Understanding and Improving Human Data Relations

Alex Bowyer

Contents

Abstract **1**

Dedication **2**

Acknowledgements **2**

Lists of Tables, Figures and Inset Boxes **5**

Tables in Chapter 3 5
Tables in Chapter 4 5
Tables in Chapter 5 5
Tables in 'Additional Reference Information' 5
Figures in Chapter 1 5
Figures in Chapter 2 6
Figures in Chapter 3 6
Figures in Chapter 4 6
Figures in Chapter 5 6
Figures in Chapter 7 7
Figures in Chapter 8 7
Figures in Chapter 9 7
Figures in 'Additional Reference Information' 8
Inset Boxes in Chapter 7 8
Inset Boxes in Chapter 8 8
Inset Boxes in Chapter 9 8

Bibliography **9**

Abstract

PCs, smartphones, and cloud computing have transformed the world: In our daily lives, we interact with many businesses and public services who (often to reduce costs) rely on data collection and processing rather than face-to-face user

interactions to inform their decisions. This creates an *imbalance of power* between those who hold personal data and the individuals about whom data is stored, who cannot easily see their data or examine how it is used. This *Digital Civics* PhD research explores, from an individualist perspective, the lived experience of this imbalanced and data-centric world. Through two qualitative case studies across public and private sectors, it discovers desires for visible, understandable and useable data, and for transparent relations with data holders that enable oversight and involvement. Case Study One focuses on *Early Help* social care: Through four workshops with supported families and social workers, perspectives on civic data use are understood. *Shared data interaction* is explored as a means to empower individuals while maintaining an effective care relationship. Case Study Two is a three-month study exploring 10 participants' experience of using *GDPR data access rights* to view their own data, resulting in insights into individual needs and the challenges of data-centric service relationships, and recommendations for improvement of policies and practices. Expanding upon prior work in *Personal Information Management*, *Human Data Interaction* and *MyData* personal data ecosystems, and informed by parallel industrial experience, a new research agenda for improving ***Human Data Relations (HDR)*** is established. The thesis then explores the *practical pursuit* of this agenda from an technologist's adversarial design stance. Drawing on first-hand knowledge acquired from the author's expert participation in research projects at BBC R&D and Hestia.ai/Sitra, *the landscape for future research and innovation* is mapped out in terms of obstacles, designerly insights and activist strategies.

Dedication

For my children Rosie, Joey, and Zach; my nephew Elliott; and my nieces Amy and Lyla. My wish is that that you and your generation might soon experience a future where technology can truly help people and empower them to thrive, and where personal data drives human flourishing more than corporate profit. I hope that this research can contribute to a better future for you all.

Acknowledgements

No-one really knows how hard a PhD is until they are already well beyond the point of no return. It is demanding, challenging and often thankless and lonely work. You spend countless hours striving to find meaning among masses of data or iterating endlessly to concisely express complex, nebulous and elusive ideas. This endeavour is made even harder when you are a mature student with financial and parental responsibilities. At times, especially during the final unfunded writing-up period, the impacts upon my life and those around me have been huge and unreasonable. For this reason, the greatest thanks of all go to my wife Joni Bowyer, who has stood by me throughout, picking up the slack again and again where I could not. She has endured the uncertainties and impacts

upon income, time, and divided attention that this unforgiving work has thrust upon our family. She has also provided practical help on countless occasions with everything from poster layout to time management to grammatical advice. I love you forever, Joni. Thank you.

The next person I want to thank is Jack Holt, who dedicated many weeks and months of his life to collaborate with me to analyse of mountains of participant data from Case Study Two and to co-write the paper (Bowyer *et al.*, 2022) with me. I have absolutely no doubt that without his dedication, it would have been impossible to complete and publish the GDPR study in any reasonable timeframe.

