Understanding and Improving Human Data Relations

Alex Bowyer

Contents

Appendices **1**
A The Pilot Study: CHI 2018 Paper 1
B Ethics Approvals 1

Index of Key Ideas, Insights and Contributions **1**

Glossary of Pre-Existing Abbreviations, Names and Terms **4**

Bibliography **8**

Appendices

A The Pilot Study: CHI 2018 Paper

For additional context, see 1.3.1. The paper is on the following pages.

B Ethics Approvals

Index of Key Ideas, Insights and Contributions

This section serves as both a glossary to explain abbreviations and existing terms used, as well as an index to easily locate key HDR-related concepts, ideas and contributions that this thesis contributes.

The Six Wants in Data Relations

Direct Data Relations:

1. Data Visibility - 6.1.1
2. Data Understanding - 6.1.2
3. Data Useability10 - 6.1.3

Indirect Data Relations:

1. Process Transparency - 6.2.1
2. Individual Oversight - 6.2.2
3. Involvement in Decision Making - 6.2.3

HDR Objectives

The objectives are introduced in 7.7. They are explored in Chapter 8]:

Objective 1 - Data Awareness & Understanding - 8.1 Objective 2 - Data Useability10 - 8.2 Objective 3 - Data Ecosystem Awareness & Understanding - 8.3 Objective 4 - Data Ecosystem Negotiability - 8.4 Objective 5 - Effective, Commercially Viable and Desirable HDR Systems - 8.5

HDR Obstacles

The obstacles are collectively shown in Figure 8.1 and are explained in the following subsections of Chapter 8]:

Closed, Insular and Introspective Practices - 8.5.2 Complex and Invisible Personal Data Ecosystems - 8.3.1 Diminishing Individual Agency - 8.4.2 Data Holder Hegemony - 8.4.1 Immobile Data - 8.2.1 Inaccessible Data - 8.1.1, 8.2.1 Insufficient Machine Understanding of Human Data - 8.5.4 Intractable Data Self, the - 8.4.3 Invisible Data - 8.1.1 Lack of Metadata - 8.3.2 Lack of Individual Demand - 8.5.1 Lack of Interoperability - 8.5.4 Lack of Provider Investment - 8.5.3 Non-Interrogable Data - 8.2.1 Personal Data Diaspora, the - 8.1.2 Unmalleable Data - 8.2.1 Unrelatable Data - 8.1.1

HDR Insights

Insights 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.

Insights 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.

HDR Approaches

The four approaches are collectively summarised in 9.6, and explained and illustrated as follows:

Approach 1 - Discovery-Driven Activism - 9.2, Figure 9.2 Approach 2 - Building the Human-centric Future - 9.3, Figure 9.3 Approach 3 - Defending User Autonomy and Hacking the Information Landscape - 9.4, Figure 9.19 Approach 4 - Teaching, Championing and Selling the HDR Vision - 9.5, Figure 9.21

Thesis Contributions

Contribution C1 - An understanding of What People Want in Direct Data Relations Contribution C2 - An Understanding of What People Want in Indirect Data Relations Contribution C3 - The Synthesis and Formulation of the Field of Human Data Relations (HDR) Contribution C4 - A clear delineation of two primary motivators for individuals seeking better HDR Contribution C5 - A map of the HDR landscape, identifying obstacles and insights Contribution C6 - Four identified trajectories for advancing Human Data Relations Contribution C7 - A reframing of data literacy for the HDR space Contribution C8 - Validation and enumeration of supported families' attitudes and needs around civic data Contribution C9 - *Shared Data Interaction* - A proposed model for more efficient and empowering social support relationships that embraces human-centricity Contribution C10 - A model to understand the five different origins of held personal data Contribution C11 - A rich understanding of the lived experience of accessing data using GDPR rights and of motivations for GDPR data access Contribution C12 - Evidence for the impact of knowledge about data handling practices on provider trust and perceived individual power Contribution C13 - Guidance for policymakers, data holders and individuals on how to improve HDR Contribution C14 - A proto-methodology for educating individuals about held data, data access and the data ecosystem

