4.3.2, 6.1.3) and provide more immediate feedback that can help people measure their progress towards improvement goals (2.2.3, 4.4.2, 4.4.3), as well as enabling ongoing individual oversight (4.4.2) and dynamic consent (4.3.4.3). In the face of providers making decisions based on unseen data using processes that cannot be observed, people feel excluded and powerless (5.4.4.1, 6.1.1). To ensure effective access (Gurstein, 2011), accountability and trust, there must be ongoing involvement, data access and transparency (5.5.1, 6.2.2).

Indirect data use enforces an uneasy trust; services that use data need a human face or point of contact, in order to grow understanding, earn trust and improve relations. There is a coldness to data. Facts, judgments and mistakes appear in print with equal weight, without explanation or context, seemingly a complete set of objective facts. In both studies, the value of human contact accompanying data access was evident; in Case Study Two participants valued the GDPR responses that felt most human (5.4.4.3) and lamented the inability to discuss data or resolve questions (5.3.3, 5.4.2.3). In (Bowyer *et al.*, 2018) participants wanted ‘*to have a conversation*’ about their data and in Case Study One participant data interface designs included buttons to chat to their support worker or ask questions (4.3.2.3). People do not want to be severed or alienated from their data, yet they feel they have no choice but to relinquish access and involvement (4.4.2, 5.4.4.1). Ideally, all data-using

organisations would have a human face or point of contact that individuals can address questions to and in whom their trust of the organisation can be embodied (4.4.2).

Without involvement, people cannot take a full and equitable role in processes that affect their life. When data is used by organisations, this inherently serves as a proxy for their involvement [Bowyer *et al.* (2018); 5.4.4.1]. People have consequently lost control and agency (Crabtree and Mortier, 2016), creating a crisis of trust and a power imbalance (Hoffman, 2014a, 2014b). Exclusion from data handling and decision-making inevitably reinforces a hierarchy, with the individual destined to have less say and influence over services that affect their daily lives. Taken together, an ongoing involvement with the data held by service providers and the processes that use that data would change indirect data interaction (6.2) into direct data interaction (6.1) and transform the individual experience of service use. Of course such a shift is difficult and costly to offer, for both small and large companies – the data-centric world emerged in part as a means to reduce costly human interactions and facilitate large-scale scaling up to serve larger user bases – but what we see suggests that the balance has tipped too far towards exclusion of individuals, and that giving them even a limited role in reviewing data, consenting and contributing to decisions, and just being informed, could carry significant benefits for both individual empowerment and organisational reputation, and trust.

6.3 Achieving Individual Empowerment

Empowerment means being able to do more. Empowerment is defined as “*the process of gaining freedom and power to do what you want or to control what happens to you*” (Cambridge Dictionary, no date). Conceptually, in line with the constructivist, pragmatist ontology outlined in 3.1, power can be thought of in line with this definition, best judged by the individual’s consideration of the question “*what can I do?*”. It is important to acknowledge that the power imbalance described by the World Economic Forum (Hoffman, 2010, 2011, 2013, 2014a, 2014b) and described in 2.1.2 encompasses both this concept of ‘*power to*’ but also the concept of ‘*power over*’. However, as argued by Pansardi, these are best considered as two sides of the same coin; two perspectives on the same set of facts: organisations have ‘*social power*’ over individuals, that affects both what those organisations can do, and what the individuals served by those organisations can do. An organisation having social power does not imply a desire to dominate; impacts on the individual’s agency are merely material effects of the power holder’s attempts to influence the behaviour of individuals in pursuit of the power holder’s own desired outcomes (Pansardi, 2012).

The more that the six data wants can be addressed (especially transparency), the more empowering it is to individuals. Through the above exploration of the six individual wants – visible, understandable, usable data paired with process transparency, individual oversight and decision-making involvement – exemplified through the Case Studies, we can see that the status quo

of the data-centric world (2.1) is that lack of these capabilities reduces individual capacity to act, individuals are *disempowered*. People are seen indirectly through data (Cornford, Baines and Wilson, 2013); they are excluded and not involved (4.5, 5.4.2.1, 5.4.3.3, 5.4.4.1). While in both Case Studies many of the findings are based on the opinions of participants as to what they believe *would* be desirable, preferable or more successful, in the case of data visibility (6.1.3) and process transparency (6.2.1) we have clear evidence that it *does* have an impact on individual's sense of power; in Case Study Two, 45% of cases saw people experience a change in power after examining privacy policies, experiencing GDPR and scrutinising responses. 29% felt a decrease in power, and 17% an increase. Notable also is the impact on trust that transparency caused - with 52% feeling more distrustful, having discovered unsatisfactory exertions of power by providers, and only 14% being sufficient reassured to increase their trust (5.3.4). It seems that data visibility and process transparency are a key first steps to shifting the power balance, but that this should be followed soon after with understanding (6.1.1), agency (6.1.3, 6.2.2) and involvement (6.2.3). While any one of the six of these data wants can be pursued to produce improvement in its own right, the combination of all six is likely to produce more than the sum of its parts, an empowered form of digital citizen.

These six wants form the blueprint of a new role for digitally empowered citizens. Giving people a role in influencing the life of their own data is a key ingredient in and of more progressive digital citizenship (Bridle, 2016). Shifting data interaction interfaces and processes to a more human-centric [2.3] model where people are controllers at the centre of their own personal data ecosystem (2.3.4) would be progressive and transformative, and not without cost, education, deployment and uptake challenges, but, as the families in Case Study One found, it is not hard to imagine data interfaces and service provider interactions that are more empowering and human-centric than the status quo - some companies targeted in Case Study Two do now offer extensive privacy hubs and explanations of data practices, while others do now offer interfaces solely for accessing data and exerting data rights, and while those approaches do not satisfy all of the data wants described in this chapter, they do show that organisations can choose to invest in involving people more with their data. In Case Study One participants recognised that it will take proactive action, not just by frontline workers but system designers and management policymakers, to challenge the prevalence of data-centric norms (4.3.4.3). If the locus of decision-making could be shifted towards supported individuals, this would give them a role to play as agents in the life of their data, allowing them to curate their own data self, the representation of them seen by the state, so that it is fair, accurate and representative [Bowyer *et al.* (2018); 4.4.3]. The principles of involvement, effective access and shared data interaction could be applied in many domains - education, health, democracy and commerce, and an emphasis on individual sociotechnical capabilities is a useful mindset to apply to both business process design and data interface design.

Empowering individuals with data should lead to a better future.

Taken together, the pursuit of the six data wants allow us to envisage a new, fully data-empowered future for individuals, who would reap the benefits of being able to gain insights and feedback from their own data in real time, while also co-operating with service providers in the stewardship of their data and involved in decisions. They would have agency, influence and negotiability in an ongoing manner. In this future, there are opportunities for organisations to reimagine customer relations and the role of data in service provision, leading to increased accuracy and consent, reduced liability, greater trust and loyalty (5.5.2), and potentially to capitalise on new demands for data insight tools (5.5.3). The ideal is that individuals would have a “*permanently open two-way communication channel [with data holders], allowing a dialogue to take place*” (Bakardjieva and Feenberg, 2001), delivering dynamic consent and ongoing involvement. Such a change in thinking, away from legal box-ticking (be it capturing informed consent or satisfying GDPR requests) and towards delivering an ongoing understanding of the collection and use of one’s data and a voice in what happens to it, would empower individuals and begin to move us towards that idealised future. In such a society, the positive impacts of transparency would be evident. The general principle that transparency allows trust to be earned (which can in turn benefit organisations themselves) could be applied in a wide range of service endeavours that are currently very closed. This would allow innovation in data access and process involvement to flourish.

7 Discussion II: Designing and Pursuing Better Human Data Relations

“*Civilizations advance not by the technology they know about, but by the technology they don’t have to know about.*” – Anonymous proverb

7.1 Introduction & Background

Through the Case Studies (Chapter 4 & 5) and the discussion in Chapter 6, a clear understanding of what people want from direct and indirect data relations (RQ1 & RQ2) has been established. In this chapter, we turn our attention from theory to practice, from what is needed to *what is possible*. Specifically, this chapter will return to the overall research question “*What relationship do people need with their personal data, and how might that be achieved?*” and look specifically at its second clause. This chapter describes practical approaches for future research and innovation, in a way that is deliberately broad and shallow, from the perspective that it is more useful to introduce a wide range of applicable ideas than to go into great detail on just a few. This is not intended to form a complete or exhaustive roadmap; it is a snapshot of ongoing work, identified challenges and known opportunities, forming an *anthology of reference material* for designers and innovators in this space. These ideas are illustrated through real world insights and activities from the four industrial and academic research projects I was part of, and from the work of other innovators and activists. This

chapter also builds upon the theoretical insights from the Case Studies in order to inform the design of future research, innovation and policy as to how the better Human Data Relations conceived in this thesis thus far might be achieved.

The approach this chapter takes is to name and illustrate what challenges and opportunities are relevant when attempting to bring about changes in the world that would bring people closer to the six HDR wants that this thesis has uncovered. There are many aspects to such a wide-reaching objective: technical, design, commercial, legal, moral, social and political and this chapter does not pretend to cover them all nor to be formal empirical research. However I have been fortunate to have undertaken, during the same time period as I have been working on this PhD but outside of the research, direct embedded work in personal data interaction related projects (3.4.3) in academic and industrial research that directly contribute to the question of how to bring about better human data relations in practice. As a result, some of these challenges and opportunities herein are described in greater detail than others, corresponding only to my proximity and depth of engagement with those ideas rather than their relative merit, complexity or impact potential.

In section 7.1.1 the external activities I undertook are described; they form a primary point of reference for insights and illustrations shared in this chapter, as they have allowed me to learn enough to provide a useful overview and highlight many important and evolving areas where different actors are trying to bring about changes that often align well to the six data wants uncovered in the previous chapters.

In section 7.1.2, I explain some important context about the nature of the ideas presented in this chapter and how to attribute them fairly.

In section 7.1.3, I introduce some additional background on *Theories of Change (ToC)*, which are used as a framing device for structuring the insights described in the main body of this chapter into a series of different possible trajectories for change.

In section 7.1.4, I consider the researcher-turned-activist stance that drives this chapter, framing the pursuit of better HDR as a *recursive public*.

In section 7.2, to provide deeper context for what follows, the concept of HDR is expanded to identify some additional insights into how people relate to data, and an important dichotomy of two distinct drivers that motivate people's needs for better relations with their data.

Section 7.3 and 7.4 form the main body of this chapter, with obstacles and insights being detailed in section 7.3 and specific opportunities into how better Human Data Relations can be pursued in practice described in 7.4. 7.4 is structured using the ToC framing described in 7.1.3, as a series of named opportunities fitting into each trajectory of change.

Section 7.5 concludes the thesis, summarising the change trajectories presented in 7.4, the thesis' contributions as a whole, and answering the overall research

question.

7.1.1 External Research Settings

[TODO Move 3.4.3 etc. to here and remove all refs to 3.4.3]

The majority of examples and learnings shared in this chapter come from my participation as an expert researcher and designer in two industrial research projects:

1. **BBC R&D's Cornmarket Project**, which explored through user experience design, technical prototyping and participatory research, how individuals might interact with data through a Personal Data Store interface (see 3.4.3.3)
2. **Sitra/Hestia.ai's #digipower Project**, a successor to Case Study Two, in which European politicians examined companies' data practices through exercising data rights and conducting technical audits (see 3.4.3.4)

In addition, my participation as an interface designer and front-end software developer in the following two academic research projects contributes secondarily to this chapter:

3. **Connected Health Cities (CHC)'s SILVER Project**, where I, along with a backend developer and a team of researchers, developed a prototype health data viewing interface for Early Help support workers (see 3.4.3.1).
4. **Digital Economy Research Centre (DERC)'s Healthy Eating Web Augmentation Project**, which explored the use of web augmentation techniques to modify the user interface of takeaway service Just Eat to insert health information, in support of healthy eating (see 3.4.3.2).

7.1.2 Attribution of Insights

While this thesis is my own original work, and many ideas presented in this chapter are fully original, some of the specific details, theories and ideas presented in this chapter arose or were developed or augmented through my close collaboration, discussion and ideation with other researchers, including:

- Jasmine Cox, Suzanne Clarke, Tim Broom, Rhianne Jones, Alex Ballantyne and others at BBC R&D;
- Paul-Olivier Dehaye, Jessica Pidoux, Francois at Hestia.ai;
- Stuart Wheater of Arjuna Technologies and Kyle Montague of Open Lab during the SILVER project; and
- Louis Goffe of Open Lab on the DERC Healthy Eating project
- earlier innovation work with Alistair Croll at Rednod, Montréal, Canada (circa 2011) and with Megan Beynon at IBM Hursley, UK (circa 2006).

Due to these collaborations and the ongoing and parallel nature of many of these projects to my PhD research, it is impossible to precisely delineate the origin of each idea or insight. In practice, ideas from my developing thesis and own

thinking informed the projects' trajectories and thinking, and vice-versa. These ideas would not have emerged in this form without my participation, so they are not the sole intellectual property of others, but equally I would not have reached the same conclusions alone, so the ideas are not solely my own either. All diagrams and illustrations were produced by me, except where specified, and the overall synthesis and framing presented in this chapter is my own original work. Where this chapter includes material from the four projects, that material is either already public, or permission has been obtained from the corresponding project teams.

7.1.3 Theories of Change

To provide a structure for cataloguing the insights conveyed by this chapter, I use a *Theory of Change (ToC)* framing. ToC is a set of methodologies commonly used by philanthropists, educators and those trying to improve the lives of disadvantaged populations (Brest, 2010); the theories can be used in different ways including planning, participatory design and field evaluation of the effectiveness of new initiatives. There are many different implementations, but common to most of them is a focus on explicitly mapping out desired outcomes (Taplin and Clark, 2012) with a clear focus on who is acting and whether the change being brought about is a change in action, or a change in thinking (Es, Guijt and Vogel, 2015). In this chapter, ToC theory will be used in a very limited way, not as a methodology but simply to provide a structural frame for proposed changes, as described below. Using ToC to perform evaluation of the effectiveness of proposed change approaches in action in society would be well beyond the scope of this thesis. Nonetheless, this frame is a useful way to map out the different approaches to changing the world in pursuit of the ideal of better HDR.

Figure 29 illustrates the aspects of ToC thinking that section 7.4 will use as its frame. Specifically, desired changes can be broken down into:

- **Internal changes:** changes in thinking, feeling, reasoning, understanding, attitudes or identity.
- **External changes:** changes in actions, behaviour, interactions, structure, policy, technological capability, processes and the external environment.

At the same time, desired changes can be broken down into:

- **Individual changes:** changes to individual thought or actions
- **Collective changes:** changes to the thoughts or actions of groups of people together, or to the systems, practices and norms of society at large.

These two splits produce four dimensions of change, and form four quadrants representing different types of change, which are shown in Figure 29 and described

¹⁷Diagram used here unchanged from *Hivos ToC Guidelines* (Es, Guijt and Vogel, 2015, p. p90) under a CC-BY-NC-SA 3.0 license, whose authors state that this diagram was adapted from earlier work by Wilber (1996), Keystone (2008) and Retolaza (2010, 2012).

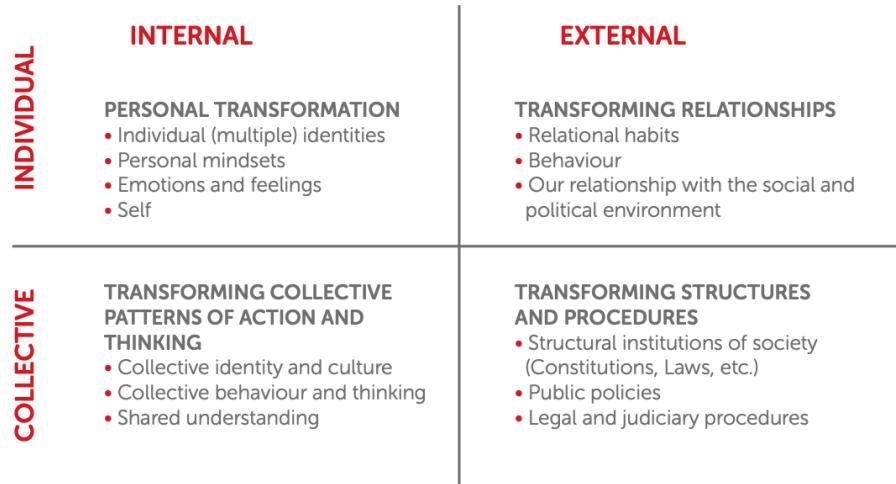


Figure 29: Figure 29: The Four Dimensions of Change¹⁷

here:

- ***Individual/Internal (II)***: This top-left quadrant represents changes to what individuals know and understand, and to how they think, feel and plan to take action.
- ***Individual/External (IE)***: This top-right quadrant represents changes to how individuals' relationships with others; acting (or being enabled to act) differently in their daily lives and when interacting within society.
- ***Collective/Internal (CI)***: This bottom-left quadrant represents changes in the shared knowledge of groups of people or to the collective identity or values of social groups.
- ***Collective/External (CE)***: This bottom-right quadrant represents changes to the structures and procedures within which people operate, including technology, law, societal norms and communications.

Key to ToC thinking is the idea that making changes in one quadrant can stimulate change in others; for example, collective learning about data attitudes and practices, such as the research conducted in this PhD, (lower left quadrant) could inform the design of new technologies, interfaces or processes (lower right quadrant), which if built could make new structures available to have an impact on improving individual-provider relationships (upper-right quadrant). The changes to those relationships could then in turn lead to individuals thinking and feeling differently (upper left quadrant), for example feeling more empowered or having greater awareness of data practices.

7.1.4 Better Human Data Relations as a Recursive Public

Before engaging with the practicalities of pursuing change, it is valuable to revisit the stance from which we approach this change. As outlined in 3.2, the research of this PhD has been grounded in participatory action research and experience-centred design; by using a Digital Civics (Vlachokyriakos *et al.*, 2016) frame to gain deep understanding of people's needs and the ways those needs are not fully met, we can see how the world needs to change. Section 3.2 already outlined that we can consider such research as political, seeking to correct an imbalance in the world. In this chapter, we look beyond identifying what change is needed, and step into the role of activist, exploring how individuals and groups can actually change the world they inhabit.

In doing so, we can consider ourselves (those who pursue better Human Data Relations, or *HDR reformers* as a shorthand) as a *recursive public* (Kelty, 2008; 'Recursive Public (Discussion Page)', no date), albeit a nascent one. This is a term originating in the free software movement to describe a "*collective, independent of other forms of constituted power, capable of speaking to existing forms of power through the production of actually existing alternatives*". This term captures the idea that through various means at our disposal: participatory research, experience-centred design, engineering software prototypes, exertion of legal rights, and efforts to raise public awareness, we seek to modify the systems and practices we live within in pursuit of our goals. This collective around better Human Data Relations does not yet exist as a named and identifiable *public* (Le Dantec, 2016) but its members congregate around emergent collectives in interconnected and overlapping spaces, most notably the MyData community (MyData, 2017) and its members, but also research and activism agendas including but not limited to: digital rights ('Open rights group: Who we are', no date), gig economy worker rights (Kirven, 2018), privacy by design (Cavoukian, 2010), data justice (Taylor, 2017; Crivellaro *et al.*, 2019), critical algorithm studies (Gillespie and Seaver, 2016), humane technology (Harris, 2013a) and explainable AI ('Explainable AI: Making machines understandable for humans', no date).

Whether these disparate groups coalesce into a single identifiable public remains to be seen, and so too whether the term this thesis offers of *Human Data Relations* is sufficient to capture that public (at least, it provides a descriptive umbrella term). Nonetheless, the breadth of research and innovation and activism happening in this space validates both the need and the desire for such a recursive public around better HDR to exist. Therefore, this chapter takes an unashamedly critical view of the status quo, favouring disruptive societal changes that would further the objectives of better Human Data Relations and providing actionable approaches that will be of use to the members of this public. The chapter asks, "How can we change the world into the one we want?"

7.2 Refining and Defining ‘Human Data Relations’ (HDR)

Chapter 6 established six ‘wants’ in HDR: **visible**, **understandable** and **usable data**; **process transparency**, **individual oversight** and **decision-making involvement**. At a simplistic level therefore ‘better’ HDR can be achieved by working to improve upon those six aspects of data interaction. However, as this section will explain, HDR can be conceptually split into two distinct motives, to which those six wants apply differently, therefore it is useful to develop the concept of HDR further. As background understanding for this duality of motivation, it is first necessary to examine more closely what role data plays in people’s lives.

7.2.1 The Role of Personal Data

In the modern world, where almost anything can be encoded as data, and given many previously analogue objects and activities now have digital equivalents, the concept of data has become broad and hard to pin down. Underlying Human Data Relations is to explain what roles data can play in people’s lives – what it **is** to people. Through the Case Studies, external work and my prior learning, I have so far identified 8 distinct lenses to consider how people might relate to it. These are modelled in Table 15.

Table 11: Table 15. **Eight lenses on data.**

Way of thinking about data	Explanation & Implications
Data as property	Data can be considered as a possession. This highlights issues of ownership, responsibility, liability and theft.
Data as a source of information about you	Knowing that data contains encoded assertions about you and can be used to derive further conjectures enables thinking about how it might be exploited by others, but also how you can explore and use it yourself for reflection, asking questions, self-improvement and planning. It invites consideration of the right to access, data protection, and issues around accuracy, fairness and misinterpretation / misuse.
Data as part of oneself	A photo or recording of you, or a typed note or search that popped into your head could be deeply personal. This lens on data highlights issues around emotional attachment/impact, privacy, and ethics.
Data as memory	Data can be considered as an augmentation to one’s memory, a digital record of your life. This lens facilitates design thinking around search and recall, browsing, summarising, cognitive offloading, significance/relevance, and the personal value of data.

