Understanding and Improving Human Data Relations

Alex Bowyer

Contents

1 Introduction **1**
1.1 Background and Motivation to this Research 1
1.1.1 Personal motivations and context 2
1.1.2 Statement of Research purpose 3
1.2 Nature and Contributions of the thesis 3
1.2.1 An Understanding of what People want from Personal Data 3
1.2.2 Establishing a new field: Human Data Relations 4
1.2.3 Additional contributions in the Early Help and Civic Data Use context 5
1.2.4 Additional contributions in the context of GDPR and Everyday Data Access 6
1.3 Publications arising from and connected to this research 6
1.3.1 Pilot Study 6
1.3.2 Primary Case studies 6
1.3.2.1 Publications from Case Study One 6
1.3.2.2 Publication from Case Study Two 7
1.3.3 Workshop papers & presentations 7
1.3.4 Publications from other work 7
1.4 The structure of this thesis 9

Bibliography **9**

1 Introduction

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

[targets give 3900w]

1.1 Background and Motivation to this Research

[why this study, what it builds on, why does the world need it, why does it

matter][750w?]

1.1.1 Personal motivations and context

This PhD and this thesis represent the culmination of a passion to get more value from our computers that has been with me for over 25 years. 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, that you could master it and make it do what you wanted it to. 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 and transform the way people interact. Keenly tracking the Web 2.0 revolution and the digitisation and disruption of so many industries since the start of the 21st century and embracing new capabilities, I became fascinated with the ways in which humans were shaping computer systems which in turn were shaping us. 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. 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 my developing ideas for better human data interaction four times at Bitnorth conferences[2]. During this time I 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. 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 IT and seek ways to research, design and help to build a better digital future, to make computers useful again. This led me, via a web science architect position at citizen science platform Zooniverse that gave further understanding of user motivations, to join the Digital Civics CDT programme (Open Lab, 2014), where I was finally able to work on this most important of problems. It has been a tremendous privilege to spend six years understanding in great detail the nature of the problems facing our data-centric society, and to be able to map out the landscape for improving

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

the way we relate to data. As well as the ability to find grounded evidence to quantify and qualify the losses of agency I had observed and theorised, this gave me space to experiment with using using GDPR to access data and push boundaries, and to design and prototype new views of data. Looking forward, this opportunity has opened doors that have enabled me to begin to put these learnings into action, working on important projects 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 Statement of Research purpose

[purpose of this thesis, and its limitations] [include the 'journey' of C4 and of C5] [and the parallel journey in practical projects] [be clear about difference between the primary part of thesis and what is not] [signposting][450w?]

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. Insufficietn 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

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

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

C8: An understanding of the lived experience of accessing data using GDPR rights

C9: Evidence for the impact of knowledge about data handling practices on provider trust and perceived individual power

C10: Guidance for policymakers, data holders and individuals on how to improve HDR

C11: A methodology for educating individuals about held data, data access and the data ecosystem (with caveat not intended or evaluated as such)

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 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 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 X.

[3]MRes result awarded: Distinction.

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 *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 a lecture 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 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:

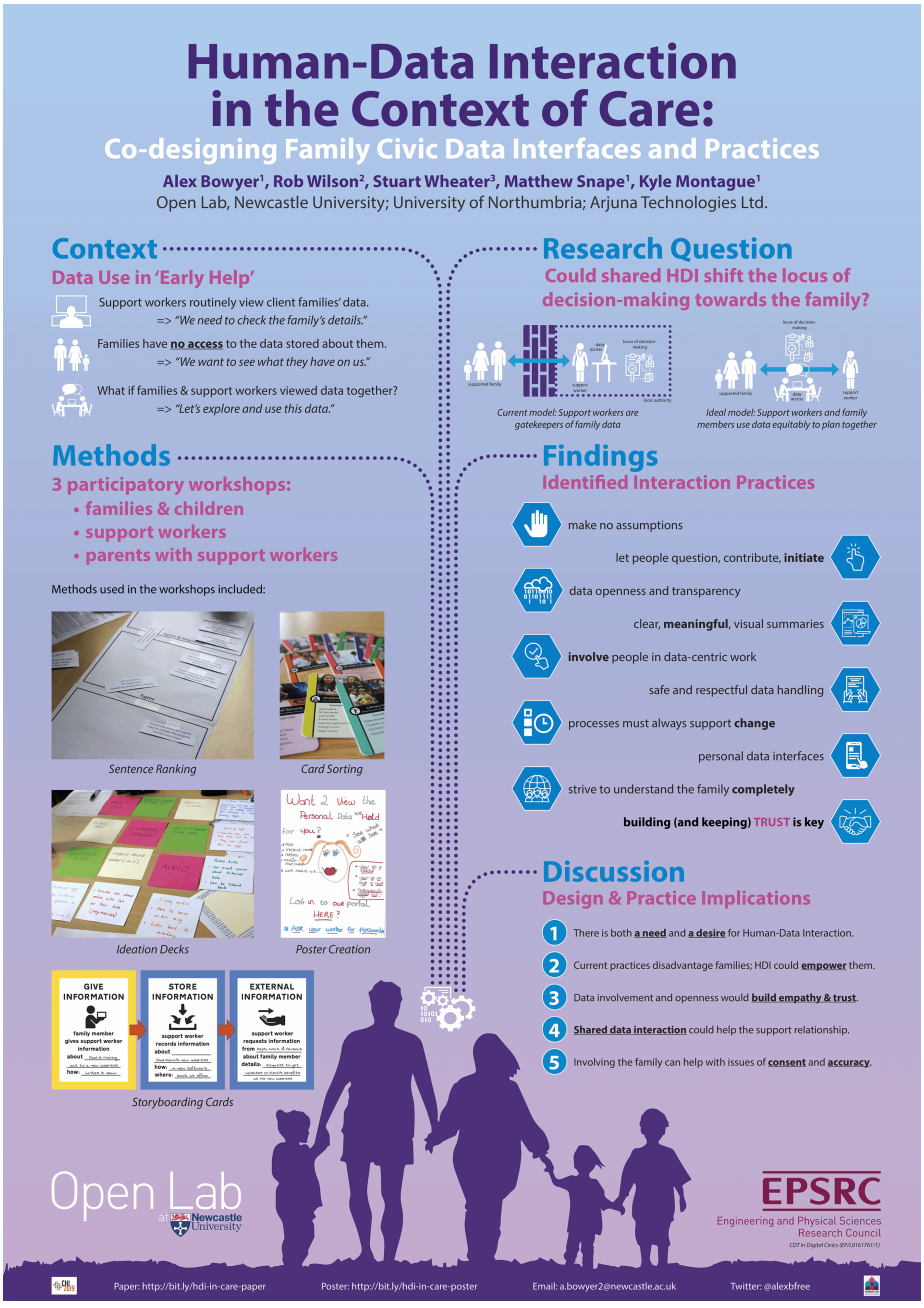

Figure 1: Figure X: Poster Presentation of Case Study One

- 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 a project leader, data access coach and researcher at Hestia.ai, I was a lead author on a research report auditing the data economy, and co-author on a research report on power mechanisms in the data economy.

1.4 The structure of this thesis

[overall description here][120w]

Chapter 2. . . [120w]

Chapter 3. . . [120w]

Chapter 4. . . [120w]

Chapter 5. . . [120w]

Chapter 6. . . [120w]

Chapter 7. . . [120w]

Bibliography

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