I would like to thank, in reverse chronological order, my supervisors and all the other faculty and staff who have supported me on this six-year journey:

- Dave Kirk, for both detailed and high-level advice in bringing my thesis to conclusion during the final year;
- Jan Smeddinck, for his calm, pragmatic encouragement and thorough feedback and on drafts and plans through the latter half of my research;
- Rob Wilson, for always reminding me to stay grounded in the data, and for his sage advice in matters sociotechnical and philosophical;
- Josephine Go Jefferies, for much-needed scrutiny and challenges to my writing and many detailed chapter draft reviews;
- Rachel Pattinson, for being the best CDT manager a postgraduate could hope for;
- Patrick Olivier, Pete Wright and Dave Kirk, for their continuing commitment to help me find ways to make the PhD financially viable through the finding and accommodating of peripheral paid work;
- Kyle Montague, for his valuable input on study design in the early stages of my PhD, and for advocating to protect my independence and integrity as a researcher during problematic negotiations with a partner organisation;
- Phil Lord, for assistance and advocacy with those same issues at a crucial time;
- Madeline Balaam, for inspiration on participatory methods right at the start, and for helping me develop a paper writing style;
- Paul, Nicola, Sara, James, Glau, Fion, Alex, and all the other School admin staff, for countless random assists;
- Rob Comber, Simon Bowen, Matt Wood and other lecturers who taught me valuable *Digital Civics* and research skills during the MRes, that helped shape me into the researcher I am.

I have been fortunate to take this journey with dozens of other researchers. I would like to especially thank:

- Tom Maskell, for his help with wrestling with concepts around data access and involvement, and companionship on many bus commutes in the early years;
- Sunil Rodger, for valuable moral support, writing camaraderie, and practi-

cal advice for the last two years; and
- Stuart Wheater, for myriad data discussions and tactical discussions that helped me get through a difficult period and often went beyond project business.

I would also like to thank, in no particular order:

- Rebecca Nicholson, Sean Peacock, Jen Manuel, Rosie Bellini, Megan Venn-Wycherley, Kieran Cutting, James Hodge, Hazel Dixon, Seb Prost, Sara Armouch and all the other *Digital Civics* PhDs whose company and mutual support I enjoyed during dozens of writing sessions and discussions;
- (Matt) Marshall, for sharing his powerful script framework for markdown-based thesis writing and document generation (Marshall, 2020), and supporting me in adapting it, making thesis development so much easier;
- Tom Nappey, Laura Pinzon Cardona and Rosie Daglish for their assistance with graphic design;
- Louis Goffe, Debbie Smart, Kat Jackson, Liam Spencer, Ruth McGovern and Kyle Montague for giving their time to help run workshops with participants in the pilot study and in Case Study One;
- Paul-Olivier Dehaye, Mike Martin, Soheil Human, Jasmine Cox, Peter Wells, Ian Forrester, Rhianne Jones, Tim Broom, Suzanne Clarke, Hannes Ricklefs, Max Leonard, Chris Gameson, Euijin Hwang, Toby Lowe, James Nicholson, Alan Mitchell, Iain Henderson, StJohn Deakins, Dalia Al-Shahrabi, Anna Scott, Sarah Knowles, and Michael Jelly, all of whom I have had the pleasure to work with, ruminate with or learn from;
- Chris Lintott and everyone in the Oxford University/Zooniverse team for giving me the smoothest possible start in postgraduate academia before I joined Open Lab;
- Kellie Morrissey, Aare Puussaar, Andy Dow, Anja McCarthy, Zander Wilson, Raghda Zahran, Michael Jelly, Jay Rainey, David Williams, Ben Wright, Paul Whittles, Jon Bowyer, and all the other lovely people who have encouraged me and supported me along the way on this journey; and
- the research participants themselves, who shall remain nameless but without which this work could not exist.

Finally, I would like to thank my parents, Jim and Rosi Bowyer. You have always been there for me, whenever I needed you. Thank you for your unquestioning love and support in an ever-changing world.

This PhD was funded by the Engineering and Physical Sciences Research Council (EPSRC)'s Centre for Doctoral Training in Digital Civics at Newcastle University in the UK (EP/L016176/1).

Lists of Tables, Figures and Inset Boxes

Lists of Tables by Chapter

Tables in Chapter 3

- 3.1 - Context One (Civic Data & Early Help): Participants involved in Research Activities leading into Case Study One.
- 3.2 - Context Two (Digital Life): Participants Involved in Digital Life Research Activities Leading into Case Study Two.