Way of thinking about data	Explanation & Implications
Data as creative work	Some of the data we produce (e.g. writing, videos, images) can be considered as an artistic creation. This lens enables thinking about attribution, derivation, copying, legacy and cultural value to others.
Data as new information about the world	Data created by others can inform us about previously unknown occurrences in our immediate digital life or the wider world. This lens is useful for thinking about discovery, recommendations, bias, censorship, filter bubbles, and who controls the information sources we use, as well as who will see and interpret data that we generate and what effects our data has on others.
Data as currency	Many data-centric services require data to be sacrificed in exchange for access to functionality, and some businesses now explicitly enable you to sell your own data. This lens highlights that data can be thought of as a tradable asset, and invites consideration of issues of data's worth, individual privacy, exploitation and loss of control.
Data as a medium for thinking, communicating and expression	Some people collect and organise data into curated collections, or use it to convey facts and ideas, to persuade or to evoke an emotional impact. This lens is useful to consider data uses such as lists, annotation, curation, editing, remixing, visualisation and producing different views of data for different audiences.

When considering HDR, it is important to recognise that people may think of their personal data through any or all of these '*lenses*' [Karger *et al.* (2005);2.2.2] at any given time, and any process or system design involving data interaction should take these into account.

Looking across this set of lenses, it is possible to identify four specific roles that data can serve:

1. Data has a role as an **artifact of value** to your life;
2. Data has a role in **informing** you about yourself, the world, and the prior or recent actions of others that may affect you;
3. Data has a role as a **usable material with which to effect change** in your life;
4. Data has a role as a **means to monitor changes** in data holders' behaviours, digital influences upon you or changes within your life.

7.2.2 Human Data Interaction or Human Information Interaction?

To unpack HDR further, it is important to highlight the difference between humans relating to data, and humans relating to information. Human Data

Interaction (HDI) concerns the way people interact with data. Mortier *et al.* (Mortier *et al.*, 2013, 2014) defined the field of HDI without distinguishing data (the digital artifact stored on computer) from information (the facts or assertions that said data can provide when interpreted). This is an important distinction. The parallel field of Human Information Interaction (HII) originated in library sciences, and considers the way humans relate to information without regard to the technologies involved (Marchionini, 2008). William Jones *et al.* called for a new sub-field of HII in an HCI context¹⁸, observing that it is important to include a focus on information interaction because HCI can “*unduly focus attention on the computer when, for most people, the computer is a means to an end – the effective use of information*” (Jones *et al.*, 2006). DIKW theory (see 2.1) highlights that **interpretation of data to obtain information** is a discrete activity. This was borne out in the findings of Case Study Two, where it became clear that participants have distinct needs from data, and from information (5.4.3.2). Access to data **and** information is critical to both understanding and usability, as detailed in section 6.1.2 and 6.1.3.

Drawing on this theory, we can see then that in considering Human Data Relations, there are in fact three distinct artifacts to consider:

1. ***data*** - the stored digital artifacts pertaining to users held by organisations for algorithmic processing and human reference, copies of which can be obtained using individual data rights.
2. ***information about individuals*** - the collection of facts and assertions about the individual and their life, which are obtained through human or algorithmic interpretation of stored data (or in some organisations' case, through analytical inference).
3. ***information about data*** (also categorised in Table 9 / 5.3.1 as ***metadata***)
- stored facts about the data, such as where it has been stored, who has accessed it, how it was collected, what it means, or when it has been shared externally.

7.2.3 The Two Distinct Motivations for Human Data Relations

By making this distinction between the two types of information which people might interact with, and considering the six wants in Chapter 6, it becomes clear that there are two very different reasons why people might want better HDR:

- (i) to acquire *information about one's data*, so that one might exert control over and make informed choices about where *the data* is held and how it is used, in order to be treated fairly and gain more control over the use of one's personal data. This is **Personal Data Ecosystem Control (PDEC)**.
- (ii) to acquire *information about oneself*, so that one might gain insights into

¹⁸The group of HCI researchers involved in this panel were (with the exception of Raya Fidel) seemingly unaware of the existing HII field in library sciences as they positioned the publication as a call for a ‘new field’.

one's own behaviour and gain personal benefits from those insights or them to make changes in one's life. This is **Life Information Utilisation (LIU)**.

The two distinct processes that individuals might go through in pursuit of these motives are exemplified in Figure 30. PDEC is a process of holding organisations to account over and managing *what happens to personal data*, often regardless of what it means, whereas LIU is more concerned with *what the data means* and its inherent value as encoded life information, regardless of where it is stored and how it is used¹⁹. This novel way of modelling the motivations for data interaction were first proposed in my 2021 workshop paper (Bowyer, 2021).

7.2.3.1 Life Information Utilisation *Life Information Utilisation* is a superset of *Self Informatics (SI)*, as defined in 2.2.3. It includes all purposes relating to self-monitoring and self-improvement through data, but also includes all other uses of personal data including creative expression, evidence gathering, nostalgia, keeping, and sharing. Many of these desires were expressed in Case Study Two (see Table 12 in 5.3.3), and also hinted at in the Early Help context (4.4.1). While the existence of digitally-encoded information clearly unlocks new possibilities, LIU has existed in some form throughout human civilisation, as seen through analogue processes such as storytelling, journalling, scrapbooking, arts and crafts.

In the LIU context, the most important wants to focus on improving are *data understandability* (6.1.2) and *data usability*¹³ (6.1.3), which relate closely to the HDI concepts of *legibility* and *agency* respectively.

7.2.3.2 Personal Data Ecosystem Control Unlike LIU, **Personal Data Ecosystem Control** is an individual need that is new; arising as a result of the emergence of the data-centric world (2.1, 2.2.4). Only when organisations began to collect and store facts about people as a substitute for direct communication and involvement did it become necessary. The more data is collected about individuals, and the more parties collect and share that data, the greater the need for individuals to learn about that data so that they might influence its use (or risk their lives being affected in unexpected or potentially unfair ways).

¹⁹Of course, there is some overlap; the reason that organisations hold data is so that they can interpret it (usually algorithmically) to inform decision-making. In this way, organisations could be seen to be doing LIU of service users' lives for their own benefit. From a human-centric perspective, this grey area is situated as part of PDEC, as from the individual perspective, how organisations understand you through information will inform decisions that affect your life. Thus, this can be considered part of the reason why one might want to exert control over use of your data, rather than being part of exploiting data to gain self-insights and personal benefits.

²¹The illustrated processes assume reliance on existing data access processes such as GDPR, where the only access is through provision of a copy of one's data. This is in fact, not ideal, as it creates divergent versions and will quickly become out-of-sync, however for the sake of simplicity this inefficiency is ignored here. Improvements upon this approach are explored in [INSERT REF]

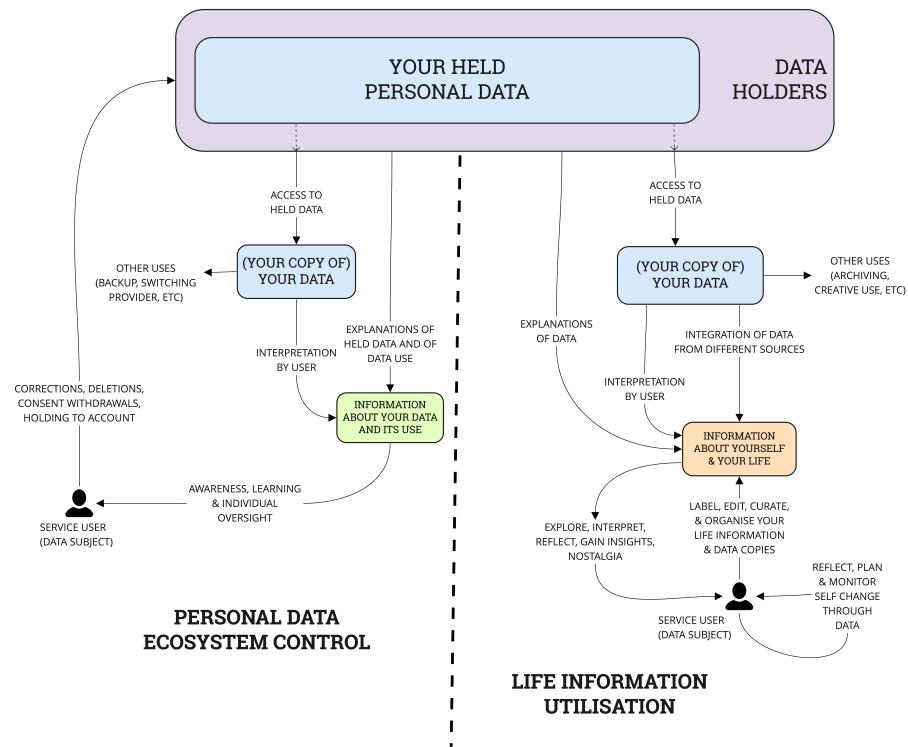


Figure 30: The Two Motivations for HDR: Controlling your personal data ecosystem and utilising your information about your life, with ‘idealised’²¹ processes illustrated

PDEC is a direct response to the power imbalance between data holders and individuals that the World Economic Forum described in 2014 [2.1.2; Hoffman (2014a)].

In the PDEC context, multiple data wants are important: visible data and transparent processes, as well as individual oversight and involvement. For simplicity, the former two wants can be referred to collectively as “*ecosystem transparency*”, and the latter two as “*ecosystem negotiability*” (drawing on the HDI concept of *negotiability*), and these terms will be used below.

7.3 The landscape of opportunity for better HDR, with its obstacles and resulting insights

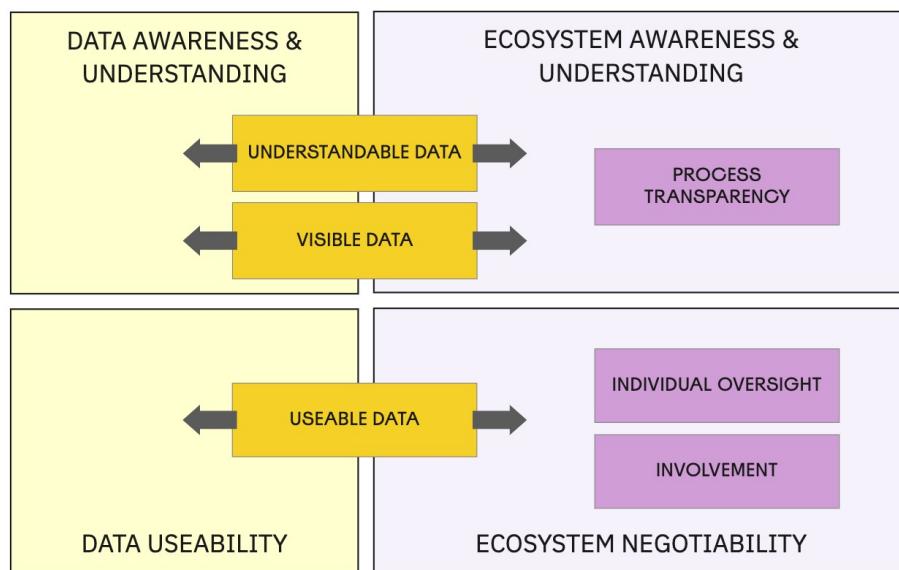


Figure 31: Figure X: Mapping the Six Wants into Objectives for the HDR Opportunity Landscape

In this section I will describe the high level obstacles to better HDR, in four sections. These are arranged into six groupings. The first four groupings correspond to the six wants identified in Chapter 6. Two additional groupings have been included to cover more general human and technical challenges that affect all endeavours in this space:

- 7.3.1: Obstacles to Data Awareness & Understanding
- 7.3.2: Obstacles to Data Useability
- 7.3.3: Obstacles to Ecosystem Transparency (including visible data & transparent processes)

- 7.3.4: Obstacles Ecosystem Negotiability (encompassing individual oversight and involvement) decision-making)
- 7.3.5: Obstacles in the Solution Space

7.3.1 Data Awareness & Understanding

7.3.1.1 Obstacle 1: Data Legibility People struggle to relate to data. It is not relatable because it is complex, not presented as meaningful information, and not easily interpretable as information. They lack tools to gain insights. To overcome this obstacle, more work is needed to make data relatable and to provide tools that can deliver valuable meaning and insights.

7.3.1.2 Insight 1: Life Information Makes Data Relatable When data is transformed into information that can be related back to moments, people, places or relationships in people's lives, it becomes instantly relatable. [from BBC: Data becomes meaningful when people are able to associate it with the real substance of their lives - people, places, organisations, causes or topics they care about. Therefore, the more associations you can find in data the more valuable it is.]

We can consider the different types of information in people's lives:

We need to model life information, not data.

7.3.1.3 Obstacle 2: The Personal Data Diaspora Every individual's personal data is scattered across multiple providers, devices, apps, held by hundreds of third parties. The complexity of a modern day digital life is unmanageable and overwhelming. People are inevitably ignorant of much of their data and its use. This can lead to resignation and apathy. To overcome this obstacle, approaches must be identified that recognise the scattered, complex reality of each individual's personal data ecosystem and begin to make it visible and understandable.

7.3.1.4 Insight 2: Ecosystem Information is an antidote to Digital Life Complexity No matter how understandable the data itself is, it is also critical that people can access information about their data ecosystem. Without this, there will always be aspects of their data that are beyond their awareness or beyond the reach of what they can access, control or manage. Many tools today do not recognise this, and build for a world that does not exist. It is important that people have tools that allow them to interact with multiple providers and data sources across their digital life.

7.3.2 Data Usability

7.3.2.1 Obstacle 3: Data isn't free Almost all data is constrained in some way, limiting its usability. It may be held by a particular provider and inaccessible. It may be stored in a format which is hard to use or change. It

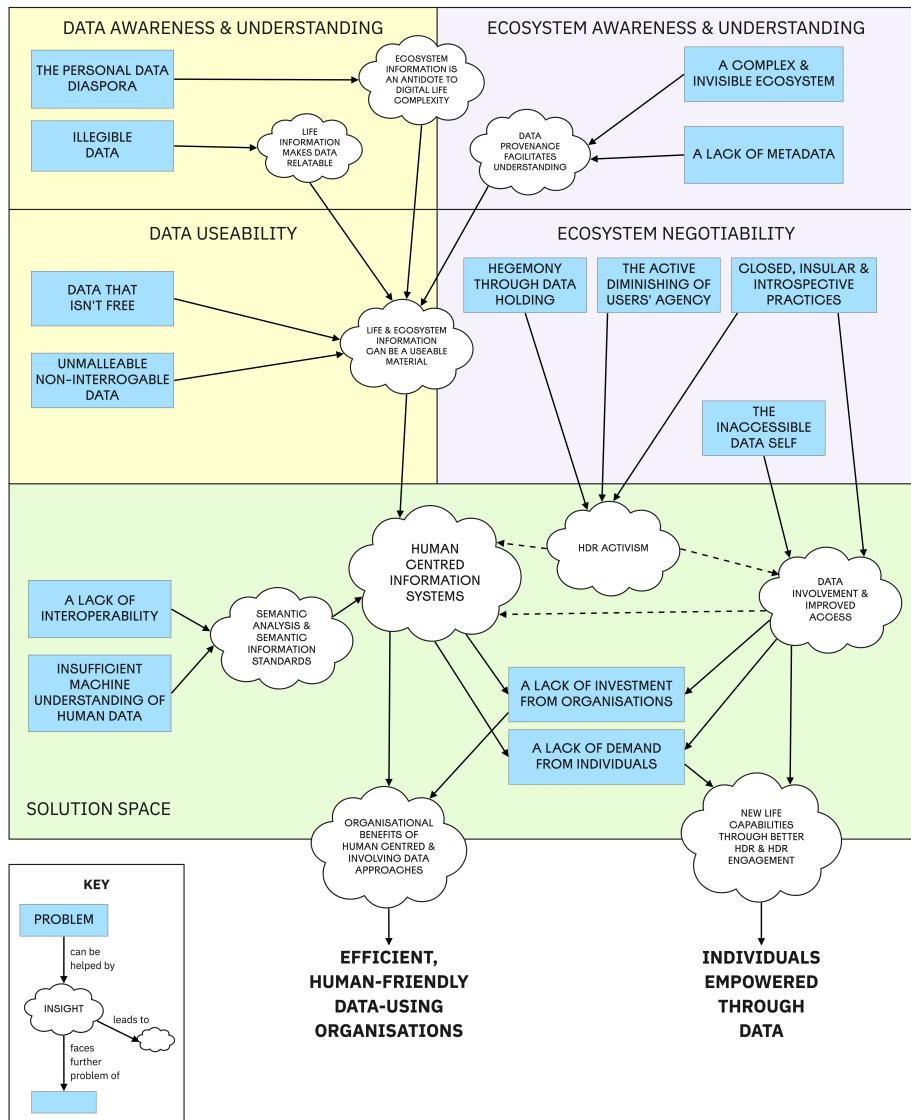


Figure 32: Figure X: Obstacles and Resulting Insights in the HDR Opportunity Landscape

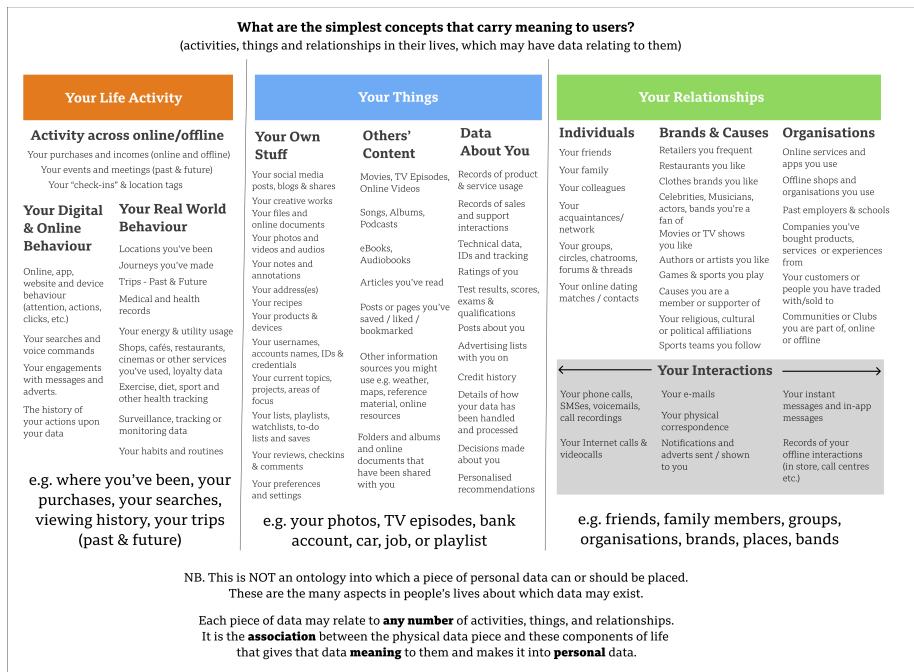


Figure 33: Figure X: Life Concept Modelling

may only be visible after a delay. It may be unchangeable. To overcome this obstacle, we need to find ways to extract data from its current constraints and to remove some of these technical or temporal limitations.

7.3.2.2 Obstacle 4: Data that is Unmalleable and Non-Interrogable

Even once an individual has gained possession or access to the relevant parts of their personal data, it can be extremely hard to use. This partly comes from a lack of malleability - the ability to break it down, look at it from different perspectives, reconstitute it in different ways. Put simply, people need to be able interrogate their data - ask questions of it. This requires more than just an ability to view visual representations of data, but an ability to interact with the data and produce new views and insights that can help to answer specific questions. Making some of the PIM and SI capabilities described in 2.2.2 and 2.2.3 can help to address this, but more capabilities can be made available and are needed to fully overcome this obstacle.

7.3.2.3 Insight 3: Life Information & Ecosystem Information should be Useable as a Material

Many computer operating systems and interfaces today treat files as the basic material that an individual can manipulate. To truly empower users to make use of their data, we need to move to a model where pieces of life information – facts (or assertions) – can be created, deleted, moved, grouped, annotated, copied, shared, modified, labelled, organised, separated or otherwise manipulated instead. So far, people access data within products. But what they need is a platform, not a product. We need an information operating system.

7.3.3 Ecosystem transparency

7.3.3.1 Obstacle 5: A Complex and Invisible Data Ecosystem The first and most obvious barrier that individuals face in managing a complex personal data ecosystem is that, to a great degree, they cannot see it. For example, it is very easy to allow a handful of communication and social media apps access to your address book or contact list, and before you know it you have created a complex and unmanageable network of connections that silently sync and propagate your addresses and phone numbers across the Internet. And there are deeper layers which are not even slightly visible to users: networks of data brokers, advertisers and digital cookie companies exchanging user identifiers, activity data and personal information about you while you browse or use apps. As Chapter 5 showed, even though people have been granted new rights to access their data and information about provider data sharing activity, the ability to effectively execute those rights to build up a meaningful picture of your personal data ecosystem is severely limited by inconsistent, incomplete or unclear responses. The strong negative practical impacts of today's complex digital lives were already described in section 2.2.4; managing the complexity is an overwhelming, unmanageable task that even personal data experts are not fully able to get a handle on. The ability to provide a user with ecosystem transparency

is hindered by the complexity and multiplicity of the data relationships they have been encouraged to set up, and by a lack of tools to provide a meaningful, or indeed any, view of those relationships. A further aspect to this obstacle is that no individual or organisation has the ability to see the whole of a user's ecosystem, and there is little commercial motive to try and solve this problem, as every provider focuses just on their own apps, websites and services.

