Understanding and Improving Human Data Relations

Alex Bowyer

Contents

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

[TODO address JG feedback]

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, 2014).

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 occurreed, 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, hew 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

[TODO update per JG feedback - split into short motivation here, and longer reflection later (7.5/8?)]

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 2.0[1], 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 conferences[2] 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

---

[1]Archived at https://web.archive.org/web/20111231165329/http://www.human20.com/
[2]http://bitnorth.com/shortbits/

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

1.1.2 Research Objectives and Purpose

[TODO address JG feedback]

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 is 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 analyis of interview data will form the core academic research of this PhD.

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 investigative 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, design 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 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** [3.2.2] of the PhD through more than just the participatory investigative 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

[TODO address JG feedback]

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**15 [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 7.1.2, 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.

[TODO Only highlight these lists here, move the full lists to 7.5/8?]

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

[TODO summarise this list] 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 intractable 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:

[TODO summarise this list?] 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.

[TODO highlight the conception of HDR literacy, possibly as distinct contribution?]

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 GPDR 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 as 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 (Härkönen and Vänskä, 2021), 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; Härkönen *et al.*, 2022).

1.3 Publications arising from and connected to this research

[TODO address JG feedback]

1.3.1 Pilot Study

My Doctoral Training programme at Open Lab began with a Masters in Research in Digital Civics. For my MRes project[3], 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 and presented at CHI 2018:

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

This study is given a special status in this thesis; while 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 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 1.1. A full journal paper of Case Study One is in prep.

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 and presented as a full first-author paper at CHI 2022, where it was awarded an *Honorable Mention*:

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

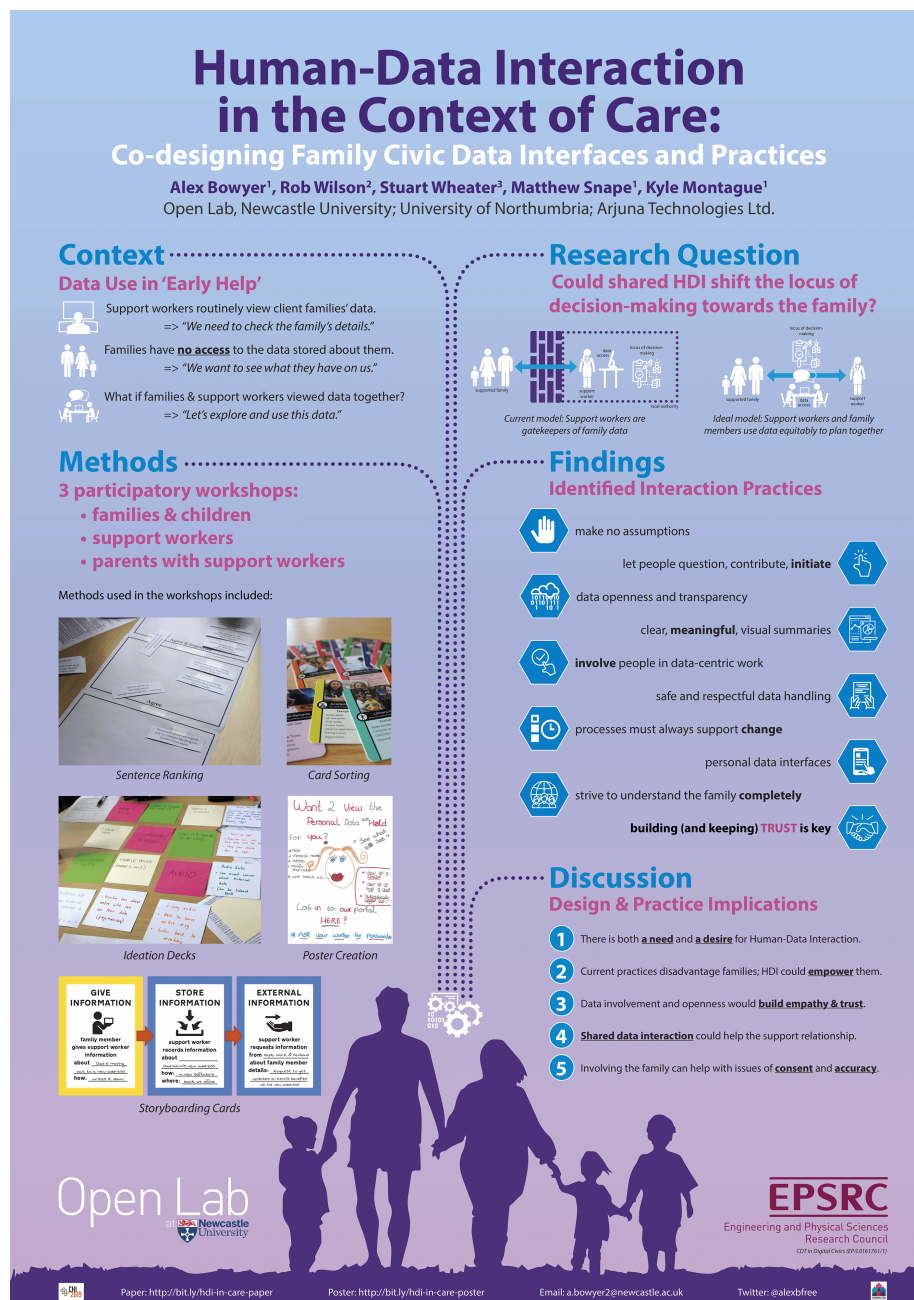[3]MRes result awarded: Distinction.