Tables in Chapter 4

- 4.1 - Case Study One Group Design Workshops
- 4.2 - Theme 1 - Meaningful Data Interaction for Families: Subthemes & Participant Quotes
- 4.3 - Theme 2 - Giving a Voice to the Family: Subthemes & Participant Quotes
- 4.4 - Theme 3 - Earning Families' Trust Through Transparency: Subthemes & Participant Quotes

Tables in Chapter 5

- 5.1 - Types of Data Holding Organisation Targeted for GDPR Requests by Study Participants
- 5.2 - Types of Personal Data Potentially Accessible from Data Holders via GDPR Rights
- 5.3 - Presence and Quality Assessments of GDPR Responses by Data Type (as Percentages)
- 5.4 - Participants' Hopes, Imagined Data Uses and Goals for GDPR, as well as Resultant Outcomes
- 5.5 - Theme 1 - Insufficient Transparency: Subthemes & Participant Quotes
- 5.6 - Theme 2 - Confusing & Unuseable Data: Subthemes & Participant Quotes
- 5.7 - Theme 3 - Fragile Relationships: Subthemes & Participant Quotes

Tables in 'Additional Reference Information'

- ARI4.1 - Example Categories of Family Civic Data
- ARI5.1 - Best and Worst Data Holders for GDPR, according to Participants' Judgements
- ARI7.1 - Eight Lenses on Personal Data

Lists of Figures by Chapter

Figures in Chapter 1

- 1.1 - Poster Presentation of Case Study One at CHI 2019

- 1.2 - The Structure of This Thesis

Figures in Chapter 2

- 2.1 - The Wisdom Curve: Making Data into Meaningful Information
- 2.2 - Li *et al.*'s Stage-based Model of Personal Informatics Systems

Figures in Chapter 3

- 3.1 - My Action Research Approach
- 3.2 - Family Facts — What is Data?
- 3.3 - Walls of Data — Sensitising Participants to the World of Commercially-held Data and GDPR
- 3.4 - Sentence Ranking — Bringing Support Workers and Families to a Shared Problem Space
- 3.5 - Family Civic Data Cards — Things to Think With
- 3.6 - Personal Data Examples — Making Data Relatable
- 3.7 - Home Interviewing: Card Sorting with a Family in Their Living Room
- 3.8 - Ideation Decks — Combining Random Design Ingredients to Generate New Ideas
- 3.9 - Group Poster Design — A Participant-designed Poster to Advertise Features of Imagined Data Interface Products
- 3.10 - Storyboarding Cards — A Collaboratively-constructed Narrative Created through Discussion from a Palette of Possible Parent and Staff Actions
- 3.11 - Thematic Analysis of Qualitative Data using *Quirkos* for Case Study One
- 3.12 - Spreadsheet-based Quantitative Analysis of Interview Data for Case Study Two
- 3.13 - Pilot Study Recruitment Poster
- 3.14 - How the Case Studies and Peripheral Activities Contribute to This Thesis

Figures in Chapter 4

- 4.1 - Participants' Shared Values Deduced from Sentence Rankings Data
- 4.2 - Current Model of Data Interaction, and Proposed Model of Shared Data Interaction

Figures in Chapter 5

- 5.1 - A Journey Map of Each Participant's Study Progression
- 5.2 - An Example Life Sketch from Interview 1, with Data Handling Companies in Red, Data Types in Blue, and Feelings in Green
- 5.3 - Sankey Overview of Participants' GDPR Requests

- 5.4 - Longitudinal Distribution of Net Changes in Participants' Perceived Power and Trust Scores
- 5.6 - Participants' Perceived Trust in Provider at Different Stages of the GDPR/Study Process

Figures in Chapter 7

- 7.1 - The Two Motivations for HDR: Controlling Your Personal Data Ecosystem and Utilising Your Information About Your Life
- 7.2 - Mapping the Six Wants into Objectives for the HDR Opportunity Landscape