7.3.3.2 Obstacle 6: A Lack of Metadata From this complexity an additional obstacle becomes evident. There is scant attention to information **about** your data. Even where data access rights are executed (or data is shared via human means such as in Chapter 4), the attention is on the data itself: what it says. Chapter 5 shows that some of the most desired information was not the data itself, but how it is used and shared and what is inferred from it, yet this was rarely forthcoming. There are many pieces of information that can be quantified about an individual's data, as illustrated in Figure X, which I created during my internship at BBC R&D:

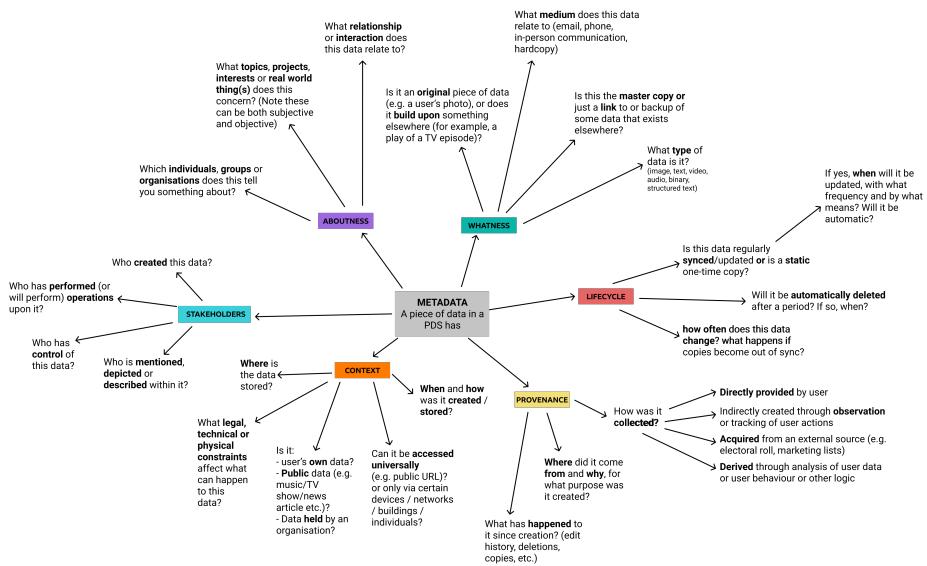


Figure 34: Figure X: Some of the many aspects of metadata that might exist about a datapoint or dataset

[EXPLAIN ASPECTS]

To provide users with meaningful transparency, many of these aspects will need to be tracked and visualised; not an easy task given the complexity and the potential to overwhelm a user, but nonetheless a vital first step on the road to giving individuals the ability to have oversight of their personal data ecosystem and take action within it.

[ADD REFERENCE BACK TO 2.2.2 METADATA]

7.3.3.3 Insight 4: Data Needs Provenance A number of researchers have independently identified the importance of keeping the history of a piece of data with it. Without context, data loses meaning (a phenomenon witnessed in Case Study Two – see 5.4.3.1). The idea that what has happened to not just an individual but to a piece of data over time is important is a key part of the thinking behind temporal PIM systems, from Lifestreams (Freeman and Gelernter, 1996) to activity streams (Hart-Davidson, Zachry and Spinuzzi, 2012) (see 2.2.2). William Odom, Siân Lindley and colleagues proposed the idea of file biographies, which view the lifetime of a file as something that should remain connected and traversal in order to understand the context of the file at its different interaction points. Significant research in this space has been undertaken by Professors Mike Martin and Rob Wilson at Northumbria University, formerly Newcastle University, who express the idea of **data with provenance**; in other words that data must carry with it the details of why it exists, how it came to be, and what has happened to it since its inception, and that provenance must be communicated alongside any visualisation of the data, if it is to be fully understood (ie. for its context []). This plays into the ideas of Gitelman, Neff and others, that data is not neutral and in fact is inherently biased, since it was created for a specific purpose with a specific agenda in mind (Gitelman, 2013; Neff, 2013). [ADD MORE DETAIL FROM MIKE MARTIN PAPER AND EMAIL HERE]. While it is not a solution in its own right, it is clear that data with provenance is very likely to be a critical and valuable part of any design that aims to help individuals with managing to get an overview of their complex and invisible personal data ecosystems.

7.3.4 Ecosystem negotiability

7.3.4.1 Obstacle 7: Provider Hegemony and the Nature of Digital Power In the pursuit of individual oversight and greater involvement, the power imbalance between individuals and data holders (2.1.2) becomes most clear. While the Internet itself initially held the promise to be a great leveller and to empower individuals, this potential has largely been suppressed. Data is owned and controlled by service providers, who also design and control the interfaces, apps, websites and devices through which individuals access those services, controlling what (if any) of the data stored behind the scenes, and of the internal processes that use that data, is visible, and how such data and processes are represented. In Jasperson *et al.*'s detailed metatriangulation review of types of power that affect technology systems (Jasperson *et al.*, 2002) we can identify a number of specific types of power that clearly are in effect in today's digital data-centric service provider context:

[ADD TYPES OF POWER FROM JASPERSON WITH CONTEXTUAL EXPLANATIONS] [structural power, resource control, centralisation etc]

(Foucault, 1975)

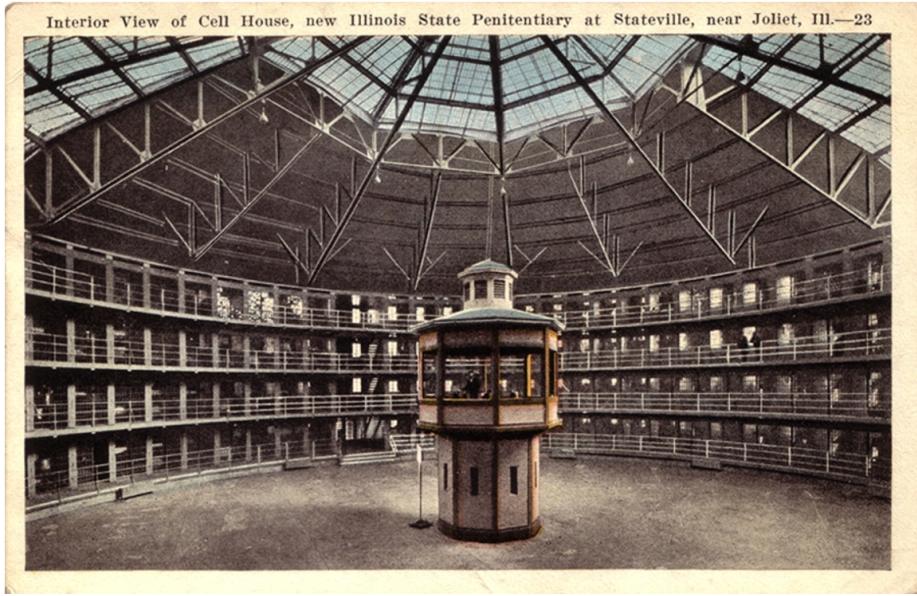
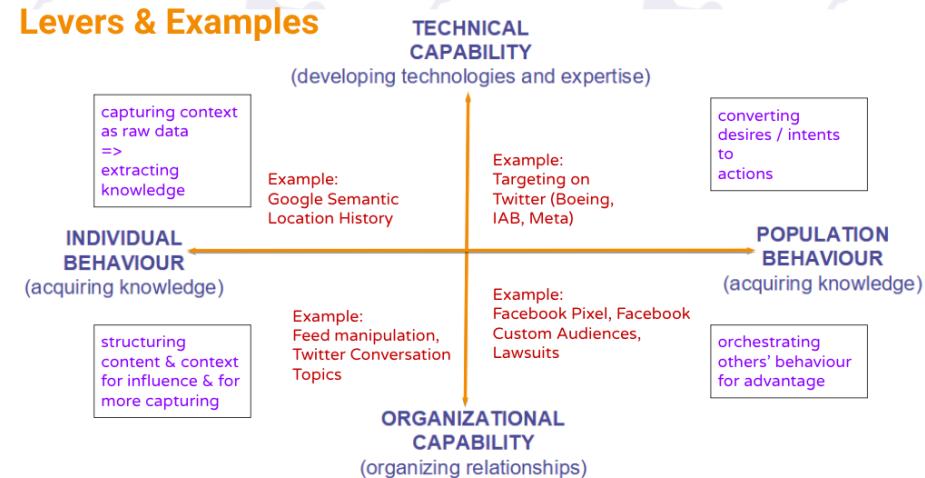


Figure 35: Figure X: The Panopticon Structure of the Illinois State Penitentiary

A helpful analogy for the relationship between provider and user can be seen in the design of Panopticon: A style of prison architecture designed to elevate the power of the prison guards to observe all the prisoners easily at any time and to diminish the ability of prisoners to operate in privacy or to see those in authority. Jeremy Bentham [REF], drawing on the philosophy of Foucault [REF], makes clear that such design is political, and shows that power can be enforced by the environment. This is a useful mental scaffold to keep in mind; as explained below [REF], we can think of today's digital landscape as similarly power-enforcing. Code is law [ADD REF Lessig], and interfaces limit what individuals can do. By holding data behind interfaces shaped to serve their own interests, the landscape is controlled by the data holders. [UPDATE THIS BASED ON OTHER WRITING ABOUT PANOPTICON]

7.3.4.2 Insight 5: Data Holders Exploit Four Levers of Power to Manipulate the Digital Landscape Sitra's #digipower investigation [REF], of which I was project leader for Hestia.ai, was a successor to my Case Study Two, but worked with high profile politicians and European influencers and added additional technical audit techniques. Its focus was not on the individual experience of data access, but on using those experiences and acquired datasets to better understand the data ecosystem. Through this research, a model was produced to understand the ways in which service providers (and in particular the larger ecosystem-level platform providers such as Google and Facebook) exert power over individuals and smaller organisations. This model is reproduced in

Figure X:



[TODO: redo this diagram]

[ADD EXPLANATION AND REFERENCE TO THE PIDOUX REPORT]

Through this landscape it is clear that the most powerful data holders exert huge influence over the digital landscape, in terms of what is knowable and what is do-able. Individuals or activists' abilities to balance the landscape are hindered by the fact that they are operating in a landscape that the incumbent platform and service providers effectively control.

A key mechanism to highlight here is that the accumulation of information is implicitly and objectively a form of power. This is consistent with participants' observations in 5.4.4.1 that data holding and limiting access to it is a source of power. In terms of this being an obstacle, we can therefore see that as long as current platforms and service providers are free to collect so much personal information, the information landscape will remain imbalanced and individuals will not be able to acquire ecosystem negotiability.

7.3.4.3 Obstacle 8: Closed, Insular and Introspective Practices Today's digital landscape is fractured[REF Splinternet]; myriad providers vie to pull users into service relationships or connected ecosystems that will encourage a flow of money and attention to their own products and services, most evident from companies such as Apple, Amazon, Facebook, Google and Microsoft (the so-called 'big five') that have multiple touchpoints into people's lives through different devices, apps and services. We can think of these different providers' sub-Internets as walled gardens or silos [REF]. Commercial motives encourage them to get users to spend time in their own proprietary spaces (so that resultant ad revenue can be captured) and in order to maintain subscription revenues it is in providers' interests to make it hard for individuals to leave or switch providers.

In effect, providers build for a world that does not exist, where every individual is imagined to only interact with that single company's interfaces. There is little incentive to open up the ecosystem when the free flow of information and of users might result in loss of income for the company in question. Users with negotiability would be more able to leave. And this also encourages keeping users in the dark (5.4.2). The less agency and negotiability that users have, the more freedom the provider has to do exactly what they want with their data. In this context, users are, as Lawrence Lessig wrote, '*pathetic dots*' [ADD REF]. Thus service providers continue to build **proprietary, incompatible silos**.

But it is not only commercial motives that encourage insular attitudes to personal data and user service provision. In the SILVER project [ADD REF] meetings with local authorities and care providers revealed deep organisational and technical barriers within the public sector, with for example health organisations being typically unwilling to share health data with social care services, but also with different councils, community services and charities typically operating separate IT systems, each attempting to construct their own digital pictures within their own databases and very little operability. The problems of this technical reality are explored further in 4.1.2. From what we have observed, the introduction of GDPR and similar regulations has made this problem worse not better, as organisations and departments become increasingly paranoid about storing or sharing data they should not, or about the risks of acting upon data without sufficient consent. We learned of practices such as the sharing of information between care organisations verbally by telephone so that no digital trail was left.

It is clear that throughout society, there is a trend towards organisations being reluctant to work together around people's data, inclined towards collecting their own databases and not sharing them.

Also mention resistance to change

7.3.4.4 Obstacle 9: A Trend of Actively Diminishing Individuals' Agency As a result of the practices and motives described above, the last decade has seen much reduction in individuals' agency. When software was sold in a box, manufacturers competed based upon which product would let the user take home the greatest range of features and capabilities. New releases with new features drove new product sales. But in the cloud computing era, a smaller set of core features done well is sufficient to guarantee an ongoing subscription revenue from a user. Cost savings in development and support costs can be made by reducing feature sets. The relentless pursuit of increased profits and further cost saving sees products lose, not gain, features. Interfaces are reshaped to serve businesses' interests first and foremost. As described in 2.3.5, the primary concern is about making user behaviours constrained, predictable and profitable, rather than meeting their needs or providing maximal value. One of the most revealing examples is seen in the case of Facebook. Users used to be free to consume their friends' posts in other clients via RSS feeds. These were removed, forcing users to use only Facebook's interfaces, where their eyeballs

can be monetized (Twitter closed its APIs too to a great degree, killing off many third party readers). On Facebook users used to have the ability to view the latest updates from a particular list of friends or of news pages. These features too were removed, presumably to increase monetization through the main feed. The ‘Friends’ page on Facebook currently shows a list of recommended new friends; to access your current friend list requires an extra click. Encouraging users to grow their networks is prioritised over user convenience.

Companies change their practices to limit users’ agency (and their own accountability to customers) too. For example, Facebook recently announced they will no longer collect historical location data from users (though they will still use location information). This makes it harder for users to see how their data has been used. Tiktok announced they will rely on legitimate interest rather than consent when it comes to using users’ activity data to personalise the app experience, removing users’ ability to withdraw consent to such use. Unchecked, it is clear that trends to reduce users’ agency and further providers’ interests will continue, creating another obstacle to be tackled.

7.3.4.5 Obstacle 10: The Inaccessible Data Self Earlier in this thesis the concept of a data self has been introduced (4.4.1, 4.4.3, 6.3). We know from both the preliminary study with families (Bowyer *et al.*, 2018) and Case Study Two that data serves as a proxy for direct human involvement of the served individual(s). Put simply, service providers try to minimise interaction with people, by maximising their usage of data to represent people. We are viewed through the distorted lens of our data selves. Despite the inherent challenge of representing people fairly and accurately in data [Bowyer *et al.* (2018); 4.4.1; 5.4.4.1], this is the default modus operandi for service provision today. This therefore represents a key obstacle to ecosystem negotiability today: how can individuals be given the ability to influence and shape the data self that providers will use to understand them and make decisions?

7.3.5 Obstacles in the Solution Space

While in the previous four subsections it was possible to identify obstacles relating to specific HDR wants, there are also some readily identifiable obstacles that will affect all our endeavours to improve HDR. Obstacles relating to human challenges are described in this section, and technical challenges are addressed in the following section, 7.3.6.

7.3.5.1 Obstacle 11: A lack of demand and HDR motivation, and perceived hard work In considering the recommendations of Case Study One (shared data interaction between the state and the individual) and of Case Study Two (new human-centric data practices by service providers), and in exploring possible new human-centric system and interface designs through my work with BBC R&D, it is evident that even if new human-centric types of computer system or service interaction practices can be created, we cannot assume that

people will be inclined to use them. Today, data is overwhelming, complex, and ‘sounds boring’. There is no denying that currently, engaging with one’s personal data economy to any degree more than that of passive consumer, is hard work. People routinely accept data sacrifice, click through T&Cs and cookie banners and are unwilling (or in some cases lack sufficient technical literacy, comprehension or skill) to do the work of asserting control over their digital lives. There is not a clear demand for holistic and novel ways of managing your digital life and exerting agency and negotiability over it. This can be seen as an obstacle that affects all HDR improvement approaches we see, and indeed is why many companies in the emergent PDE economy (2.3.4) struggle to find a business model. But this should not deter disruptive innovation nor does it indicate that such offerings would not be useful. As Henry Ford famously said, “If I had asked people what they wanted, they would have said faster horses.” Nonetheless, it is a clear overarching obstacle to overcome.

7.3.5.2 Insight 6: New Life Capabilities and Pain Relievers Drive User Demand Through work at the BBC R&D exploring how to better connect people with their data, it became clear that there is a way to combat such indifference and apathy of users. It emerges from the realisation that the way people find value in data is to connect it to their lives. The more that people see relatable life information and can imagine ways to harness that information in their everyday life, the more motivated they will be. [include the three concentric circles diagram a bit like the one Rhianne used]

As an example, myself and BBC colleague Jasmine Cox imagined focusing on address books and contact lists as a strong relatable starting point that could easily generate a user demand. Many people face a complexity they cannot easily manage when it comes to the automated syncing and sharing of potentially sensitive contact information between devices, apps and providers, and developing human-centric personal information management capabilities to bring that messy situation under control would offer a clear and tangible benefit to users.

Another example that is helpful to consider is my the example from my 2011 article: that of a vacation, as shown in Figure X (Bowyer, 2011). Today, all the information around such a holiday is scattered into multiple systems - emails, online provider bookings, chat logs, cloud synced photos, web browser bookmarks, smartphone location logs, etc. It is not hard to imagine that a system that was able to bring all related information about that vacation together in one central place could deliver huge value to users and be very compelling. Such context-targeted human-centric offerings can have a much greater chance of generating interest and impact than offerings that merely allow you to “organise your data” or some other abstract phrasing.

As with any public offering of a product or service, it is important to start with identifying a problem or need, and to demonstrate a potential tool or solution that can help. In particular, there is a need to let people do **new** things that they could not do before. This has been identified as a key ingredient of user

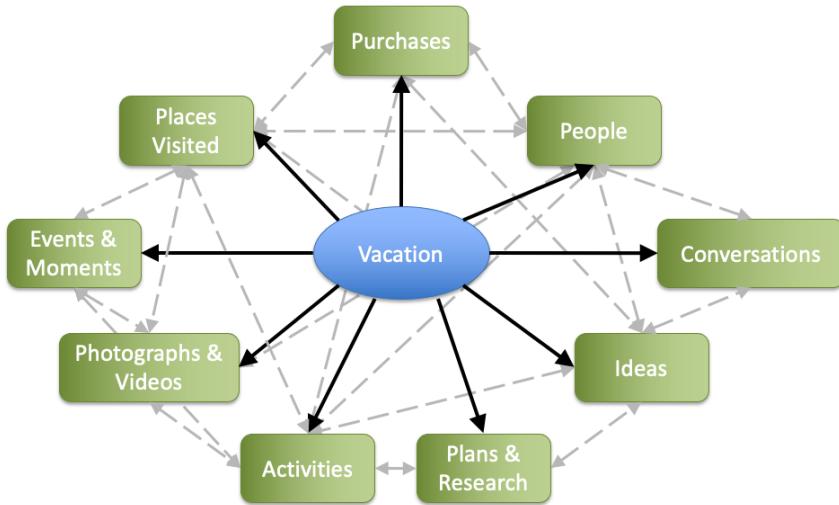


Figure 36: Figure X: Conceptual Semantic Grouping of Related Data for a Vacation

empowerment (Meschtscherjakov, Wilfinger and Tscheligi, 2014; Schneider *et al.*, 2018). This became a driving influence for design thinking on the BBC R&D Cornmarket project. It is not enough to believe that “If you build it, they will come.”

7.3.5.3 Obstacle 12: A lack of Interoperability Obstacle 8 (7.3.4.3) already touched on the issues around different companies developing different standalone walled garden or silo user experiences, from a sociotechnical or systemic standpoint. But there is a very specific technical problem that must be acknowledged across all HDR improvement approaches, and that is that it is very difficult to build technical systems that connect and exchange data with each other. This was witnessed first hand by our development team on the SILVER health data interface project [REF] which endeavoured to build a bridge making health data available to Early Help support workers. Not only are there a lack of standards, with each organisation using their own databases and formats for storing data, but often the programming interfaces (APIs) that would be needed to interface between different systems (sometimes legacy systems) do not exist, are insufficient. Furthermore, there can be issues around licensing and consent when data passes from one domain to another. Data sharing agreements must be established, especially in the public sector which is by its nature more liable to scrutiny and accountability. But at an abstract level the technical obstacle, the problem is one that has always faced the tech industry, which is that there often is no universally agreed way to represent important concepts - in this case

human-centric information concepts such as events, social media posts, website visits, location history information, app activity, etc. And any entity that does create a standard then faces the challenge of trying to persuade others that their standard is the best one to use. In general, standards work best when established by non-commercial industrial standards bodies (for example the World Wide Web Consortium (W3C) or International Organisation for Standardization (ISO) and then mandated through policy such as European Union law. Such standards must be established with input from industry experts.