Major Concepts of This Thesis

Auditing Data Holders - 9.2.2 **Categories of Family Civic Data** - ARI4.1 **Categories of Personal Data** - Figure 3.6 **Data Access & Understanding Services** - 9.2.4, 9.4 **Data Literacy in an HDR Context** - see HDR Literacy **Data Cards** - Figure 3.5, 3.5.2, 4.2.1, 4.4.2, Bowyer *et al.* (2018) **Data Wants** - Chapter 6 **Data Relations, Direct** - 6.1 **Data Relations, Indirect** - 6.2 **Digital Self Curation** - 4.4.3, 5.5.2, 6.3 **Ecosystem Detection** - Insight 4 **Ecosystem Information** - 7.7, Insight 4 **Ecosystem Information Display** - Insight 4 **Ecosystem Negotiability** - 7.7, 8.4 **Empowerment (in the context of data wants)** - 6.3 **Free Data Interfaces** - Bowyer (2018) **Hacking the Seams** - Insight 12 **Human Data Relations (HDR)** - 7.2 **HDR Literacy** - 9.5.1 **Human Information Operating System** - 8.2 **Human Information** - see Life Information and Ecosystem Information **Inclusive Data Flows** - 9.3.5 **Information Standards** - 5.5.1, 8.5 **Landscape of HDR Opportunity** - Fig-

ure 8.1, 9, **Figure 9.23** Life Concepts** - 8.1.1 **Life Information Utilisation** - 7.6, Figure 7.1 **Life Information** - Insight 1 **Life Interface Design** - 9.3.1 **Life Partitioning** - 9.3.2 **Locus of Decision Making (LDM)** - 4.4.3 **Meaning in Data** - Figure 2.1, 4.3.2, 5.4.2, 6.1.2, Insight 1 **Perceived Individual Power** - 5.3.4 **Personal Data Diaspora, the** - 8.1.2 **Personal Data Ecosystem Control** - 7.6, Figure 7.1 **Personal Data as a Proxy for Involvement** - 5.4.4, Bowyer *et al.* (2018) **Personal Data Stewardship** - 4.3.3, 5.6 **Proxy Representations of Immobile Data** - Insight 4 **Shared Data Interaction** - 4.2.4 **Surface Information Injustices** - 9.5.1, Insight 12 **Storyboarding Action Cards** - ARI4.3 **Trust - 4.3.4, 4.4.1, 5.3.4, 5.4.4, 6.2.1, Insight 13** Types of Personal Data (by origin)** - Table 5.2 **Useability**10 (as distinct from Usability) - 6.1.3]

Glossary of Pre-Existing Abbreviations, Names and Terms

Action (stage of Personal Informatics) - see SI **Action Research** - a mode of research where cycles of investigation shape future studies **Accessibility Tags (ARIA)** - tags within HTML code that screenreaders use and which can be exploited for seam hacking **Activism** - using vigorous campaigning to bring about political or social change **Agency** - the ability, described in HDI, to act for oneself in a system, see HDI **Barriers Cascade** - a series of obstacles in SI **BBC R&D** - the Research & Development division of the British Broadcasting Corporation, where I did a research internship **Boundary Objects** - tangible objects and representations that help different populations that may think in different terms to collaborate effeectively **Card Sorting** - a technique used in Participatory co-design where participants arrange cards to convey their thinking **CHC - Connected Health Cities** - government initiative behind the SILVER project, which I worked on **Civic Hacking** - technologists or enthusiasts working to reconfigure the way society works **Civil Libertarianism** - argues for the supremacy of individual rights and personal freedoms over imposition by authority **Collection (stage of Personal Informatics)** - see SI **Consent, Dynamic** - ongoing and changeable expression of preference **Consent, Informed** - initial one-time expression of preference **Constructivism** - a belief that new knowledge is formed by developing one's own mental models in order to explain new experiences **Conceptual Anchors** - the mental scaffolds which we use to organise our thinking and human information **Context-aware Computing** - designing systems that take account of the user's situation and varying needs **Co-experience** - bringing participants towards a shared perspective **Cornmarket** - codename for the BBC R&D PDS project I interned with **Critical Algorithm Studies** - research into making computer systems and their behaviour more understandable and accountable **Data Access Request** - see Subject Access Request **Data Brokers** - third parties that buy and sell user's data **Data Controller** - an organisation responsible for collecting and storing user data **Data Download**