Figure 1: Figure 1.1: Poster Presentation of Case Study One

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

1.3.3 Workshop papers & presentations

During the PhD, I gave a number of additional presentations and published three workshop papers which included material from, or directly contributed to, this thesis and its argments.

- "Designing For Human Autonomy: The next challenge that civic HCI must address" (Bowyer, 2017) - a short talk I presented to my peers at Open Lab in January 2017 laying out the landscape of reduced agency and possible avenues for improving humans' relationships to their data that my PhD would explore;
- "Free Data Interfaces: Taking Human-Data Interaction to the Next Level" (Bowyer, 2018b) - a CHI 2018 workshop paper formalising my pre-PhD design thinking and outlining a vision for unconstrained and useful data interaction interfaces;
- "A Grand Vision for Post-Capitalist HCI: Digital Life Assistants" (Bowyer, 2018a) - a CHI 2018 workshop paper where I imagined a form of digital computer assistant that is far more helpful and human-data-centric than the digital voice assistants of today;
- "Personal Data Use: A Human-centric Perspective" (Bowyer, 2020) - 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" (Bowyer, 2021b) - in April 2021, I presented my thesis in Newcastle University's 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" (Bowyer, 2021a) - A workshop paper I presented at CHI 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 [3.4.1.1], I contributed to work published through Newcastle University's internal report to CHC (not publicly available) and the overall impact report (Connected Health Cities, no date, pp. 129–130), and more directly published demonstration videos (Bowyer and Wheater, no date) 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

[TODO address JG feedback]

The overall structure of this thesis is illustrated in Figure 1.2. 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 academic investigation of the two research questions. [Chapter 7] looks beyond the theoretical to the practical, setting the stage for future research and innovation in the HDR landscape, 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. Each chapter finishes with a summation section. Chapter 8 concludes the thesis.
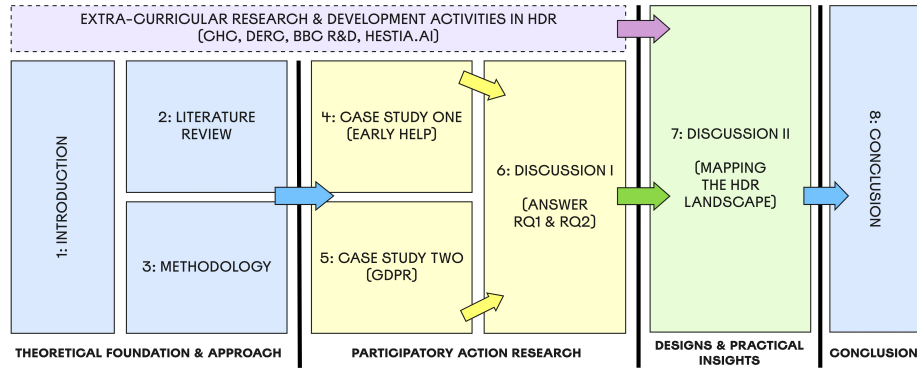

Figure 2: Figure 1.2: The Structure of This Thesis

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 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 (Collins Dictionary, no date) of human-centred innovation, leading to the primary academic Research Question (RQ) of this thesis: *"What relationship do people need with their personal data?"* [2.4].

Chapter 3 describes the methodology used in this research, explaining first the constructist 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.

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.

Chapter 6 synthesises the two case studies, and answers RQ and RQ2, bringing the thesis's central academic 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 looks beyond the original research question, and is a deliberately broad, shallow and open-ended chapter synthesising designerly insights acquired through practical experience. Here I move beyond a traditional thesis structure, with the goal of providing real-world ideas on *how such human-centric empowerment might be achieved in practice*. In tackling this more practical question, the thesis becomes a *valuable and actionable anthology of reference material* for future researchers, activists and innovators. First [7.1], the peripheral contexts that I worked in alongside this PhD from with the additional insights arise are introduced. In [7.2], the findings of the thesis are expanded upon to frame the pursuit of this thesis's *'data wants'* as a defined field of future research, called *Human Data Relations* (HDR), whose practitioners act as a *recursive public*, pursuing four *'landscape objectives'* that could address the data wants in reality.