7.3.5.4 Obstacle 13: Insufficient machine understanding of human data Following on from the previous obstacle, but a subtly different point, is that it is technically difficult for machines to handle human information. Without deliberate coding, software can only understand streams of binary data as files or datasets, and does not understand what people, places, events or entities the facts within the data relate to. Therefore, it is necessary to consider how algorithms and systems can be designed to include an understanding of the semantics (meaning) of the information within the files and data records they handle. For example, the data record representing a post on Twitter looks entirely different to the data record representing a post on Facebook. No algorithm can recognise or unify these disparate pieces of data as two instances of the same semantic concept until its specifics of the data format can be mapped back to a common semantic abstraction of a “social media post”. [find meaning in user’s data]

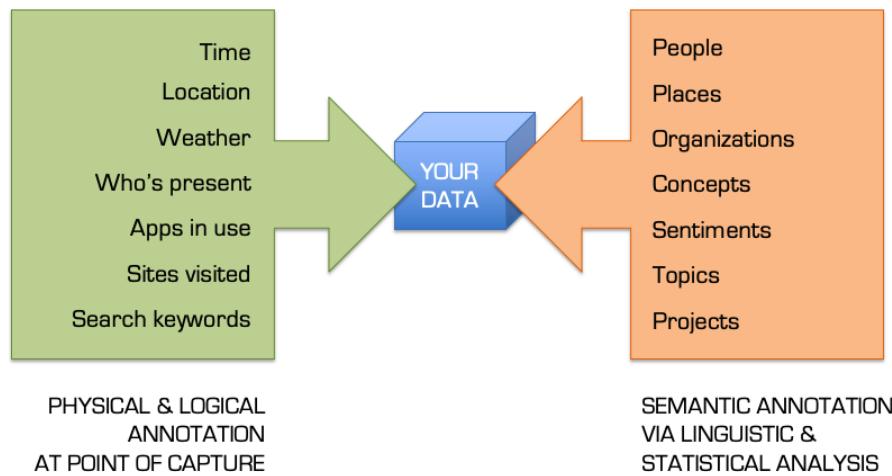


Figure 37: Figure X: Annotating Data with Semantic Context

7.3.5.5 Insight 7: Semantic Analysis and Information Standards can Transform Data Storage and facilitate human-centric interface building

This leads to the next insight: that to build systems and interfaces that are able to deal in human concepts and represent the elements of everyday life requires building systems that store semantic context and semantic associations, not just raw bundles of data. This is advocated by the Web's inventor Tim Berners-Lee in his vision of a Semantic Web (Berners-Lee, Hendler and Lassila, 2001) and by proponents of networked PIM systems (2.2.2). There is a need to develop standard ways to digitally model facts and assertions about users' lives, so that those disparate pieces of data can be unified, connected, correlated and compared. Sizable industries have built up around Content Analytics and Enterprise Content Management. Through the capture of metadata at the point of data recording, and through subsequent programmatic analysis of stored data, as illustrated in Figure X (Bowyer, 2011), we can begin to teach computers what the data we store represent. Machine learning technologies and Artificial Intelligence have pushed machine understanding of human words, images and content to impressive levels in recent years and such technologies can certainly be helpful, but in fact at the core what we are talking about here is something much simpler than AI; It is simply about labelling datapoints in as many different ways as possible so that those datapoints can be associatively retrieved from many different angles, and providing humans with ways to amend incorrect labels and to reclassify data or apply new semantic associations.

7.3.5.6 Insight 8: Better HDR can deliver business value through increased accuracy and consent, and decreased liability.

7.4 Four approaches to improving HDR, and their specific opportunities

Now, having established some of the key obstacles to improving HDR, we can move to considering what opportunities exist to pursue the HDR wants and to overcome those obstacles. This section will first introduce a framing for those opportunities, and then illustrate specific opportunities in detail.

In Figure X, the ToC frame introduced above in 7.1.3 / Figure 29 is used as a canvas upon which to position the different trajectories for changes that could improve HDR. By enumerating the possible types of activity that can bring about change, each of the four quadrants's core change trajectory can be named, as shown in purple, forming the backbone of the roadmap for improving HDR, which can be summarised thus:

- ***Learn & Discover*** (CI): In this quadrant, individuals, researchers, activists and other stakeholders work in groups to understand data attitudes and user needs, and to gain collective knowledge of data collection and usage practices which are sometimes hidden.
- ***Defend & Create*** (CE): In this quadrant, activists work to ensure current HDR capabilities are not eroded, while researchers, designers, technologists and social innovators design and create new technologies, operating models,

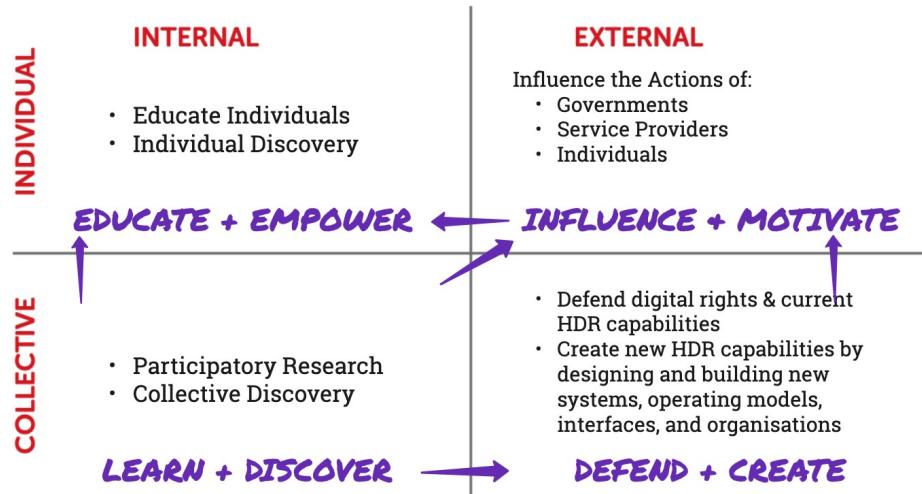


Figure 38: Figure X: ToC chart showing Trajectories of Change for improving Human Data Relations

organisations and interface designs, the structures enabling a world with better HDR. Given the majority of the opportunities for change identified in this chapter occur in this quadrant, this quadrant will be further subdivided into four different types of activity:

- Defend the Status Quo and Push for Compliance
- Create New Structures and Systems
- Create New Information Interfaces
- Create New Capabilities
- **Influence & Motivate** (IE): In this quadrant, individuals' relationships with data and with data holders, as well as data holders and policymakers' relationships, can improve. For us, as external actors seeking change, the task is to influence the many parties by showcasing and facilitating newly created structures and capabilities, and to harness our new collective knowledge to advocate the benefits of changing data-related behaviours.
- **Educate & Empower** (II): In this quadrant, individuals' ways of thinking about data and data holders grow and evolve towards a state we could call 'feeling empowered'. This change is driven through education, improving Human Data Relations literacy, and through the experience of new capabilities and changed relationships with data and with data holders that can empower individuals to hold a more aware and equitable position in those relationships and in their digital life.

[TODO: do we need a summary diagram here?] [Figure X: SUMMARY OF OPPORTUNITIES]

7.4.1 Learn & Discover

Research such as that conducted in this PhD is an example of the collective, internal focused activity that can be done in this quadrant to further the goals of better HDR: Groups of people working together using a variety of techniques such as participatory co-design, interview-based quantitative studies, design prototype evaluation and other HCI techniques can gain new understandings of individual needs and experiences in HDR. However rather than mapping out such possibilities, this section will focus on more novel approaches that go beyond traditional HCI research towards activities that are potentially more socially impactful.

7.4.1.1 Opportunity 1: Auditing Data Holders Helps with: Ecosystem transparency

Through the emergence of new tracking tools such as TrackerControl and Apple's App Activity Reports, individuals can observe the actual behaviour of the apps they use, providing a new means to identify potential data sharing destinations, to assess whether providers are meeting their promises, and to uncover new questions that can be asked of providers using data access rights. By collectively examining and comparing such data, it is possible to begin to map out the data ecosystem, as was done in the digipower investigation [REF].

7.4.1.2 Opportunity 2: Collective Investigation Helps with: Ecosystem transparency

This sort of combination of individual observations is just one of many ways in which individuals can, through working together, discover more information about data usages and practices. Collectives offer a powerful means to examine how providers categorise users and process their data, for example by comparing field values to understand the range of possible values or inferences a data holder might have stored, or by comparing variations in information presentations, data rights handling or customer service experiences to reverse engineer provider practices.

7.4.1.3 Opportunity 3: A 'Data Understanding' Industry Helps with: Data Understanding, Ecosystem Transparency

Given the complexity of today's digital landscape and the forces that hinder better HDR, there is scope for an industry to develop around 'data understanding' services. This can encompass everything from self-service tools people can use to gain insights over their data (such as those provided by Ethi or Hestia.ai), to workshops helping consumer organisations, journalists, regulators, lawyers and other interested parties to collectively gain understanding and value from data so that they might better achieve their goals, as well as serving a general educational purpose for example in schools. This industry is beginning to emerge,

but faces challenges in funding, scalability, governance, and credibility and should be supported (REF Pidoux et al).

7.4.2 Defending the Status Quo and Pushing for Compliance

Given the shifting power balance of the information landscape outlined in 7.3.4.1, 7.3.4.2 and 7.3.4.4, it is clear that there is an opportunity, perhaps a need, for HDR reformers to carry out activities that monitor and publicise any changes that providers make that reduce individuals' HDR capabilities. Having identified such changes it is then easier for those HDR reformers, and indeed the wider public, to fight to protect and maintain current capabilities, as we see in the Right to Repair movement [REF] or the Net Neutrality movement [REF]. [mention also the idea of pushing to make sure what should be done, is done - e.g. in GDPR returns]

7.4.2.1 Opportunity 4: Tapping the Seams in pursuit of a Free Information Landscape Compounding the impacts of reducing agency described in 7.3.4.4 is the 'dumbing down' of technology. Apple, for example, encourages users to consider technology as 'magical', rather than as understandable tools to be harnessed and understood; Such thinking is manifested in their hardware design too: phones that cannot be opened up, expanded or repaired [REF]; the removal of accessory ports, disk drives, and headphone jacks [REF]; increased controls over what can be installed on users' hard drives and which areas of disk can be modified [REF]. These changes simplify the technology and bring it to a more mainstream audience, something that the iPhone and iPad must be given due credit for - but it happens at the cost of reducing user agency. Companies like Apple increasingly encourage users to think of technology as a black box, which you cannot and should not look inside.



Figure 39: Figure X: The 'Black Box' view that many modern technology providers present

[TODO rephrase this para as a better introduction to Seams] An important concept to understand in this space is that of *seams*. In 'The Politics of Seams' Storni outlines that current designs are incompatible with empowerment-in-use,

and highlights the role of design seams (and their removal) as being a key determiner of user power [REF]. He says that the designer passes some power to the user through their design, but also, that users should be able to take some power on their own terms (repurposing etc). He talks [says what] about the problems of technology as magic/design as conjuring:

“Magical design prioritises pleasing and surprising a passive user who can only use the solution as authorised” – Cristiano Storni (Storni, 2014)

Therefore part of what we need to be doing is (a) highlight and (b) removing seams/creating new seams between disconnected parts....

Groups of HDR reformers can combine development skills, innovation and disruptive design approaches to find and publicise new ways to circumvent providers' efforts to control and limit their users' agency, as illustrated by the use of web scrapers and web augmentation approaches to try and obtain information or functionality from providers that would otherwise be inaccessible. [also mention device tenancy (zeynep) and firefox containers/taking back power in the browser/browser as seam (reference Goffe et al)]

7.4.2.2 Opportunity 5: Collective Activism Collectives can also exert external influence in the adjacent ‘Defend & Create’ quadrant by using their learnings through data to demand change, as seen in the case of Uber drivers working together to obtain data on algorithmic judgements that affect their work and using that as evidence to help them demand fairer working conditions [REF]. [individuals collectively pressuring to improve GDPR responses, drive data portal improvements, etc] [traditional means e.g. press, public campaigns but also new ways e.g. mass GDPR or targeted GDPR] [mention dehaye’s pressure, leading to FB Off Site Activity, my success with Spotify] [noyb as example, also Privacy International, Bits of Freedom,] [cite examples from Mahieu papers][Mahieu, Asghari and Van Eeten (2018);mahieu2020a;Mahieu and Ausloos (2020)] [mention pooldata, data unions]

7.4.3 Create New Structures and Systems

7.4.3.1 Opportunity 6: A central home for your personal data Helps with: Data Understanding, Data Useability, Ecosystem Transparency, General Human Challenges Affected by: Lack of interoperability (7.3.6.1/12)

As others have identified, one of the most promising models for giving people a new and improved relationship with their data is to create a place where one's personal data can be stored and aggregated in one place, a personal data locker (see 2.3.4). This prospect was explored through the BBC R&D Cornmarket project during my internship, as detailed in 3.4.3.3 [OR MOVE SOME OF THAT TEXT HERE]. As alluded to in the quote opening Chapter 1, people's data is scattered (see also 2.2.4 and (Abiteboul, André and Kaplan, 2015)), and simply providing the ability to bring data from sources together

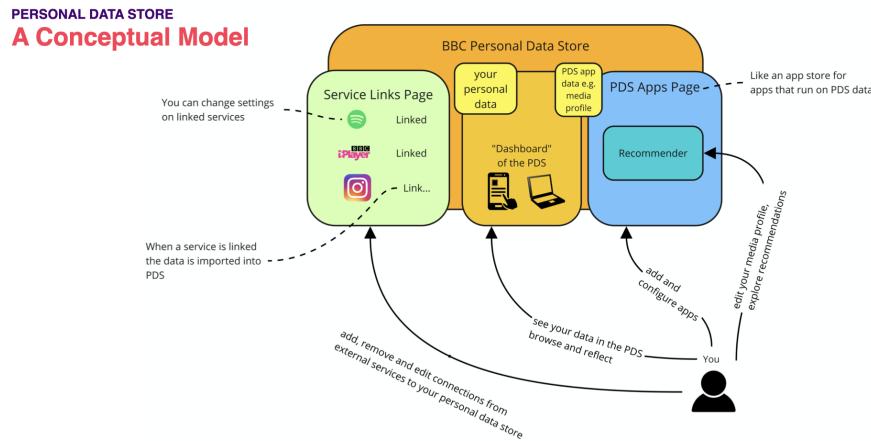
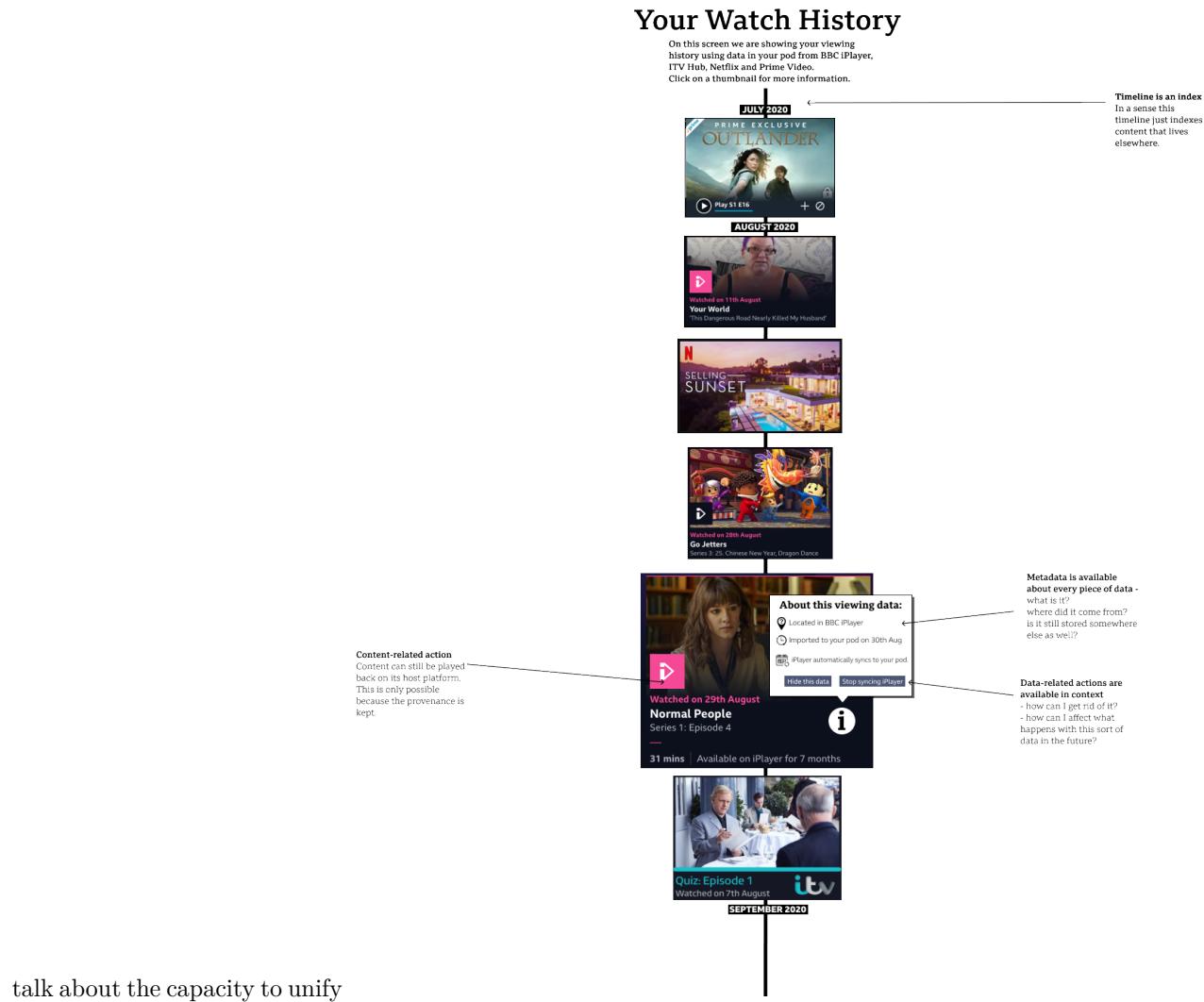


Figure 40: Figure X: A conceptual model for a Personal Data Store system that I developed for the BBC Cornmarket project

in one place can improve people's understanding of their data and its ecosystem. This integration requires technical standardisation but also [BRING IN SOME TEXT FROM BBC BLOG ARTICLE] [REF previous appetite for PDS <https://journals.sagepub.com/doi/full/10.1177/2053951720935616>]



talk about the capacity to unify

[add quote from BBC research where people liked the concept of a place for your data]

7.4.3.2 Opportunity 7: Modelling Data as Life Information

Helps with: Life Information as Material, Limited machine understanding of data

As part of the BBC R&D Cornmarket project, I carried extensive information modelling and design work with colleagues on how today's common types of data might be modelled as life information in order to help with the stated goals.

First we need to consider what a piece of data is, which is different from what data format it is or what semantic concept it represents.

What GENERAL types of personal data are we aware of?

DATA WILL HAVE
MORE THAN ONE
TYPE!

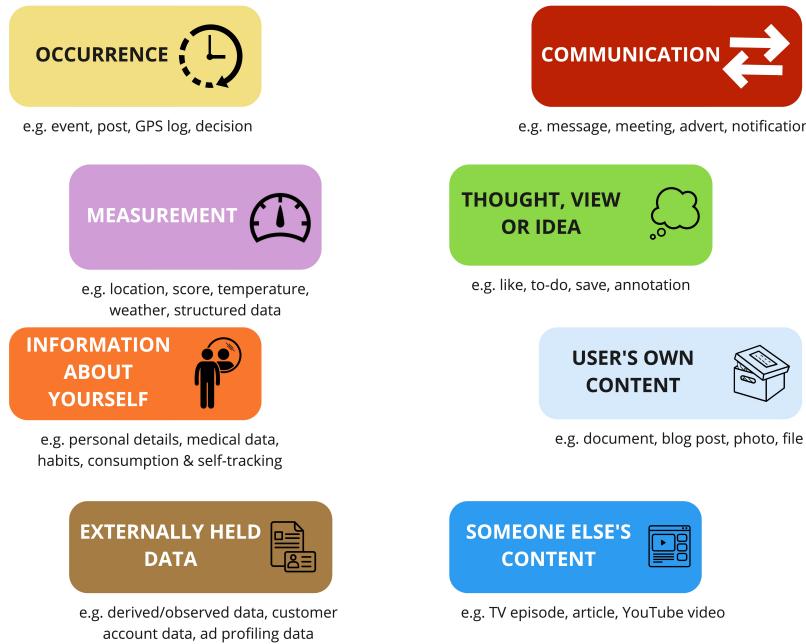


Figure 41: Figure X: High Level Data Types

It is possible to use some abstraction of commonalities to group together pieces of data that can perform a similar role:

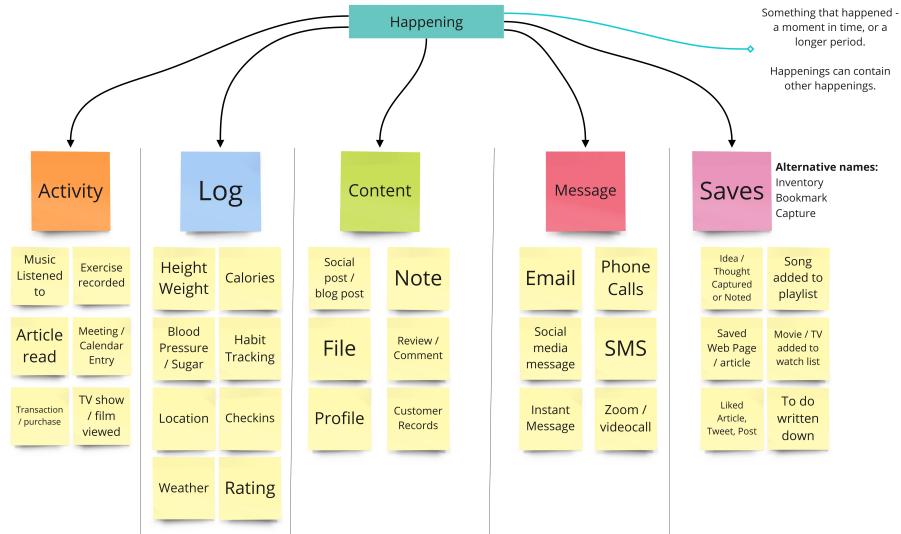


Figure 42: Figure X: Happenings

It can also be useful to model the different attributes of data in terms of what can be done to it.