Portal - a website or service that allows users to access held data in a 'self-service' manner **Data Flow Auditing** - the use of apps or system services to monitor the communication and data sharing behaviours of consumer apps **Data Justice** - research into the relationship between datafication and social justice **Data Portability Request** - a particular kind of access request focussed on retrieving data in a machine-readable and useable format **Data Processor** - one who handles user data on behalf of a data controller **Data Provenance** - the history and origin of a piece of data **Data Self** - the representation of an individual in data through which state or commercial actors understand that person when making decisions **Data Subject** - the individual about whom data is stored **Data Transcendence** - the idea that data should not be tied to a single machine, but should move freely to the places it is needed **Data (general)** - digitally-encoded human information **Data, Civic / Family Civic** - the data stored about families by social care organisations such as Early Help programmes **Data, Acquired** - personal data that has been obtained from an official or public source or a third party **Data, Derived** - new data that has been extrapolated through interpretation of existing data **Data, Metadata** - data about the data itself, or about the incident recorded in data **Data, Observed** - data collected about individuals automatically, as a by-product of other actions or in the background **Data, Personal** - data about or related to identifiable individuals **Data, Volunteered** - personal data that has been knowingly shared by that individual with an organisation **Data, Trapped** - data that is hard to access due to technical, commercial or other restrictions **DERC** - Digital Economy Research Centre, funders of the Healthy Eating Web Augmentation project **Design, Co-** - researchers and participants collaboratively exploring problems and solutions **Design, Adversarial** - the design of systems or processes that challenge current norms **Design, Design After** - the ability to repurpose products or services for new objectives that might not have been initially considered **Design, Disrespectful** - design practices (often in user interfaces) that prioritise commercial needs over human convenience **Design, Experience-centred** - design that focusses on human psychology and lived experience of the situation **Design, Magical** - the presentation of technology offerings as powerful and mysterious, that need not be examined or understood **Design, User-centred Design** - design approach that builds up an understanding of user needs that is subsequently used by technical experts **Design, Value-centred Design** - designing in ways that focus on positive impacts on human life **Device Tenancy** - the idea that we are no longer owners of our devices, and lack control and autonomy over their use **Digital Civics** - a multi-disciplinary research field encompassing HCI, governance, education, planning, social science and computer science, practiced by Open Lab, where I studied this PhD **Digital Self** - see Data Self **DIKW pyramid** - see Wisdom Curve **DPA** - Data Protection Authority** - the official authority for regulating data use in a given country **DPO** - Data Protection Officer** - the individual legally responsible for managing data use within a company and for handling access requests **EPSRC** - Engineering and Physical Sciences Research Council, funders of this PhD research **Early Help** - a set of programmes in UK local

