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

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). Left unchecked, the situation will not improve for individuals, as 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.

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.

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 a lifelong passion to help people get more value from our computers that has began over 25 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 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 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. The revolutionary potential of a Web 2.0 '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 IT and 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

---

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

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 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

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 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 publi-

cations 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

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.

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.

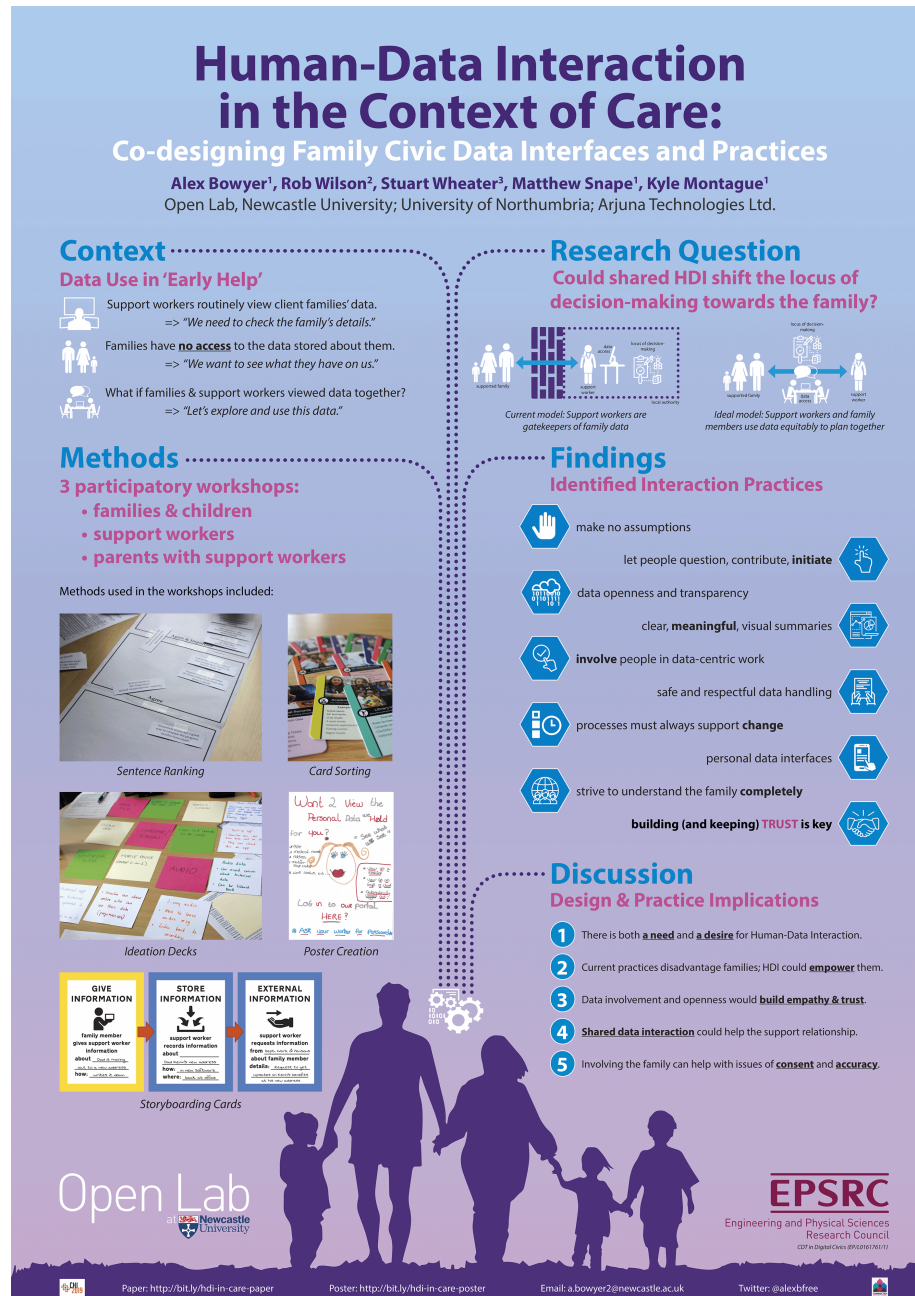---

[3]MRes result awarded: Distinction.

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

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:

- 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

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.
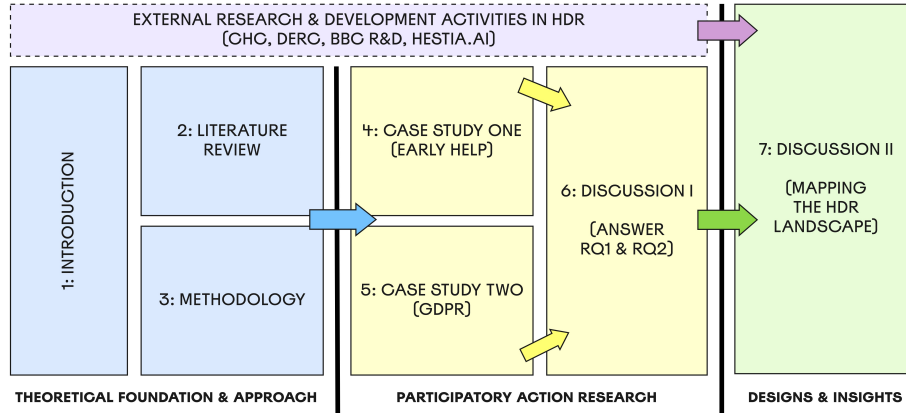

Figure 2: Figure X: The Structure of this Thesis

Chapter 2... [120w]

Chapter 3... [120w]

Chapter 4... [120w]

Chapter 5... [120w]

Chapter 6... [120w]

Chapter 7... [120w]

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.

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. *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.*

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.

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.

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.

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).

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.