We can imagine a simplified model of presenting information to users:

7.4.3.3 Opportunity 8: Algorithmic Meaning Extraction and Learning Systems

Helps with: engagement/efforts

Key idea to share: that the system should try and automatically associate data to entities. also: calendar/contact as start point. conjecture and assertion to reduce effort. learning, correcting, like an assistant [ref] word2vec as an example of the sophistication of what is being done and harnessed for provider purposes[REF]

[citation: diagram by Alex Peysakhovic, from CS 4803 / 7643: Deep Learning Guest Lecture: Embeddings and word2vec, a guest lecture at Georgia Tech by Facebook AI research engineer Ledell Wu, given Feb 18th 2020. https://www.cc.gatech.edu/classes/AY2020/cs7643_spring/slides/L13_EMBEDDING_WORLD2VEC_FINAL_VERSION.pdf, archived at https://web.archive.org/web/20211018015836/https://www.cc.gatech.edu/classes/AY2020/cs7643_spring/slides/L13_EMBEDDING_WORLD2VEC_FINAL_VERSION.pdf]

one possible flow of how to identify data

show how attributes types etc (ref back to earlier diagrams) can be detected:

What can we let users do with data? Different types of data have different properties...



Figure 43: Figure X: Attributes of Data

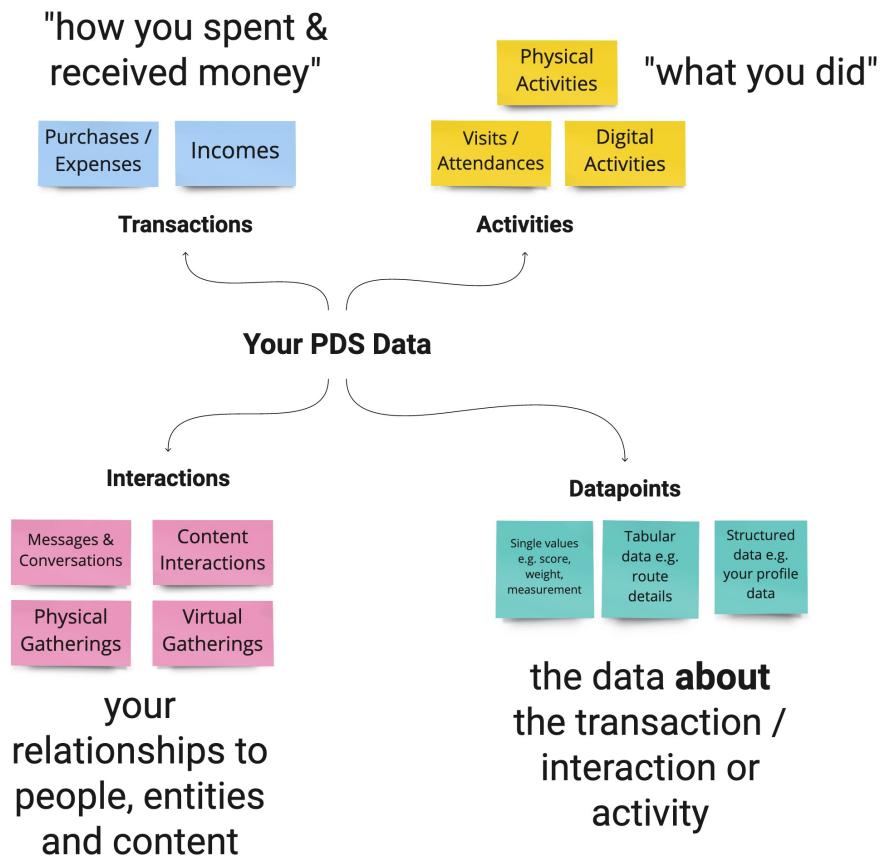


Figure 44: Figure X: Simple Life Information Presentation Model

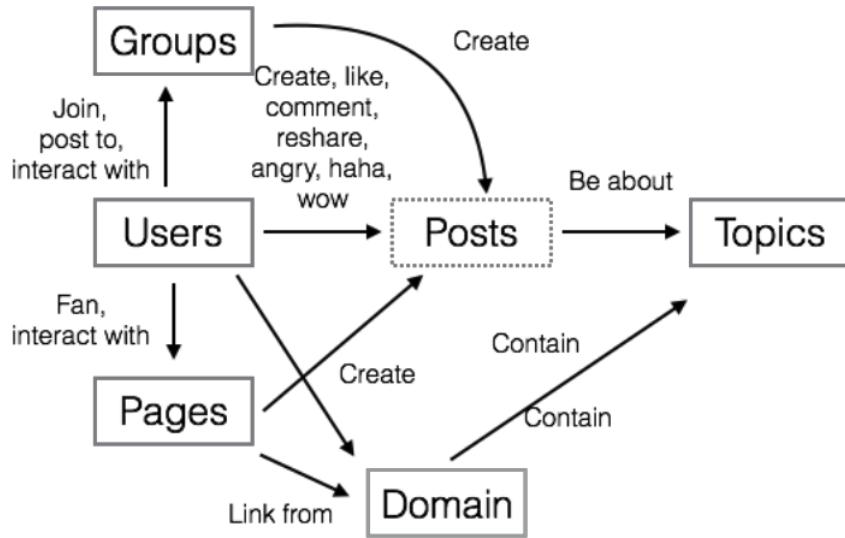


Figure 45: Figure X: A design for Facebook’s world2vec model, semantically modelling data from social media posts on Facebook

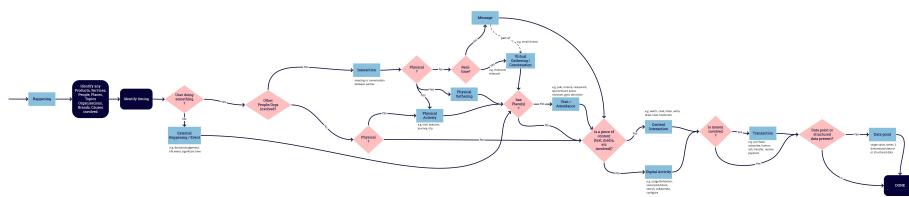


Figure 46: Figure X: Determining the nature of a piece of data

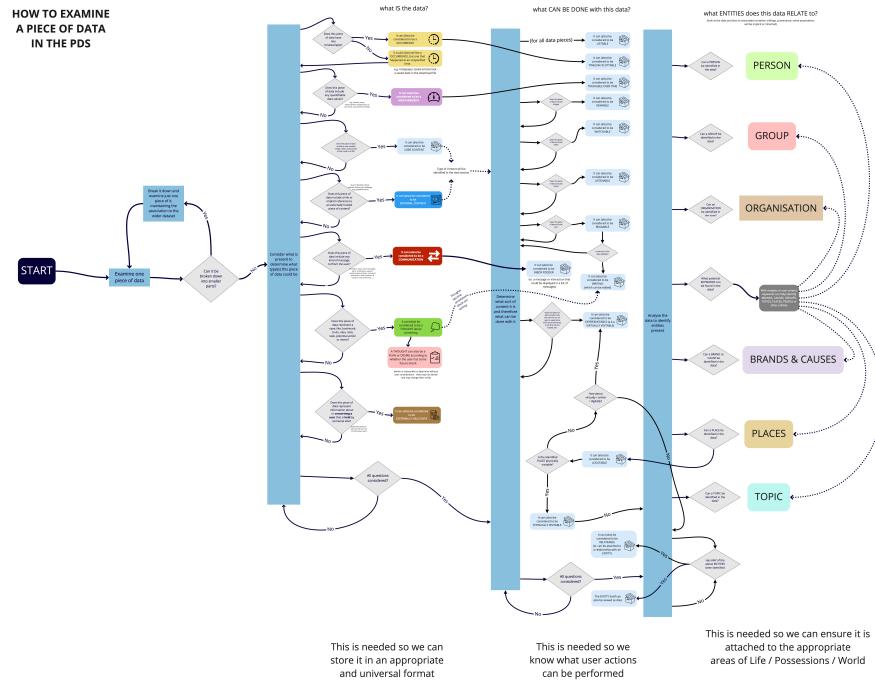


Figure 47: Figure X: Identifying the attributes of data

show how different types of entity can be identified. important to establish associations.

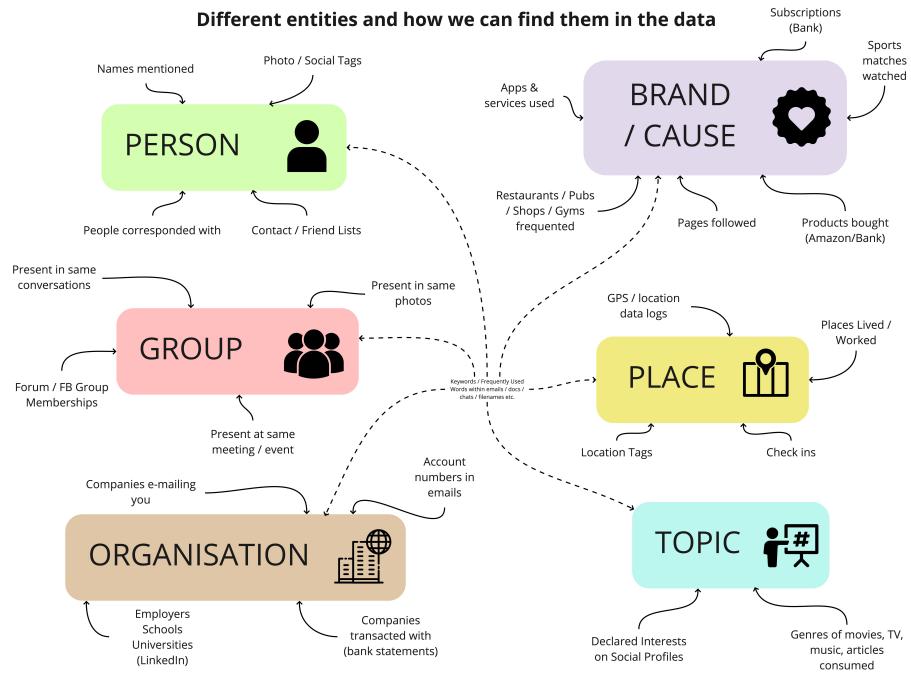
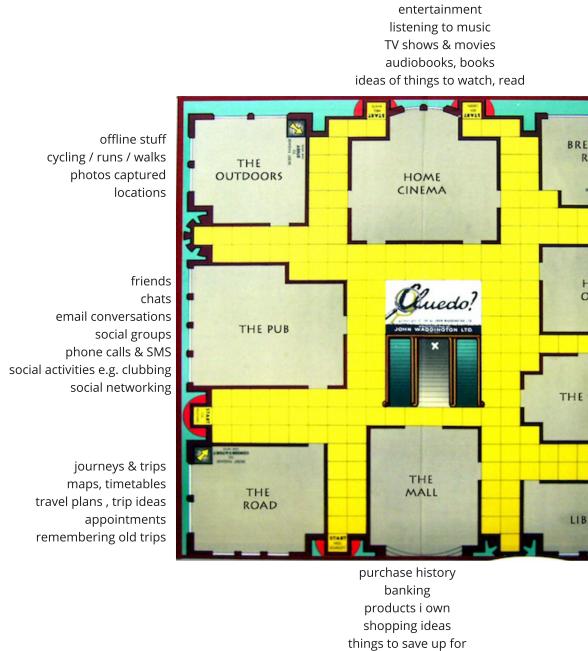


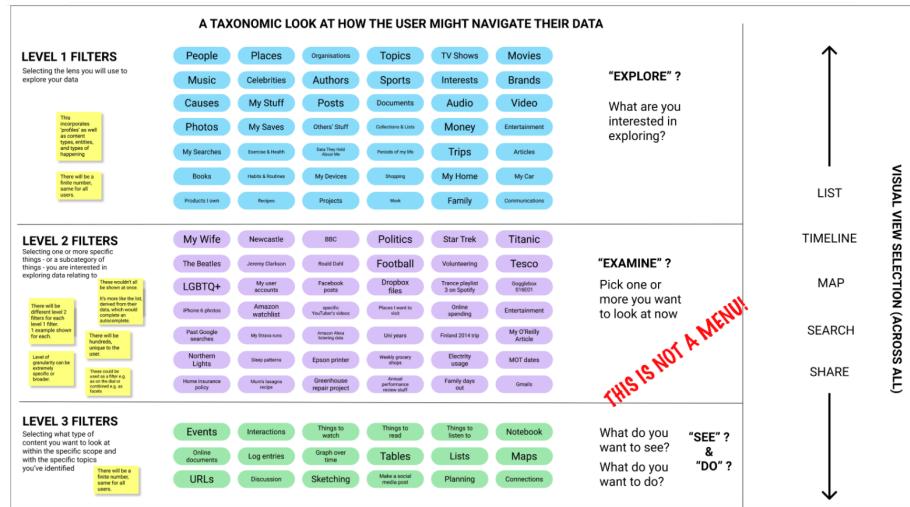
Figure 48: Figure X: Identifying entities in data

7.4.4 Create New Information Interfaces



7.4.4.1 Opportunity 8: Life Information interfaces (add credit for cluedo board)

Life partitioning would allow conceptually the user to navigate information according to what semantic concepts it is or relates to.



(add credit for cluedo board)

A mockup of how this might look in a user interface (ignore first frame for now)

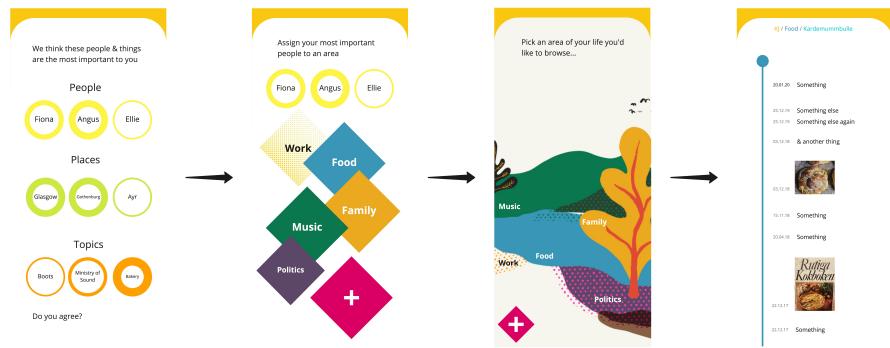


Figure 49: Figure X: Browsing By Areas of Life

A mockup of a life interface dashboard (by Alex Ballantyne)

The screenshot shows the Zero app's dashboard with several sections:

- Insights:** Includes a summary card for Amazon movie genre (Adventure) and Netflix viewing habits ("Netflix believes you love to binge Fantasy").
- Health Profile:** Shows Step count (4,600 to go), Sleep (8h 22min, 8h 27min active target), and Calories (1123.01, 1090.00 to go).
- Financial Profile:** Displays Balance (£1277 across all key accounts) and Today (£36.00 to go).
- Savings:** Shows £623.45 / £800 target (41% of the way there).
- Your donated data:** Shows a card for the NHS asking for permission to use physical activity data for research.
- Media Profile:** Shows 1h 43min today, with breakdowns for Netflix, Spotify, BBC Sounds, and BBC iPlayer.
- Social Profile:** Shows daily average on social media platforms: 2h 12m (target: 4h 53m).
- How about trying...**: A card for the Zero app, encouraging users to track their carbon impact.
- Activities:** Shows activity data for Monzo, Spotify, TFL, and Netflix.
- From your archive:** Highlights from this time last year (05/02/2020), showing "Smash miles" (56 miles), "Sorry miles" (13 miles), and "Pain miles" (121 miles).
- Delete my data:** A button to delete all data, with a warning message below it.

(to do: make this fit on a portrait page)

7.4.4.2 Opportunity 9: Ecosystem Detection & Visualisation Key idea: detecting the ecosystem. Example: the subscription detector.

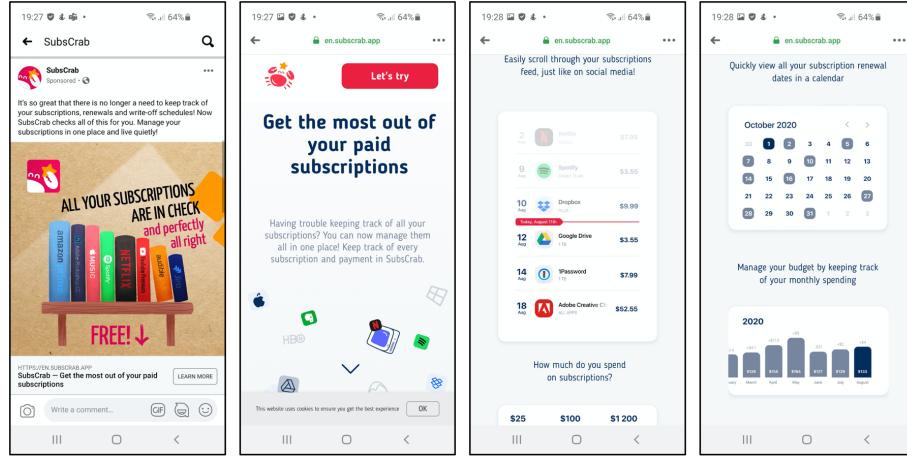


Figure 50: Figure X: Subscrab: An example application for ecosystem detection and visualisation

7.4.4.3 Opportunity 10: Inclusive Information Flows Key idea: Rivers of flowing info. Including people.

7.4.5 Create New Capabilities

7.4.5.1 Opportunity 11: Exploratory Actions & Asking Tools It is important to think about the capabilities people will have (expand on and map this back to all the PIM calls in 2.2.2)

refer back to dashboard mockup & data needs to be interrogable and malleable. What if there aren't visualisations for your questions. But also, visualisations raise questions

Here are a series of data cards we used in a user research activity at BBC. (initially conceived by me, condensed and reworked by myself in collaboration with Chris Gameson)

Key idea: verbs. types of question. asking tools not just predetermined insights.

7.4.5.2 Opportunity 12: Self Profiling & Curating your Digital Self Key idea: self profiling (a la BBC) but also more generally, the "I want a bicycle" VRM idea (unless already covered)

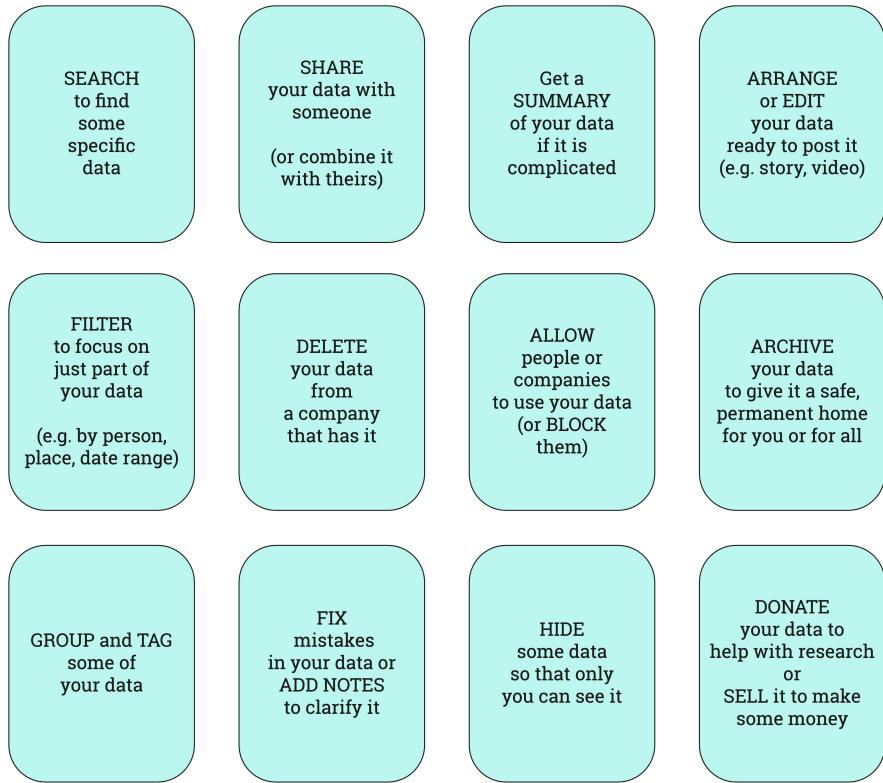


Figure 51: Figure X: Data actions

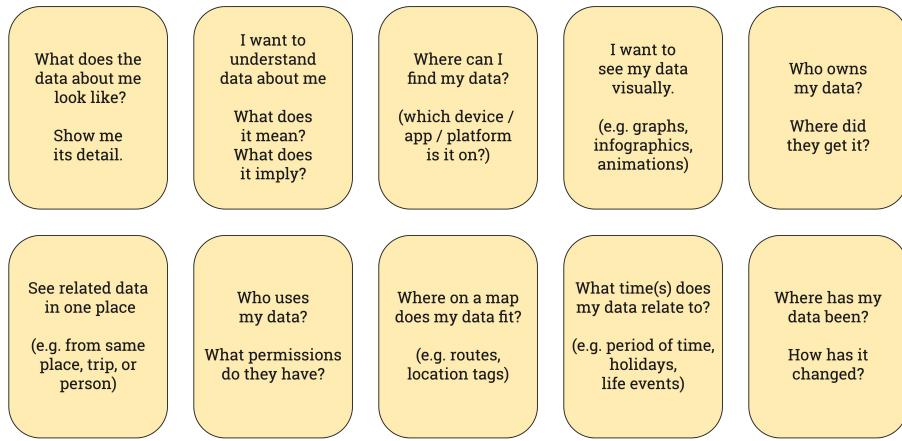


Figure 52: Figure X: Data questions

7.4.6 Influence & Motivate

7.4.6.1 Opportunity 13: Regulating the Information Landscape Key point: regulation of *landscape* -> new moves DSA maybe also some ref back to end C5 and to the GDPR Guidelines

7.4.6.2 Opportunity 14: Information Unification and Schematisation Key idea: generalised types of data (refer back to relevant model above) [reference to Solid Shapes etc]

7.4.6.3 Opportunity 15: Life/Ecosystem Information as Boundary Objects (Multi-stakeholder Design) [POSSIBLY CUT THIS ONE]

Key idea: Data as bringing different people together , ref living lab. also use of cards in my research and at BBC (and ref Urquhart?)