Figures in Chapter 8

- 8. 1 - Obstacles and Resulting Insights in the HDR Opportunity Landscape
- 8. 2 - Life Concept Modelling
- 8. 3 - Mock-up of a Unified TV Viewing History Interface
- 8. 4 - SubsCrab: An Example Application for Ecosystem Detection and Visualisation
- 8. 5 - Some of the Many Aspects of Metadata that Might Exist About a Datapoint or Dataset
- 8. 6 - The Panopticon Structure of the Illinois State Penitentiary
- 8. 7 - Human Values (from BBC R&D research)
- 8. 8 - A Contact-and-Calendar-centric PDS Approach
- 8. 9 - The Scattered Data Relating to a Vacation
- 8.10 - Mock-up of a Unified Interface for a Vacation
- 8.11 - Annotating Data with Semantic Context

Figures in Chapter 9

- 9. 1 - Theory of Change [ToC]: The Four Dimensions of Change
- 9. 2 - HDR Approach 1: Discovery-Driven Activism
- 9. 3 - HDR Approach 2: Building the Human-centric Future
- 9. 4 - Conceptual Model for a Personal Data Store System
- 9. 5 - High Level Data Types
- 9. 6 - Life Information Modelled as Happenings
- 9. 7 - A Simple PDS Life Information Presentation Model
- 9. 8 - Mock-up of Life Information Presented in a PDS Interface
- 9. 9 - Life Partitioning Analogy using a Cluedo™ board
- 9.10 - Mock-up: Browsing by Areas of Life
- 9.11 - Identifying Entity Associations in Data
- 9.12 - Facebook's *world2vec* Model, Semantically Modelling Human Information from Social Media Posts on Facebook
- 9.13 - Identifying the Attributes of Data
- 9.14 - Determining the Nature of a Piece of Data
- 9.15 - Attributes of Data
- 9.16 - Actions One Might Perform on Life Information

- 9.17 - Questions One Might Ask of Life Information
- 9.18 - Example Taxonomies for Life Information Navigation
- 9.19 - HDR Approach 3: Defending User Autonomy and Hacking the Information Landscape
- 9.20 - The Modern 'Black Box' View of Technology
- 9.21 - HDR Approach 4: Winning Hearts and Minds: Teaching, Championing and Selling the Vision
- 9.22 - SILVER Health Data Viewing Interface
- 9.23 - Summary of Generalised Change Strategies for Pursuing Better HDR, Using the ToC Model

Figures in 'Additional Reference Information'

- ARI3.1 - Private Data Viewing Monitor with Viewing Glasses
- ARI4.1 - Extract of Sample Scenario Storyboarding Exercise walkthrough
- ARI4.2 - Example Backing Mat for Storyboard Decks
- ARI5.1 - Screenshot from *Quirkos* During Coding Process
- ARI5.2 - Screenshot from *Quirkos* at End of Coding Process
- ARI5.3 - Screenshot from *Workflowy* During Theme Construction
- ARI7.1 - Screenshot from a Web Augmented version of the *Just Eat* Website

List of Inset Boxes by Chapter

Inset Boxes in Chapter 7

- Definition - A Definition of Human Data Relations (HDR)

Inset Boxes in Chapter 8

- Insight 1 - Life Information Makes Data Relatable.
- Insight 2 - Data Needs to be United and Unified.
- Insight 3 - Data Must be Transformed into a Versatile Material.
- Insight 4 - Ecosystem Information is an Antidote to Digital Life Complexity.
- Insight 5 - We Must Know Data's Provenance.
- Insight 6 - Data Holders use Four Levers of Infrastructural Power.
- Insight 7 - Human-centred Information Systems Must Serve Human Values, Relieve Pain and Deliver New Life Capabilities.
- Insight 8 - We Need to Teach Computers to Understand Human Information.

Inset Boxes in Chapter 9

- Insight 9 - Individual GDPR Requests can Compel Companies to Change Data Practices.
- Insight 10 - Collectives can Compare and Unify their Data and Use it to Demand Change.

- Insight 11 - Automating the Identification of Entities can enhance Machine Understanding and Unburden Life Interface Users.
- Insight 12 - The 'Seams' of Digital Services need to be identified, exploited and protected.
- Insight 13 - It is Possible (and Necessary) to Demonstrate Business Benefits of Transparency and Human-centricity.

Bibliography

Bowyer, A. *et al.* (2022) 'Human-GDPR interaction : Practical experiences of accessing personal data', *CHI '22*.
Marshall, M. (2020) 'Markdown thesis'. available at: https://gitlab.com/mrshll1001/markdown-thesis.