The landscape of HDR is mapped out in two parts. First [7.3], I outline the identified *obstacles* to pursuit of the HDR objectives. Then [7.4], using a *Theories of Change* (ToC) framing, my identified *opportunities for progress* are introduced, divided into *four different trajectories of change* that could be executed to pursue better HDR. 7.3 and 7.4 are interspersed with actionable *insights* that could help tackle the obstacles and pursue the change trajectories. 7.5 is a summation of the chapter.

Chapter 8 is a brief conclusion of the thesis, summarising its contributions and positioning HDR and this thesis as call to arms for activist research and innovation that can tackle the power imbalance in society. [TODO check this once Chapter 8 is split out separately]

Bibliography

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

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.

Bødker, S. (2015) 'Third-wave HCI, 10 years later—participation and sharing', *Interactions*, 22(5), pp. 24–31. doi: 10.1145/2804405.

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. (2017) 'Designing for human autonomy: The next challenge that civic HCI must address'. Available at: https://eprints.ncl.ac.uk/273832.

Bowyer, A. (2018a) 'A grand vision for post-capitalist HCI: Digital life assistants"', *CHI Workshops 2018.* Available at: https://eprints.ncl.ac.uk/273826.

Bowyer, A. (2018b) 'Free Data Interfaces: Taking Human- Data Interaction to the Next Level', *CHI Workshops 2018.* Available at: https://eprints.ncl.ac.uk/273825.

Bowyer, A. *et al.* (2018) 'Understanding the Family Perspective on the Storage, Sharing and Handling of Family Civic Data', in *Conference on human factors in computing systems - proceedings.* New York, New York, USA: ACM Press, pp. 1–13. doi: 10.1145/3173574.3173710.

Bowyer, A. *et al.* (2019) 'Human-data interaction in the context of care: Co-designing family civic data interfaces and practices', in *Conference on human factors in computing systems - proceedings.* doi: 10.1145/3290607.3312998.

Bowyer, A. (2020) 'Personal data use: A human-centric perspective'. Available at: https://eprints.ncl.ac.uk/273834.

Bowyer, A. (2021a) 'Human-Data Interaction has two purposes: Personal Data Control and Life Information Exploration'. Available at: https://eprints.ncl.ac.uk/273832#.

Bowyer, A. (2021b) 'My thesis in 3 minutes: Understanding and designing human data relations'. Available at: https://www.youtube.com/watch?v=YFHXc_TfM5c.

Bowyer, A., Pidoux, J., *et al.* (2022) *Digipower technical reports: Auditing the data economy through personal data access.* doi: 10.5281/zenodo.6554177.

Bowyer, A., Holt, J., *et al.* (2022) 'Human-GDPR interaction : Practical experiences of accessing personal data', *CHI '22.*

Bowyer, A. and Wheater, S. (no date) 'SILVER family health data interface: Demo screencasts / showcase'. Available at: https://eprints.ncl.ac.uk/273839.

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.

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.

Collins Dictionary (no date) 'bleeding edge'. Available at: https://www.collinsdictionary.com/dictionary/english/bleeding-edge (Accessed: 29 July 2022).

Connected Health Cities (no date) *Connected health cities impact report: 2016 - 2020.* Available at: https://web.archive.org/web/20211225192408/https://www.chc-impact-report.co.uk/ (Accessed: 25 December 2021).

Fu, S. *et al.* (2020) 'Social media overload, exhaustion, and use discontinu-

ance: 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.

Goffe, L. *et al.* (2021) 'Appetite for disruption: Designing human-centred augmentations to an online food ordering platform', in *34th british HCI conference*, pp. 155–167.

Goffe, L. *et al.* (2022) 'Web augmentation for well-being: The human-centred design of a takeaway food ordering digital platform', *Interacting with Computers*.

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.

Härkönen, T. *et al.* (2022) *Tracking digipower: How data can be used for influencing decision-makers and steering the world*. Sitra. Available at: https://www.sitra.fi/en/publications/tracking-digipower/.

Härkönen, T. and Vänskä, R. (2021). Sitra. Available at: https://www.sitra.fi/en/projects/digipower-investigation/#what-is-it-about.

Hoffman, W. (2014) *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.

Lemley, M. A. (2021) 'The splinternet', *Duke Law Journal*, pp. 1397–1428. Available at: https://perma.cc/92LZ-B8DN%5D.

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.

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.

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).

Pidoux, J. *et al.* (2022) *Digipower technical reports: Understanding influence and power in the data economy*. doi: 10.5281/zenodo.6554155.

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.

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/.

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.

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.

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.