7.4.6.4 Opportunity 16: The Business Value of Transparency and Human Centricity key idea: tackling resistance reducing liability, improved consent . less waste cost on broadcast advertising (could express it as a development of all the ad personalisation today). selling the benefits.

7.4.7 Educate & Empower

7.4.7.1 Opportunity 17: Life Information & Personal Data Ecosystem Literacy Key point: define it, distinguish it from technical skills/literacy as well as from numbercrunching Literacy

7.4.7.2 Opportunity 18: Individual Discovery: Mapping Your Personal Data Ecosystem Key point: empowering individuals as investigators. Can help them with tools or learning programs.

7.5 Thesis Conclusion

[reiterate the answer to the question - the key 4 roles, 3 capabilities and N approaches needed for better human data relations] [how to group together the approaches]

[clarify the contribution of the thesis, with backreferences - 2 case studies, RQ answers, and the HDR roadmap]

[highlight future value/societal implications of the work]

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.
- Abiteboul, S., André, B. and Kaplan, D. (2015) *Managing your digital life with a Personal information management system*. 5. ACM, pp. 32–35. doi: 10.1145/2670528.
- ‘About The Quantified Self’ (no date). Available at: <https://quantifiedself.com/about/what-is-quantified-self/> (Accessed: 22 March 2021).
- Abowd, G. D. et al. (1999) ‘Towards a better understanding of context and context-awareness’, in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, pp. 304–307. doi: 10.1007/3-540-48157-5_29.
- Abowd, G. D. (2012) ‘What next, ubicomp?: celebrating an intellectual disappearing act’, in *Proceedings of the 2012 ACM conference on ubiquitous computing*. New York, New York, USA: ACM Press, pp. 31–40. doi: <http://dx.doi.org/10.1145/2370216.2370222>.
- Abowd, G. D. and Mynatt, E. D. (2000) *Charting Past, Present, and Future Research in Ubiquitous Computing*. 1, pp. 29–58. Available at: <https://www.cc.gatech.edu/fce/pubs/abowd-mynatt-tochi-millenium.pdf>.
- 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).
- Alizadeh, F. et al. (2019) ‘GDPR-reality check on the right to access data’, in *ACM international conference proceeding series*. New York, New York, USA: ACM Press, pp. 811–814. doi: 10.1145/3340764.3344913.
- ‘AllofMe Company Profile’ (2007). Available at: <https://www.crunchbase.com/organization/allofme> (Accessed: 23 March 2021).
- ‘AllofMe.com Teaser Clip’ (2008). YouTube. Available at: <https://www.youtube.com/watch?v=JWyqt4WL6xE> (Accessed: 21 March 2021).
- Andrews, R. (2005) ‘GTD : A New Cult for the Info Age’, *Wired*. Available at: <https://www.wired.com/2005/07/gtd-a-new-cult-for-the-info-age/>.
- Apple (2009) ‘iPhone 3G Commercial: "There's an app for that"’. YouTube. Available at: <https://www.youtube.com/watch?v=mFlITzqRBWY>.
- Arfelt, E., Basin, D. and Debois, S. (2019) ‘Monitoring the GDPR’, in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, pp. 681–699. doi: 10.1007/978-3-030-29959-0_33.
- ‘Article 13: Information to be provided where personal data are collected from the data subject’ (2020). doi: 10.1093/oso/9780198826491.003.0044.
- Article 29 Data Protection Working Party (2010) *Opinion 3/2010 on the principle of accountability*. Brussels: European Commission.
- Aslam, H. et al. (2016) ‘Harnessing Smartphones as a Personal Informatics Tool towards Self-Awareness and Behavior Improvement’, *Proceedings - 2016 IEEE 14th International Conference on Dependable, Autonomic and Secure Computing, DASC 2016, 2016 IEEE 14th International Conference on Perva-*

- sive 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.
- Ausloos, J. (2019) ‘GDPR Transparency as a Research Method’, *SSRN Electronic Journal*, (May), pp. 1–23. doi: 10.2139/ssrn.3465680.
- Ausloos, J. and Dewitte, P. (2018) *Shattering one-way mirrors-data subject access rights in practice*. Available at: www.irissproject.eu https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3106632.
- Ausloos, J. and Veale, M. (2020) ‘Researching with Data Rights’, *Technology and Regulation*, pp. 136–157.
- Bakardjieva, M. and Feenberg, A. (2001) ‘Involving the Virtual Subjects’, *Ethics and Information Technology*, 2, pp. 233–240. doi: 10.1023/A:1011454606534.
- Baker, J. (2018) ‘What’s a GDPR complaint? No one really knows’. Available at: <https://iapp.org/news/a/whats-the-definition-of-a-gdpr-complaint-spoiler-alert-no-one-knows/>.
- 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.
- Barbosa Neves, B. and Casimiro, C. (2018) *Connecting Families?: Information & Communication Technologies, Generations, and the Life Course*. Policy Press.
- Barreau, D. K. (1995) ‘Context as a factor in personal information management systems’, *Journal of the American Society for Information Science*, 46(5), pp. 327–339. doi: 10.1002/(SICI)1097-4571(199506)46:5<327::AID-ASI4>3.0.CO;2-C.
- Barreau, D. and Nardi, B. A. (1995) ‘Finding and reminding’, *ACM SIGCHI Bulletin*, 27(3), pp. 39–43. doi: 10.1145/221296.221307.
- Bate, A. and Bellis, A. (2018) *The Troubled Families programme (England)*. July.
- Battarbee, K. and Koskinen, I. (2005) ‘Co-experience: user experience as interaction’, *CoDesign*. Taylor & Francis, 1(1), pp. 5–18.
- Baxter, G. and Sommerville, I. (2011) ‘Socio-technical systems: From design methods to systems engineering’, *Interacting with Computers*. OUP, 23(1), pp. 4–17. doi: 10.1016/j.intcom.2010.07.003.
- BBC 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.
- Bergman, O. et al. (2008) ‘Improved search engines and navigation preference in personal information management’, *ACM Transactions on Information Systems*, 26(4). doi: 10.1145/1402256.1402259.
- Bergman, O. et al. (2012) ‘How do we find personal files?: The effect of OS, presentation & depth on file navigation’, *Conference on Human Factors in*

- Computing Systems - Proceedings*, pp. 2977–2980. doi: 10.1145/2207676.2208707.
- Bergman, O. (2013) ‘The Effect of Folder Structure on Personal File Navigation’, *Journal of the American Society for Information Science and Technology*, 64(July), pp. 1852–1863. doi: 10.1002/asi.
- Bergman, O., Beyth-Marom, R. and Nachmias, R. (2003) ‘The user-subjective approach to personal information management systems’, *Journal of the American Society for Information Science and Technology*, 54(9), pp. 872–878. doi: 10.1002/asi.10283.
- Berners-Lee, T., Hendler, J. and Lassila, O. (2001) ‘The Semantic Web’, *Scientific American*, 284(5), pp. 34–43. Available at: <https://jstor.org/stable/10.2307/26059207>.
- Binns, R. (2022) ‘Tracking on the Web, Mobile and the Internet-of-Things’. Available at: <http://arxiv.org/abs/2201.10831>.
- Bjerknes, Gro. et al. (1987) *Computers and democracy : a Scandinavian challenge*. Aldershot [Hants, England]; Brookfield [Vt.], USA: Avebury, p. 434. Available at: <http://www.worldcat.org/title/computers-and-democracy-a-scandinavian-challenge/oclc/614994092?referer=di&ht=edition>.
- Björgvinsson, E., Ehn, P. and Hillgren, P.-A. (2010) ‘Participatory design and “democratizing innovation”’, in *Proceedings of the 11th biennial participatory design conference*, pp. 41–50.
- Bødker, S. (2006) ‘When second wave HCI meets third wave challenges’, *ACM International Conference Proceeding Series*, 189(October), pp. 1–8. doi: 10.1145/1182475.1182476.
- Bødker, S. (2015) ‘Third-wave HCI, 10 years later—participation and sharing’, *Interactions*, 22(5), pp. 24–31. doi: 10.1145/2804405.
- Boud, D., Keogh, R. and Walker, D. (1985) *Reflection: Turning experience into learning*. Routledge.
- Bowker, G. C. (2005) *Memory practices in the sciences*. MIT Press, p. 261.
- Bowker, G. C. et al. (2015) *Boundary objects and beyond : working with Leigh Star*. MIT Press, p. 548. Available at: <https://books.google.co.uk/books?hl=en&lr=&id=nmSkCwAAQBAJ&oi=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. (2012) ‘We deserve the time and space to be human’, in Gupta, V. (ed.) *The Future We Deserve*. PediaPress GmbH. Available at: <https://eprints.ncl.ac.uk/273829>.
- Bowyer, A. (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.
- Bowyer, A. et al. (2019) ‘Human-data interaction in the context of care: Co-designing family civic data interfaces and practices’, in *Conference on human factors in computing systems - proceedings*. doi: 10.1145/3290607.3312998.

- Bowyer, A. (2021) ‘Human-Data Interaction has two purposes: Personal Data Control and Life Information Exploration’. Available at: <https://eprints.ncl.ac.uk/273832#>.
- Bowyer, A., Pidoux, J., et al. (2022) *Digipower technical reports: Auditing the data economy through personal data access*. doi: 10.5281/zenodo.6554177.
- Bowyer, A., Holt, J., et al. (2022) ‘Human-GDPR interaction : Practical experiences of accessing personal data’, *CHI ’22*.
- Brandt, E. and Messeter, J. (2004) ‘Facilitating collaboration through design games’, in *Proceedings of the eighth conference on participatory design artful integration: Interweaving media, materials and practices - PDC 04*. New York, New York, USA: ACM Press, p. 121. doi: 10.1145/1011870.1011885.
- Braun, V. and Clarke, V. (2006) ‘Using thematic analysis in psychology’, *Qualitative Research in Psychology*. Taylor & Francis, 3(2), pp. 77–101. doi: 10.1191/1478088706qp063oa.
- Brest, P. (2010) ‘The Power of Theories of Change’, *Stanford Social Innovation Review*, 8(2), pp. 47–51.
- Bridle, J. (2016) ‘Algorithmic Citizenship, Digital Statelessness’, *GeoHumanities*. James Bridle, 2(2), pp. 377–381. doi: 10.1080/2373566x.2016.1237858.
- Brooks, D. (2013) ‘The Philosophy of Data’. Available at: <https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html>.
- Brown, D. (2015) ‘Here’s what ‘fail fast’ really means’. Available at: <https://venturebeat.com/2015/03/15/heres-what-fail-fast-really-means/>.
- Brynjolfsson, E. and Oh, J. H. (2012) ‘The attention economy: Measuring the value of free digital services on the internet’, *International Conference on Information Systems, ICIS 2012*, 4, pp. 3243–3261.
- Bufalieri, L. et al. (2020) ‘GDPR: When the right to access personal data becomes a threat’. doi: 10.1109/icws49710.2020.00017.
- Bunge, M. (1999) *Social Science Under Debate: A Philosophical Perspective*. University of Toronto Press. Available at: <https://books.google.co.uk/books?id=MLjZzJLbpkC>.
- Burgess, M. (2021) ‘Why Amazon’s £636m GDPR fine really matters’, *Wired*. Available at: <https://www.wired.co.uk/article/amazon-gdpr-fine>.
- Burkeman, O. (2011) ‘SXSW 2011: The internet is over’. Available at: <https://www.theguardian.com/technology/2011/mar/15/sxsw-2011-internet-online> (Accessed: 23 March 2021).
- Bush, V. (1945) ‘As we may think’, *The Atlantic Monthly*, 3(2), pp. 35–46. doi: 10.1145/227181.227186.
- Cambridge Dictionary (no date) ‘Empowerment’. Available at: <https://dictionary.cambridge.org/dictionary/english/empowerment>.
- Campbell, P. L. (2011) ‘Peirce, pragmatism, and the right way of thinking’, *Sandia National Laboratories, Albuquerque*. Citeseer.
- Carter, J. (2015) ‘Who are the digital disruptors redefining entire industries?’ Available at: <https://www.techradar.com/uk/news/world-of-tech/who-are-the-digital-disruptors-redefining-entire-industries-1298171> (Accessed: 23 March 2021).
- Caruthers, M. (2018) ‘World Password Day: How to Improve Your Passwords’.

- Available at: <https://blog.dashlane.com/world-password-day/> (Accessed: 5 May 2021).
- Cavoukian, A. (2010) 'Privacy by design: the definitive workshop. A foreword by Ann Cavoukian, Ph.D', *Identity in the Information Society*, 3(2), pp. 247–251. doi: 10.1007/s12394-010-0062-y.
- Cavoukian, A. (2012) 'Privacy by Design and the Emerging Personal Data Ecosystem', (October), pp. 1–39.
- Chan, R. (2019) 'The cambridge analytica whistleblower explains how the firm used facebook data to sway elections', *Business Insider*. Available at: <https://www.businessinsider.com/cambridge-analytica-whistleblower-christopher-wylie-facebook-data-2019-10?r=US&IR=T>.
- Chang, A. (2018) 'The Facebook and Cambridge Analytica scandal, explained with a simple diagram - Vox'. Available at: <https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram>.
- Cheetham, M. et al. (2018) 'Embedded research: A promising way to create evidence-informed impact in public health?', *Journal of Public Health (United Kingdom)*. Oxford University Press, 40(suppl_1), pp. i64–i70. doi: 10.1093/pubmed/fdx125.
- Chevalier, J. M. and Buckles, D. J. (2008) *SAS2: A guide to collaborative inquiry and social engagement*. SAGE Publishing India.
- 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.
- Chung, C. F. et al. (2016) 'Boundary negotiating artifacts in personal informatics: Patient-provider collaboration with patient-generated data', *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 27, pp. 770–786. doi: 10.1145/2818048.2819926.
- CitizenMe (2021) 'Become a Citizen and unlock the value of your data'. Available at: <https://www.citizenme.com/for-citizens/> (Accessed: 23 August 2021).
- Claburn, T. (2021) 'Android's Messages, Dialer apps quietly sent text, call info to Google'. Available at: https://www.theregister.com/2022/03/21/google_messages_gdpr/.
- Clarke, N. et al. (2019) 'GDPR: an impediment to research?', *Irish Journal of Medical Science (1971-)*. Springer, 188(4), pp. 1129–1135.
- Cogran, P. and Kinsley, S. (2012) 'Paying Attention: towards a critique of the attention economy', *Cultural Machine*, 13.
- Collins English Dictionary (no date a) 'Useability'. Available at: <https://www.thefreedictionary.com/useability>.
- Collins English Dictionary (no date b) 'Useable'. Available at: <https://www.thefreedictionary.com/useable>.
- Comandè, G. and Schneider, G. (2021) 'Can the GDPR make data flow for research easier? Yes it can, by differentiating! A careful reading of the GDPR shows how EU data protection law leaves open some significant flexibilities for

- data protection-sound research activities', *Computer Law & Security Review*. Elsevier, 41, p. 105539.
- Connected Health Cities (2017) 'SILVER Project: Smart Interventions for Local Residents'. Available at: <https://www.connectedhealthcities.org/research-projects/troubled-families/> (Accessed: 14 May 2021).
- Copeland, E. (2015) *Small Pieces Loosely Joined: How smarter use of technology and data can deliver real reform of local government*. Available at: www.policyexchange.org.uk <https://policyexchange.org.uk/publication/small-pieces-loosely-joined-how-smarter-use-of-technology-and-data-can-deliver-real-reform-of-local-government/>.
- Cormack, A. (2016) 'Is the Subject Access Right Now Too Great a Threat to Privacy?', *European Data Protection Law Review*, 2(1), pp. 15–27. doi: 10.21552/edpl/2016/1/5.
- Cormack, A. (2021) 'Thinking with GDPR: A guide to better system design', *Information Services & Use*, 41(1-2), pp. 61–69. doi: 10.3233/isu-210107.
- Cornford, J., Baines, S. and Wilson, R. (2013) 'Representing the family: how does the state 'think family'?', *Policy & Politics*, 41(1), pp. 1–19. doi: 10.1332/030557312X645838.
- Corra, M. and Willer, D. (2002) 'The gatekeeper', *Sociological Theory*. SAGE Publications Sage CA: Los Angeles, CA, 20(2), pp. 180–207.
- Coughlan, T., Leder Mackley, K., et al. (2013) 'Current issues and future directions in methods for studying technology in the home', *PsychNology Journal*, 11(2), pp. 159–184.
- Coughlan, T., Brown, M., et al. (2013) 'Methods for studying technology in the home', in *CHI'13 extended abstracts on human factors in computing systems*, pp. 3207–3210.
- Council of the European Union (2015) 'Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)'. Brussels. Available at: <http://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf>.
- Crabtree, A. and Mortier, R. (2016) 'Personal Data, Privacy and the Internet of Things: The Shifting Locus of Agency and Control', *SSRN Electronic Journal*, pp. 1–20. doi: 10.2139/ssrn.2874312.
- Crabtree, A. and Tolmie, P. (2018) 'The practical politics of sharing personal data', in *Personal and Ubiquitous Computing*. Springer-Verlag (2), pp. 293–315. doi: 10.1007/s00779-017-1071-8.
- Crivellaro, C. et al. (2019) 'Not-equal: Democratizing research in digital innovation for social justice', *Interactions*, 26(2), pp. 70–73. doi: 10.1145/3301655.
- Croll, A. (2009) 'The Three Economies of Online Currency'. Available at: <https://solveforinteresting.com/the-three-currencies-of-the-online-economy/>.
- 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/.

- ‘Data Brokers: Everything You Need to Know’ (2022). Available at: <https://www.avast.com/c-data-brokers> (Accessed: 11 February 2022).
- ‘datacy - About Us’ (no date). Available at: <https://www.datacy.com/personal/about-us> (Accessed: 22 March 2019).
- 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.
- Dehaye, P.-O. (2018) ‘Post-hearing questions by Senator Blumenthal to Mark Zuckerberg’. Available at: <https://wiki.personaldata.io/wiki/Item:Q1800>.
- Dehaye, P.-O. (2021) ‘HestiaLabs’. Geneva, Switzerland. Available at: <https://hestialabs.org/en/> (Accessed: 23 August 2021).
- ‘Delicious’ (2003). Available at: [https://en.wikipedia.org/wiki/Delicious_\(website\)](https://en.wikipedia.org/wiki/Delicious_(website)).
- Department for Education (2018) *Working Together to Safeguard Children*. March, p. 393. doi: 10.1080/13561820020003919.
- 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>.
- ‘Digi.me’ (no date). Available at: <https://digi.me/> (Accessed: 23 August 2021).
- Dijck, J. van (2014) ‘Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology’, *Surveillance and Society*. Surveillance Studies Network, 12(2), pp. 197–208. doi: 10.24908/ss.v12i2.4776.
- DiSalvo, C. (2010) ‘Design, Democracy and Agonistic Pluralism’, *Proceedings of the Design Research Society Conference 2010*, pp. 366–371.
- DiSalvo, C. (2012) *Adversarial Design*. MIT Press (Design thinking, design theory). doi: 10.7551/mitpress/8732.003.0007.
- Dourish, P. et al. (2000) ‘Extending document management systems with user-specific active properties’, *ACM Transactions on Information Systems*, 18(2), pp. 140–170. doi: 10.1145/348751.348758.
- Dourish, P. (2001) *Where the action is: the foundations of embodied interaction*. MIT press.
- Dourish, P. (2003) ‘The appropriation of interactive technologies: Some Lessons From Placeless Documents’, *Computer Supported Cooperative Work*, 12(4), pp. 465–490.
- Dourish, P. (2004) ‘What we talk about when we talk about context’, *Personal and Ubiquitous Computing*, 8(1), pp. 19–30. doi: 10.1007/s00779-003-0253-8.
- Edwards, A. and Elwyn, G. (2006) ‘Inside the black box of shared decision

- making: distinguishing between the process of involvement and who makes the decision', *Health Expectations*. Wiley/Blackwell (10.1111), 9(4), pp. 307–320. doi: 10.1111/j.1369-7625.2006.00401.x.
- 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.
- Engelbart, D. C. (1962) 'Augmenting human intellect: A conceptual framework'. Menlo Park, CA, USA: Stanford Research Institute.
- Es, M. van, Guijt, I. and Vogel, I. (2015) 'Hivos ToC Guidelines: Theory of Change Thinking in Practice'. The Hague, The Netherlands: Hivos.
- 'Ethi' (no date). Available at: <https://www.ethi.me/>.
- Etzel, B. (1995) 'New strategy and techniques to cope with information overload', in *IEE colloquium on information overload*. IEE (223), pp. 2–2. doi: 10.1049/ic:19951427.
- European Commission (2014) *Research and Innovation in the field of ICT for Health, Wellbeing and Ageing Well: an overview*, p. 39.
- European Union Agency for Fundamental Rights (2020) 'Your Rights Matter: Data Protection and Privacy 2020', p. 20. doi: 10.2811/031862.
- Evans, W. (2021) 'Amazon's dark secret: It has failed to protect your data'.
- 'Exist.io' (no date). Available at: <https://exist.io/> (Accessed: 23 August 2021).
- 'Explainable AI: Making machines understandable for humans' (no date). Available at: <https://explainableai.com/> (Accessed: 16 June 2022).
- 'Facebook - Data Policy' (no date). Available at: <https://www.facebook.com/about/privacy> (Accessed: 9 August 2021).
- 'Facebook–Cambridge Analytica Data Scandal' (2014). Available at: https://en.wikipedia.org/wiki/Facebook\textendashCambridge_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.
- Field, F. (2010) *The Foundation Years: preventing poor children becoming poor adults*. Available at: [www.frankfield.co.uk http://www.inspiredbybabies.org.uk/Page2NationalrelevantDocsresources/Field%20Preventing%20poor%20children%20becoming%20poor%20adults%202011.pdf](http://www.inspiredbybabies.org.uk/Page2NationalrelevantDocsresources/Field%20Preventing%20poor%20children%20becoming%20poor%20adults%202011.pdf)
- 'Finland: Broadband Access Made Legal Right In Landmark Law' (2010). Available at: https://www.huffpost.com/entry/finland-broadband-access_n_320481 (Accessed: 23 March 2021).
- Firth, E. (2019) 'Personal data has value in so many different ways'. digi.me. Available at: <https://blog.digi.me/2019/09/04/personal-data-has-so-much-more-value-than-pure-cash/>.
- Foucault, M. (1975) 'Discipline and punish: The birth of the prison', *New York*. Pantheon Books.
- Foulonneau, Muriel. and Riley, Jenn. (2008) *Metadata for digital resources : implementation, systems design and interoperability*. Chandos Pub, p. 203.

- Fowler, M. and Highsmith, J. (2001) 'The agile manifesto', *Software Development*. [San Francisco, CA: Miller Freeman, Inc., 1993-, 9(8), pp. 28–35.
- Freeman, E. and Gelernter, D. (1996) 'Lifestreams: A Storage Model for Personal Data', *SIGMOD Record (ACM Special Interest Group on Management of Data)*. Association for Computing Machinery (ACM), 25(1), pp. 80–86. doi: 10.1145/381854.381893.
- Friedman, B. and Hendry, D. G. (2019) *Value Sensitive Design: Shaping Technology with Moral Imagination*. MIT Press (The MIT press). Available at: <https://books.google.co.uk/books?id=8ZiWDwAAQBAJ>.
- Friedman, R. L. (2006) 'Deweyan Pragmatism', *William James Studies*, 1. Available at: <https://williamjamesstudies.org/deweyan-pragmatism/>.
- Frost, A. (2019) 'Forget Folders: The Best Ways to Organize Your Files with Tags and Labels'. Available at: <https://zapier.com/blog/how-to-use-tags-and-labels/>.
- Fu, S. et al. (2020) 'Social media overload, exhaustion, and use discontinuance: Examining the effects of information overload, system feature overload, and social overload', *Information Processing and Management*, 57(6). doi: 10.1016/j.ipm.2020.102307.
- Gelernter, D. (1994) 'The cyber-road not taken: Lost on the info-highway? Here's some stuff that could really change your life.', *The Washington Post*, 3.
- Gellman, B. (2013) 'Edward Snowden, after months of NSA revelations, says his mission's accomplished', *The Washington Post*, 23. Available at: http://www.washingtonpost.com/world/national-security/edward-snowden-after-months-of-nsa-revelations-says-his-missions-accomplished/2013/12/23/49fc36de-6c1c-11e3-a523-fe73f0ff6b8d_story.html%5Cnhttp://www.washingtonpost.com/world/national-security/edward-
- Gemmell, J., Bell, G. and Lueder, R. (2006) 'MyLifeBits: A personal database for everything', *Communications of the ACM*, 49(1), pp. 88–95. doi: 10.1145/1107458.1107460.
- Gener8 (2021) 'Gener8'. Available at: <https://gener8ads.com/> (Accessed: 23 August 2021).
- Gillespie, T. and Seaver, N. (2016) 'Critical Algorithm Studies - A Reading List'. Available at: <https://socialmediacollective.org/reading-lists/critical-algorithm-studies/>.
- Gitelman, L. (2013) *Raw data is an oxymoron*. Edited by Lisa Gitelman. MIT Press, p. 182. Available at: <https://mitpress.mit.edu/books/raw-data-oxymoron>.
- Glavic, B. et al. (2021) 'Trends in Explanations: Understanding and Debugging Data-driven Systems', *Foundations and Trends®in Databases*. Now Publishers, Inc., 11(3), pp. 226–318. doi: 10.1561/XXXXXXX.Boris.
- Goffe, L. et al. (2021) 'Appetite for disruption: Designing human-centred augmentations to an online food ordering platform', in *34th british HCI conference 34*, pp. 155–167.
- Goffe, L. et al. (2022) 'Web augmentation for well-being: The human-centred design of a takeaway food ordering digital platform', *Interacting with Computers*.
- Golembewski, M. and Selby, M. (2010) 'Ideation decks', in *Proceedings of the 8th ACM conference on designing interactive systems - DIS '10*. New York, New

- York, USA: ACM Press, p. 89. doi: 10.1145/1858171.1858189.
- Gonscherowski, S. and Bieker, F. (2018) ‘Who You Gonna Call When There’s Something Wrong in Your Processing? Risk Assessment and Data Breach Notifications in Practice’, in *IFIP international summer school on privacy and identity management*. Springer, pp. 35–50.
- ‘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.
- 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.
- Hall, W., Tinati, R. and Jennings, W. (2018) ‘From brexit to trump: Social media’s role in democracy’, *Computer*. IEEE, 51(1), pp. 18–27.
- Hamon, R. et al. (2021) ‘Impossible Explanations? Beyond explainable AI in the GDPR from a COVID-19 use case scenario’, in *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pp. 549–559.
- Harbird, R. (2006) ‘Novel Applications for Information Technology in Risk Assessment for Children’s Social Care in the UK’, *Rn*. Available at: http://www.cs.ucl.ac.uk/research/researchnotes/documents/RN_06_11.pdf.
- Harris, T. (2013a) ‘A Call to Minimize Distraction Respect Users’ Attention’. Available at: <http://www.minimizeddistraction.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%E2%80%8A-%E2%80%8Afrom-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.
- Hayes, G. R. (2011) ‘The relationship of action research to human-computer interaction’, *ACM Transactions on Computer-Human Interaction*, 18(3), pp. 1–20. doi: 10.1145/1993060.1993065.
- ‘HDI Lab, Heerlen’ (2020). Available at: <https://hdilab.com/>.
- ‘HDI Network Plus, University of Glasgow’ (2018). Available at: <https://hdinetwork.org/>.
- Hemp, P. (2009) ‘Death by Information Overload’. Available at: <https://hbr.org/2009/09/death-by-information-overload> (Accessed: 23 March 2021).
- Henderson, I. and Group, B. W. (2020) ‘Customer — Supplier Engagement Framework Explained’, pp. 1–7. Available at: <https://me2ba.org/wp-content/uploads/2020/09/customer-supplier-engagement-framework-updated-9-28.pdf>.

- Hendler, J. and Berners-Lee, T. (2010) ‘From the Semantic Web to social machines: A research challenge for AI on the World Wide Web’. doi: 10.1016/j.artint.2009.11.010.
- Herselman, M. et al. (2016) ‘A Digital Health Innovation Ecosystem for South Africa’, in *2016 IST-africa conference, IST-africa 2016*. doi: 10.1109/ISTAFRICA.2016.7530615.
- Hixon, J. G. and Swann, W. B. (1993) ‘When Does Introspection Bear Fruit? Self-Reflection, Self-Insight, and Interpersonal Choices’, *Journal of Personality and Social Psychology*, 64(1), pp. 35–43. doi: 10.1037/0022-3514.64.1.35.
- ‘Hobson’s Choice’ (no date). Available at: <https://www.britannica.com/dictionary/Hobson%27s-choice>.
- Hoffman, W. (2010) ‘Rethinking Personal Data’. Available at: <https://web.archive.org/web/20110220013300/http://www.weforum.org/issues/rethinking-personal-data>.
- Hoffman, W. (2011) *Personal data : The emergence of a new asset class*. World Economic Forum, pp. 1–40. Available at: <http://www.weforum.org/reports/personal-data-emergence-new-asset-class>.
- Hoffman, W. (2013) *Unlocking the Value of Personal Data: From Collection to Usage Prepared in collaboration with The Boston Consulting Group Industry Agenda*. February. World Economic Forum.
- Hoffman, W. (2014a) *Rethinking Personal Data : A New Lens for Strengthening Trust*. May. World Economic Forum, p. 35. Available at: http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf.
- Hoffman, W. (2014b) *Rethinking personal data: Trust and context in user-centred data ecosystems*. May. World Economic Forum, p. 35. Available at: http://www3.weforum.org/docs/WEF_RethinkingPersonalData_TrustandContext_Report_2014.pdf.
- Honeyman, M., Dunn, P. and Mckenna, H. (2016) *A digital NHS?*
- Hoofnagle, C. J., Sloot, B. van der and Borgesius, F. Z. (2019) ‘The European Union general data protection regulation: What it is and what it means’, *Information and Communications Technology Law*. Taylor & Francis, 28(1), pp. 65–98. doi: 10.1080/13600834.2019.1573501.
- Hosch, W. L. (2017) ‘Web 2.0’. Available at: <https://www.britannica.com/topic/Web-20> (Accessed: 26 April 2021).
- Hotho, A., Nürnberg, A. and Paaß, G. (2005) ‘A brief survey of text mining’, in *Ldv forum*. Citeseer (1), pp. 19–62.
- Huberman, M. and Miles, M. B. (2002) *The qualitative researcher’s companion*. Sage.
- ‘Human Data Interaction Project at the Data to AI Lab, MIT’ (2015). Available at: <https://hdi-dai.lids.mit.edu/>.
- Human, S. and Cech, F. (2021) ‘A human-centric perspective on digital consenting: The case of GAFAM’, *Smart Innovation, Systems and Technologies*, 189, pp. 139–159. doi: 10.1007/978-981-15-5784-2_12.
- Hutton, D. M. (2012) ‘Turing’s Cathedral: The Origins of the Digital Universe’. Emerald Group Publishing Limited.
- Hwang, E. (2021) ‘Sketching Dialogue : Incorporating Sketching in Emphatic

- Semi-structured Interviews for HCI'.
- 'Information' (no date). Available at: <https://en.wikipedia.org/wiki/Information>.
- Information Commissioner's Office (2014) 'Data controllers and data processors: what the difference is and what the governance implications are', p. 20. Available at: <https://ico.org.uk/for-organisations/guide-to-data-protection/introduction-to-data-protection/some-basic-concepts/>.
- Information Commissioner's Office (2018) 'Your data matters - Your rights'. Available at: <https://ico.org.uk/your-data-matters/>.
- Information Commissioner's Office (2021a) 'Your right of access'. Available at: <https://ico.org.uk/your-data-matters/your-right-to-get-copies-of-your-data/> (Accessed: 23 August 2021).
- Information Commissioner's Office (2021b) 'Your right to data portability'.
- 'Infovark Company Profile' (2007). Available at: <https://www.crunchbase.com/organization/infovark>.
- Jasperson, J. (Sean). *et al.* (2002) 'Review: Power and Information Technology Research: A Metatriangulation Review'. Society for Information Management; The Management Information Systems Research Center. doi: 10.2307/4132315.
- 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.
- Jilek, C. *et al.* (2018) 'Context spaces as the cornerstone of a near-transparent and self-reorganizing semantic desktop', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11155 LNCS, pp. 89–94. doi: 10.1007/978-3-319-98192-5_17.
- Johnson, S. L., Kim, Y. M. and Church, K. (2010) 'Towards client-centered counseling: Development and testing of the WHO Decision-Making Tool', *Patient Education and Counseling*. Elsevier Ireland Ltd, 81(3), pp. 355–361. doi: 10.1016/j.pec.2010.10.011.
- Jones, T. (2011) 'Designing for second screens : The Autumnwatch Companion'. Available at: <https://www.bbc.co.uk/blogs/researchanddevelopment/2011/04/the-autumnwatch-companion--de.shtml>.
- Jones, W. *et al.* (2006) '"It's about the information stupid!": Why we need a separate field of human-information interaction', *Conference on Human Factors in Computing Systems - Proceedings*, pp. 65–68. doi: 10.1145/1125451.1125469.
- Jones, W. (2011a) 'The Future of Personal Information Management Part I: Our Information, Always and Forever'.
- Jones, W. (2011b) 'The Future of Personal Information Management Part I: Our Information, Always and Forever', p. 72.
- Kalvet, T. (2005) 'Digital divide and the ICT paradigm generally and in estonia', in *Encyclopedia of developing regional communities with information and communication technology*. IGI Global, pp. 182–187. doi: 10.4018/978-1-59140-575-7.ch032.
- Karger, D. R. *et al.* (2005) 'Haystack: A customizable general-purpose information management tool for end users of semistructured data', in *2nd biennial con-*

- ference on innovative data systems research, *CIDR* 2005, pp. 13–27. Available at: <https://s3.amazonaws.com/academia.edu.documents/46870765/haystack.pdf>.
- Karger, D. R. and Jones, W. (2006) ‘Data unification in personal information management’, *Communications of the ACM*, 49(1), p. 77. doi: 10.1145/1107458.1107496.
- Kasirzadeh, A. and Clifford, D. (2021) *Fairness and Data Protection Impact Assessments*. Association for Computing Machinery (1), pp. 146–153. doi: 10.1145/3461702.3462528.
- Kaye, J. et al. (2015) ‘Dynamic consent: a patient interface for twenty-first century research networks’, *European Journal of Human Genetics*. Nature Publishing Group, 23(2), pp. 141–146. doi: 10.1038/ejhg.2014.71.
- Kelly, K. and Wolf, G. (2007) ‘What is the quantified self’. Available at: <https://web.archive.org/web/20100507215130/http://www.kk.org/quantifiedself/2007/10/what-is-the-quantifiable-self.php>.
- Kelly, R. (2020) ‘The Biggest ICO Fines Ever Issued’. Available at: <https://digit.fyi/data-protection-2020-the-biggest-fines-ever-issued-by-the-ico/>.
- Kelty, C. M. (2008) *Geeks and Recursive Publics*. Duke University Press, pp. 27–63.
- Kensing, F. and Blomberg, J. (1998) ‘Participatory design: Issues and concerns’, *Computer supported cooperative work (CSCW)*. Springer, 7(3), pp. 167–185.
- Kirven, A. (2018) ‘Whose gig is it anyway: Technological change, workplace control and supervision, and workers’ rights in the gig economy’, *U. Colo. L. Rev.* HeinOnline, 89, p. 249.
- Klatzky, S. R. (1970) ‘Automation, size, and the locus of decision making: the cascade effect’, *The Journal of Business*. JSTOR, 43(2), pp. 141–151. Available at: <https://www.jstor.org/stable/pdf/2352107.pdf?refreqid=excelsior%3A24bde6bf7de0eccf42c6ea11f8446d38>.
- Klein, B. et al. (2004) ‘Enabling flow - A paradigm for document-centered personal information spaces’, in *Proceedings of the eighth IASTED international conference on artificial intelligence and soft computing*, pp. 187–192. Available at: <https://www.semanticscholar.org/paper/Enabling-flow%3A-%7BA%7D-paradigm-for-document-centered-Klein-Agne/22be4a7b25e75de235e5d96bad6ab4ab4583daac>.
- Kostkova, P. (2015) ‘Grand Challenges in Digital Health’, *Frontiers in Public Health*. Frontiers Media SA, 3. doi: 10.3389/fpubh.2015.00134.
- Kriisk, K. and Minas, R. (2017) ‘Social rights and spatial access to local social services: The role of structural conditions in access to local social services in Estonia’, *Social Work and Society*, 15(1). Available at: <https://www.socwork.net/sws/article/view/503/1007>.
- Krishnan, A. (2010) ‘Pervasive Personal Information Spaces’. University of Waikato. Available at: <https://researchcommons.waikato.ac.nz/handle/10289/4590>.
- Krishnan, A. and Jones, S. (2005) ‘TimeSpace: Activity-based temporal visualisation of personal information spaces’, *Personal and Ubiquitous Computing*, 9(1), pp. 46–65. doi: 10.1007/s00779-004-0291-x.
- Kröger, J. L., Miceli, M. and Müller, F. (2021) ‘How Data Can Be Used Against

- People: A Classification of Personal Data Misuses', *SSRN Electronic Journal*, (December). doi: 10.2139/ssrn.3887097.
- Lansdale, M. W. (1988) 'The psychology of personal information management', *Applied Ergonomics*, 19(March 1988), pp. 55–66. doi: 10.1016/0003-6870(88)90199-8.
- Lansdale, M. and Edmonds, E. (1992) 'Using memory for events in the design of personal filing systems', *International Journal of Man-Machine Studies*, 36(1), pp. 97–126. doi: 10.1016/0020-7373(92)90054-O.
- Larsson, S. (2018) 'Algorithmic governance and the need for consumer empowerment in data-driven markets', *Internet Policy Review*, 7(2). doi: 10.14763/2018.2.791.
- Le Dantec, C. A. (2016) *Designing publics*. MIT Press.
- Lecluijze, I. et al. (2015) 'Co-production of ICT and children at risk: The introduction of the Child Index in Dutch child welfare', *Children and Youth Services Review*. Elsevier Ltd, 56, pp. 161–168. doi: 10.1016/j.childyouth.2015.07.003.
- Lemley, M. A. (2021) 'The splinternet', *Duke Law Journal*, pp. 1397–1428. Available at: <https://perma.cc/92LZ-B8DN%5D>.
- 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.
- Lewin, K. (1951) 'Problems of research in social psychology', *Field theory in social science: Selected theoretical papers*, pp. 155–169.
- Li, I. (2009) 'Designing Personal Informatics Applications and Tools that Facilitate Monitoring of Behaviors', *Uist*. Available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.232.8536>.
- Li, I., Forlizzi, J. and Dey, A. (2010) 'Know thyself: Monitoring and reflecting on facets of one's life', *Conference on Human Factors in Computing Systems - Proceedings*, pp. 4489–4492. doi: 10.1145/1753846.1754181.
- Lindley, S. E. et al. (2018) 'Exploring new metaphors for a networked world through the file biography', *Conference on Human Factors in Computing Systems - Proceedings*, 2018-April, pp. 1–12. doi: 10.1145/3173574.3173692.
- 'List of target companies for GDPR requests' (no date). Available at: <https://wiki.personaldata.io/wiki/Item:Q2369> (Accessed: 22 September 2021).
- Lomas, N. (2020) 'UK's ICO faces legal action after closing adtech complaint with nothing to show for it'. Available at: <https://techcrunch.com/2020/11/05/uk-ico-faces-legal-action-after-closing-adtech-complaint-with-nothing-to-show-for-it/>.
- Lowe, T. and Wilson, R. (2015) 'Playing the game of outcomes-based performance management', *Is Gamesmanship Inevitable*.
- Luger, E. and Rodden, T. (2013) 'An informed view on consent for ubicomp', in *UbiComp 2013 - proceedings of the 2013 ACM international joint conference on pervasive and ubiquitous computing*. New York, New York, USA: ACM Press,

- pp. 529–538. doi: 10.1145/2493432.2493446.
- Mahieu, R. L. P., Asghari, H. and Van Eeten, M. (2018) ‘Collectively exercising the right of access: Individual effort, societal effect’, *Internet Policy Review*, 7(3), pp. 1–23. doi: 10.14763/2018.3.927.
- Mahieu, R. L. P. and Ausloos, J. (2020) ‘Harnessing the collective potential of GDPR access rights : towards an ecology of transparency’, *Internet Policy Review*. Available at: <https://policyreview.info/articles/news/harnessing-collective-potential-gdpr-access-rights-towards-ecology-transparency/1487>.
- Malomo, F. and Sena, V. (2017) ‘Data Intelligence for Local Government? Assessing the Benefits and Barriers to Use of Big Data in the Public Sector’, *Policy and Internet*, 9(1), pp. 7–27. doi: 10.1002/poi3.141.
- Malone, T. W. (1983) ‘How do people organize their desks?: Implications for the design of office information systems’, *ACM Transactions on Information Systems*, 1(1), pp. 99–112. doi: 10.1145/357423.357430.
- Mannay, D. and Morgan, M. (2015) ‘Doing ethnography or applying a qualitative technique? Reflections from the ‘waiting field’’, *Qualitative research*. Sage Publications Sage UK: London, England, 15(2), pp. 166–182.
- Marchionini, G. (2008) ‘Human-information interaction research and development’, *Library and Information Science Research*, 30(3), pp. 165–174. doi: 10.1016/j.lisr.2008.07.001.
- Marshall, C. C. and Jones, W. (2006) ‘Keeping encountered information’, *Communications of the ACM*, 49(1), pp. 66–67. doi: 10.1145/1107458.1107493.
- McCarthy, J. and Wright, P. (2004) ‘Technology as experience’, *Interactions*, 11(5), pp. 42–43. doi: 10.1145/1015530.1015549.
- McCullagh, K. (2009) ‘Protecting ‘privacy’ through control of ‘personal’ data processing: A flawed approach’, *International Review of Law, Computers and Technology*, 23(1-2), pp. 13–24. doi: 10.1080/13600860902742562.
- McNamee, R. (2019) *Zucked: Waking up to the Facebook Catastrophe*, p. 336.
- Melendez, S. and Pasternack, A. (2019) ‘The data brokers quietly buying and selling your personal information’. Available at: <https://www.fastcompany.com/90310803/heres-the-data-brokers-quietly-buying-and-selling-your-personal-information>.
- Merriam-Webster Dictionary (no date a) ‘Usability’. Available at: <https://www.merriam-webster.com/dictionary/usability>.
- Merriam-Webster Dictionary (no date b) ‘Usable’. Available at: <https://www.merriam-webster.com/dictionary/usable>.
- Meschtscherjakov, A., Wilfinger, D. and Tscheiligi, M. (2014) ‘Mobile attachment-Causes and consequences for emotional bonding with mobile phones’, *Conference on Human Factors in Computing Systems - Proceedings*, pp. 2317–2326. doi: 10.1145/2556288.2557295.
- Microsoft (2021) ‘Project Bali’. Available at: <https://www.microsoft.com/en-us/research/project/bali/> (Accessed: 23 August 2021).
- Miettinen, R. (2013) *Innovation, human capabilities, and democracy: Towards an enabling welfare state*. Oxford University Press.
- 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.
- Morgan, J. (2020) ‘Making your Solid Apps interoperable with ShapeRepo.com’. Available at: <https://medium.com/@JacksonMorgan/making-your-solid-apps-interoperable-with-shaperepo-com-8da512936073>.
- Morozov, E. (2013) *To save everything, click here: The folly of technological solutionism*. Public Affairs.
- Mortier, R. *et al.* (2013) ‘Challenges & opportunities in human-data interaction’, *University of Cambridge, Computer Laboratory*. Citeseer. doi: 10.5210/fm.v17i5.4013.
- Mortier, R. *et al.* (2014) ‘Human-data interaction: The human face of the data-driven society’, Available at SSRN 2508051. doi: 10.2139/ssrn.2508051.
- 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.
- Neff, G. (2013) ‘Why Big Data Won’t Cure Us’, *Big Data*, 1(3), pp. 117–123. doi: 10.1089/big.2013.0029.
- Negroponte, N. and Bolt, R. A. (1978) *Spatial data management system*. MASSACHUSETTS INST OF TECH CAMBRIDGE ARCHITECTURE MACHINE GROUP.
- Nelson, T. (2006) ‘Lost in hyperspace’, *New Scientist*, 191(2561). doi: 10.1002/elsc.200620112.
- Nelson, T. H. (1965) ‘Complex information processing’, pp. 84–100. doi: 10.1145/800197.806036.
- Nielsen, J. (2012) ‘Usability 101: Introduction to Usability’. Available at: <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>.
- Norman, D. A. and Draper, S. W. (1986) ‘User Centered System Design; New Perspectives on Human-Computer Interaction’. L. Erlbaum Associates Inc.
- Northern Health Science Alliance (2020) *Connected Health Cities Impact Report*. Newcastle upon Tyne, UK: Northern Health Science Alliance, pp. 129–130.

- O'Donnell, B. (2020) 'Zoom, the office and the future: What will work look like after coronavirus?' Available at: <https://eu.usatoday.com/story/tech/columnist/2020/09/07/zoom-work-from-home-future-office-after-coronavirus/5680284002/>.
- O'Donoghue, T. and Rabin, M. (2001) 'Choice and procrastination', *The Quarterly Journal of Economics*. MIT Press, 116(1), pp. 121–160.
- Odom, W. et al. (2018) 'Time, Temporality, and Slowness', pp. 383–386. doi: 10.1145/3197391.3197392.
- OFSTED (2015) *Early help: whose responsibility?*, p. 32. Available at: www.ofsted.gov.uk https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/410378/Ea
- Open Lab (2014) 'Centre for doctoral training in digital civics: About the CDT'. Newcastle University. Available at: <https://web.archive.org/web/20180727024332/https://digitalcivics.io/apply/> (Accessed: 6 July 2022).
- 'Open rights group: Who we are' (no date). Available at: <https://www.openrightsgroup.org/who-we-are/> (Accessed: 16 June 2022).
- Organisation for Economic Co-operation and Development (1980) *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. Available at: <https://www.oecd.org/digital/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldatal.htm>.
- Pansardi, P. (2012) 'Power to and power over: Two distinct concepts of power?', *Journal of Political Power*, 5(1), pp. 73–89. doi: 10.1080/2158379X.2012.658278.
- Papert, S. (1980) 'Mindstorms: children, computers, and powerful ideas'. Basic Books, Inc.
- Peikoff, L. (1993) *Objectivism: The Philosophy of Ayn Rand*. Penguin Publishing Group (Ayn rand library). Available at: <https://books.google.co.uk/books?id=G6DDlqNftGcC>.
- 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/>.
- Pidoux, J. et al. (2022) *Digipower technical reports: Understanding influence and power in the data economy*. doi: 10.5281/zenodo.6554155.
- 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.
- Pollock, R. (2011) 'Building the (Open) Data Ecosystem – Open Knowledge Foundation Blog'. Available at: <https://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/> (Accessed: 23 July 2019).
- Pór, G. (1997) 'Designing Knowledge Ecosystems for Communities of Practice', in *Advancing organizational capability via knowledge management*.
- Price Ball, M. (no date) 'About Us'. Available at: <https://www.openhumans.org/about/> (Accessed: 31 March 2021).
- 'Privacy' (no date). Available at: <https://privacy.linkedin.com/> (Accessed: 9 August 2021).
- 'Privacy - Apple (UK)' (no date). Available at: <https://www.apple.com/uk/privacy/> (Accessed: 9 August 2021).

- ‘Privacy & Terms – Google’ (no date). Available at: <https://policies.google.com/> (Accessed: 9 August 2021).
- Puussaar, A., Clear, A. K. and Wright, P. (2017) ‘Enhancing personal informatics through social sensemaking’, in *Proceedings of the 2017 CHI conference on human factors in computing systems*, pp. 6936–6942.
- Quinn, P. (2021) ‘Research under the GDPR—a level playing field for public and private sector research?’, *Life Sciences, Society and Policy*. Springer, 17(1), pp. 1–33.
- Raskin, J. (2000) *The humane interface: new directions for designing interactive systems*. Addison-Wesley Professional.
- Reason, P. and Bradbury, H. (2001) *Handbook of action research: Participative inquiry and practice*. Sage.
- ‘Recursive Public (Discussion Page)’ (no date). Available at: https://wiki.p2pfoundation.net/Recursive_Public (Accessed: 16 June 2022).
- Ries, E. (2011) *Wiki: The Lean Startup: How Today’s Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. Crown. Available at: http://en.wikipedia.org/wiki/Lean_Startup.
- Rivera-Pelayo, V. et al. (2012) ‘A framework for applying Quantified Self approaches to support reflective learning’, *Proceedings of the IADIS International Conference Mobile Learning 2012, ML 2012*, pp. 123–131.
- Roche, M. (2011) ‘Full internet ban for sex offenders ruled unlawful’. Available at: <https://ukhumanrightsblog.com/2011/08/12/full-internet-ban-for-sex-offenders-ruled-unlawful/> (Accessed: 23 March 2021).
- Rogers, Y. (2006) ‘Moving on from Weiser’s Vision of Calm Computing: Engaging UbiComp Experiences’, *LNCS*, 4206, pp. 404–421. Available at: [http://www.inf.ufg.br/\\$/sim\\$vagner/courses/mobilecomputing/docs/papers/03-Rogers_Ubicomp06.pdf](http://www.inf.ufg.br/$/sim$vagner/courses/mobilecomputing/docs/papers/03-Rogers_Ubicomp06.pdf).
- Ross, G. (2005) ‘An introduction to Tim Berners-Lee’s Semantic Web’. Available at: <https://www.techrepublic.com/article/an-introduction-to-tim-berners-lees-semantic-web/>.
- Rughinis, R. et al. (2021) ‘From social netizens to data citizens: Variations of GDPR awareness in 28 european countries’, *Computer Law & Security Review*. Elsevier, 42, p. 105585.
- Saha, D. and Mukherjee, A. (2003) ‘Pervasive computing: A paradigm for the 21st century’. IEEE. doi: 10.1109/MC.2003.1185214.
- Sauermann, L., Bernardi, A. and Dengel, A. (2005) ‘Overview and outlook on the semantic desktop’, in *CEUR workshop proceedings*.
- Savage, A. and Hyde, R. (2014) ‘Using freedom of information requests to facilitate research’, *International Journal of Social Research Methodology*. Routledge, 17(3), pp. 303–317. doi: 10.1080/13645579.2012.742280.
- Schneider, H. et al. (2018) ‘Empowerment in HCI - A survey and framework’, in *Conference on human factors in computing systems - proceedings*. Association for Computing Machinery. doi: 10.1145/3173574.3173818.
- Schumacher, K., Sintek, M. and Sauermann, L. (2008) ‘Combining fact and document retrieval with spreading activation for semantic desktop search’, in *Lecture notes in computer science (including subseries lecture notes in artificial*

- intelligence and lecture notes in bioinformatics*), pp. 569–583. doi: 10.1007/978-3-540-68234-9_42.
- Searls, D. (2008) ‘The Intention Economy: What Happens When Customers Get Real Power’. Available at: https://web.archive.org/web/20101226073246/http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/2009_03_24_lunchtalk.ppt.
- Searls, D. (2012) *The intention economy: when customers take charge*. Harvard Business Press (04), pp. 50-2168-50-2168. doi: 10.5860/choice.50-2168.
- Seligman, C. and Darley, J. M. (1976) ‘Feedback as a means of decreasing residential energy consumption’, *Journal of Applied Psychology*, 62(4), pp. 363–368. doi: 10.1037/0021-9010.62.4.363.
- 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. (III). and Marshall, C. C. (1999) ‘Formality Considered Harmful : Experiences , Emerging Themes , and Directions on the Use of Formal Representations in Interactive Systems’, pp. 333–352.
- Shneiderman, B. (1996) *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualisations*.
- 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. (1971) ‘Designing Organizations for an Information-Rich World’, *Computers, Communication, and the Public Interest.*, pp. 37–72.
- Simon, H. A. and Newell, A. (1958) ‘Heuristic Problem Solving: The next advance in operations research’. doi: 10.1057/978-1-349-94848-2_792-1.
- 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.
- Soja, E. (2015) ‘Supporting Healthcare of the Elderly through ICT: Socio-demographic Conditions and Digital Inclusion’, in *Knowledge economy society - challenges and development trends of modern economy, finance and information technology.*, pp. 279–290.
- Spagnuelo, D., Ferreira, A. and Lenzini, G. (2019) ‘Accomplishing Transparency within the General Data Protection Regulation.’, in *ICISSP*, pp. 114–125.
- Spector, P. E. (1982) ‘Behavior in organizations as a function of employee’s locus of control’, *Psychological Bulletin*, 91(3), pp. 482–497. doi: 10.1037/0033-2909.91.3.482.
- Spencer, D. and Warfel, T. (2004) ‘Card sorting: A definitive guide’, *Boxes and arrows*, 2(2004), pp. 1–23.

- Spiekermann, S. and Korunovska, J. (2017) ‘Towards a value theory for personal data’, *Journal of Information Technology*, 32(1), pp. 62–84. doi: 10.1057/jit.2016.4.
- Spinuzzi, C. (2005) ‘The methodology of participatory design’, *Technical Communication*. Society for Technical Communication, 52(2), pp. 163–174.
- Star, S. L. (1989) ‘The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving’, in *Distributed artificial intelligence*. Elsevier, pp. 37–54. doi: 10.1016/b978-1-55860-092-8.50006-x.
- Star, S. L. (2010) ‘This is not a boundary object: Reflections on the origin of a concept’, *Science Technology and Human Values*, 35(5), pp. 601–617. doi: 10.1177/0162243910377624.
- Steinberg, S. G. (1997) ‘Lifestreams’, *Wired*. Available at: <https://www.wired.com/1997/02/lifestreams/>.
- Steinfeld, N. (2016) “I agree to the terms and conditions”:(how) do users read privacy policies online? An eye-tracking experiment’, *Computers in human behavior*. Elsevier, 55, pp. 992–1000.
- Steyaert, J. and Gould, N. (2009) ‘Social work and the changing face of the digital divide’, *British Journal of Social Work*, 39(4), pp. 740–753. doi: 10.1093/bjsw/bcp022.
- Storni, C. (2014) ‘The problem of De-sign as conjuring: Empowerment-in-use and the politics of seams’, *Proceedings of the 13th Participatory Design Conference on Research Papers - PDC '14*. New York, New York, USA: ACM Press, pp. 161–170. doi: 10.1145/2661435.2661436.
- Strohmayer, A. et al. (2021) *Trust and Abusability Toolkit: Centering Safety in Human-Data Interactions*.
- 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><https://decodeproject.eu/publications/me-my-data-and-ithe-future-personal-data-economy%0Ahttps://media.nesta.org.uk/documents/decode-02.pdf%0Ahttps://decodeproject.eu/publications/me-my-data-and-ithe-fu>.
- Taplin, D. H. and Clark, H. (2012) ‘Theory of change basics: A primer on theory of change’, *ActKnowledge*, p. 9. Available at: http://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/ToC Basics.pdf.
- Taylor, A. S. et al. (2015) ‘Data-in-place: Thinking through the relations between data and community’, in *Conference on human factors in computing systems - proceedings*, pp. 2863–2872. doi: 10.1145/2702123.2702558.
- Taylor, L. (2017) ‘What is data justice? The case for connecting digital rights and freedoms globally’, *Big Data and Society*, 4(2). doi: 10.1177/2053951717736335.
- 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.

- 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 (2016a) ‘Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data’, pp. 16–32. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679> <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=ES>.
- The European Parliament and the Council of the European Union (2016b) ‘Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data’. Available at: <https://eur-lex.europa.eu/eli/reg/2016/679/oj> <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=ES>.
- ‘The GDPR: Does it Benefit Consumers in Any Practical Way?’ (2020). Atebits.com. Available at: <https://www.atebits.com/the-gdpr-does-it-benefit-consumers-in-any-practical-way/>.
- ‘The personal computer revolution’ (no date) in *Britannica*. Available at: <https://www.britannica.com/technology/computer/The-personal-computer-revolution>.
- Thompson, R. (2011) ‘Radicalization and the use of social media’, *Journal of strategic security*. JSTOR, 4, pp. 167–190.
- 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/>.
- Tregeagle, S. and Darcy, M. (2008) ‘Child welfare and information and communication technology: Today’s challenge’, *British Journal of Social Work*, 38(8), pp. 1481–1498. doi: 10.1093/bjsw/bcm048.
- Tufekci, Z. (2017) ‘We’re building a dystopia just to make people click on ads’, *TED*. TED. Available at: https://www.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads.
- Tunikova, O. (2018) ‘Are We Consuming Too Much Information?’ Available at: https://medium.com/@tunikova_k/are-we-consuming-too-much-information-b68f62500089 (Accessed: 23 March 2021).
- Udaptor (2021) ‘Udaptor Assistant’. Available at: <https://udaptor.io/assistant.html> (Accessed: 23 August 2021).
- US Department of Health Education and Welfare (1973) ‘Records Computers and the Rights of Citizens’.
- Varghes, S. (2019) ‘Pokémon Go was a warning about the rise of surveillance capitalism’. Available at: <https://www.wired.co.uk/article/the-age-of-surveillance-capitalism-facebook-shoshana-zuboff>.
- Various Authors (2018) ‘Our Digital Lives’, in *TED talks*. TED. Available at: https://www.ted.com/playlists/26/our_digital_lives.
- Véliz, C. (2020) *Privacy Is Power: Why and How You Should Take Back*

- Control of Your Data.* Transworld Publishers Limited, p. 224. Available at: <https://b-ok.lat/book/11000161/b53144>.
- Vlachokyriakos, V. et al. (2016) ‘Digital civics: Citizen empowerment with and through technology’, *Conference on Human Factors in Computing Systems - Proceedings*, 07-12-May-, pp. 1096–1099. doi: 10.1145/2851581.2886436.
- Wagner, A. (2012) ‘Is internet access a human right?’ Available at: <https://www.theguardian.com/law/2012/jan/11/is-internet-access-a-human-right> (Accessed: 23 March 2021).
- Waldman, A. E. (2020) ‘Data Protection by Design ? A Critique of Article 25 of the GDPR’, 1239(2019), pp. 147–168.
- Wallace, D. P. (2007) *Knowledge management: Historical and cross-disciplinary themes*. Libraries unlimited.
- Weiser, M. (1991) ‘The computer for the 21st century’, *Scientific American*, 265(3), pp. 94–105. doi: 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/F86D0A71E-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.
- ‘Whose data is it anyway?’ (2019). 04: UBDI. Available at: <https://www.ubdi.com/blog/whose-data-is-it-anyway> (Accessed: 31 March 2021).
- Wiki.personaldata.io (no date) ‘Subject Access Request Template’. Available at: <https://wiki.personaldata.io/wiki/Template:Access> (Accessed: 21 September 2021).
- Williams, H. et al. (2015) ‘Dynamic consent: a possible solution to improve patient confidence and trust in how electronic patient records are used in medical research.’, *JMIR medical informatics*. JMIR Publications Inc., 3(1), p. e3. doi: 10.2196/medinform.3525.
- Wilson, L., Wilson, R. and Martin, M. (2020) *Health and Care Practitioner Insights: Understanding Information Sharing in Constellations of Care - Report on Amy’s Page workshop series*. Great North Care Record. Available at: <https://www.greatnorthcarerecord.org.uk>.
- Wilson, R. et al. (2011) ‘Re-Mixing Digital Economies in the Voluntary Community Sector? Governing Identity Information and Information Sharing in the Mixed Economy of Care for Children and Young People*’, *Social Policy and Society*. Cambridge University Press, 10(3), pp. 379–391. doi: 10.1017/s1474746411000108.
- ‘WinFS’ (no date). Available at: <https://en.wikipedia.org/wiki/WinFS>.
- Wong, J. and Henderson, T. (2018) ‘How Portable is Portable ? Exercising the GDPR ’s Right to Data Portability’, *Acm*, pp. 911–920.
- Woods, D. W. and Böhme, R. (2022) ‘The commodification of consent’, *Comput-*

- ers & Security*. Elsevier Ltd, 115, p. 102605. doi: 10.1016/j.cose.2022.102605.
- Woolgar, S. (2014) 'Configuring the User: The Case of Usability Trials', *The Sociological Review*, 38(1_suppl), pp. 58–99. doi: 10.1111/j.1467-954x.1990.tb03349.x.
- Wright, P. and McCarthy, J. (2008) 'Empathy and experience in HCI', *Conference on Human Factors in Computing Systems - Proceedings*, pp. 637–646. doi: 10.1145/1357054.1357156.
- Xie, A., Ho, J. C. F. and Wang, S. J. (2021) 'Data City: Leveraging Data Embodiment Towards Building the Sense of Data Ownership', pp. 365–378. doi: 10.1007/978-3-030-73426-8_22.
- Zichichi, M., Ferretti, S. and D'Angelo, G. (2020) 'On the Efficiency of Decentralized File Storage for Personal Information Management Systems'. Available at: <https://arxiv.org/abs/2007.03505>.
- Zins, C. (2015) 'What is the meaning of "data", "information", and "knowledge"?' , *Institute of Knowledge Sharing*, 3(1).
- Ziogas, G. (2020) 'The Inventor of the World Wide Web Says the Internet Is Broken'. Available at: <https://medium.com/digital-diplomacy/the-inventor-of-the-world-wide-web-says-the-internet-is-broken-fbce1c8bf6cf>.
- Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile. Available at: <https://books.google.co.uk/books?id=W7ZEDgAAQBAJ>.
- Zuckerman, E. (2021) *Mistrust: Why Losing Faith In Institutions Provides The Tools To Transform Them*. New York, NY, USA: W. W. Norton & Company, pp. 1–3. doi: 10.1017/ipo.2021.30.
- Zúñiga, H. de, Garcia-Perdomo, V. and McGregor, S. C. (2015) 'What is second screening? Exploring motivations of second screen use and its effect on online political participation', *Journal of communication*. Oxford University Press, 65(5), pp. 793–815.