authorities designed to identify 'at risk' families and help them with targeted interventions **Effective Access** - ensuring that individuals have all necessary skills, systems and capabilities necessary to see and understand information **Embodied Interaction** - allowing users to create their own practices in information interaction **Empowerment in Use** - having freedom to use products and services in the way you want **Entities** - people, places, organisations, brands, topics or other identifiable 'things' that could be a stakeholder of, or related to, a piece of data **Entity Extraction** - the process of identifying real-world entities in data **Explainable AI** - algorithms whose decision making processes are described to system users **Faceted Search** - the ability to search information by its shared aspects **File Biography** - the lifetime of past actions on a computer file **Files, why they need to die** - article by myself (Bowyer, 2011) **GDPR** - General Data Protection Regulation - the EU's 2018 regulations that give users rights over the collection and use of their personal data **Gatekeeper** - One who controls the flow of data or information between an organisation and an individual **HCI - Human Computer Interaction** - research and practice that explores how people relate to and use computer systems **HDI - Human Data Interaction** - a subdiscipline of HCI that focuses on people's relationship with data, rather than with the system **Hestia.ai** - Swiss company working in the data access and understanding services space, which I currently work for **HII - Human Information Interaction** - a discipline in library sciences that considers how humans relate to information regardless of technology used **Humane Technology** - a movement focused on making technology that is more sensitive to people's lives and needs **ICO - Information Commissioner's Office** - the UK's Data Protection Authority **Ideation Deck** - a participatory design technique that uses 'ingredients' cards shuffled in a grid to generate new ideas **Individualism** - the pursuit of one's own objectives as a primary objective **Information (general)** - facts and assertions understood by interpreting data **Information, Human** - information about people that can be related to their lives or to their digital world **Information, Life** - information about people's lived experience that can be found within data **Information, Ecosystem** - information about people's data, where it is stored, and how it is used and shared **Information Landscape** - the general terrain of available information that a user can see and interact with through the services and apps they use **Infrastructural Power** - see Power, Infrastructural **Integration (stage of Personal Informatics)** - see Self Informatics **Interoperability** - getting systems to connect and exchange information through data standards or conversion **Interoperability, Adversarial** - making systems connect together in ways that were not intended by manufacturers **Legibility** - the ability, as defined in HDI, of being able to understand stored data **Lifelogging** - the practice of maximal data capture for personal SI benefit **Lenses** - different ways of focussing on some data or information according to the aspect of interest or the current role **Life Sketching** - a process of mapping out mental models of one's life on paper **MyData** - an organisation whose members pursue a human-centric change agenda **Negotiability** - the ability described in HDI to flexibly adapt and change one's preferences as the world or digital system changes **NER - Named**

Entity Recognition - see Entity Extraction **Open Lab** - the research lab in Newcastle University in which I conducted this PhD research **Orienteering** - an associative process of information-finding **PDS - Personal Data Store** - See Personal Data Lockers **PIM - Personal Information Management** - the 1990s/2000s discipline that focused on new ways to manage and interact with data and information **PIM systems, contextual** - PIM systems that organise information according to what context it relates to **PIM systems, networked** - PIM systems that focus on the relationships between different pieces of information **PIM systems, semantic** - PIM systems that focus on the underlying meaning of the stored data **PIM systems, spatial** - PIM systems that focus on arranging data in a virtual space for easier management **PIM systems, subjective** - PIM systems that focus on the varied individual needs of users **PIM systems, temporal** - PIM systems that represent information using timelines or other visualisations that highlight change over time **PIMS** - Personal Information Management Services** - See Personal Data Lockers **Participatory Action Research** - see Action Research **PDE - Personal Data Economy** - the emergent marketplace of companies innovating and offering services relating to the management, self-exploitation or harnessing of one's personal data **Personal Data Ecosystem** - the network of systems, accounts, files and digital information that constitutes an individual's digital life **Personal Data Lockers** - a place to store personal data so that it can be united, unified and interpreted by the data subject **Personal Data Vault** - See Personal Data Lockers **Personal Informatics** - see SI **Preparation (stage of Personal Informatics)** - see SI **Perspectives** - different presentations or aspects of information that support different mental models, focus or tasks **Point of Severance** - the point at which data is handed over, beyond which data subjects lose visibility, control and influence **Power - Behavioural Influence** - persuading others to carry out the desired behaviour **Power - Interpretative Influence** - determining how reality is externally represented **Power - Network Centrality** - becoming an indispensable hub of a wider ecosystem **Power, Authority** - ownership of technology or infrastructure **Power, Disciplinary** - using an influential position to affect others' mental models **Power, Infrastructural** - a model of understanding how providers exert power over their users, created as part of the digipower investigation **Power, Interpretive** - creating the internal representations of reality within an organisation **Power, (power to)** - an individual's ability to act (see Agency) **Power, (power over)** - a dominant actor's ability to limit or manipulate the actions of others **Power, Obscure** - where the subservient cannot tell when they are watched **Power, Pervasive** - where the one in power can see everything all the time **Power, Processual** - changing processes for competitive advantage **Power, Rational** - controlling decision-making processes **Power, Resource Control** - controlling the flow of resources **Power, Social** - power where the power holder attempts to influence the behaviour of individuals in pursuit their desired outcomes **Power, Socially-shaped** - influencing a wide audience to settle upon a preferred interpretation **Power, Systems/Structural** - see Infrastructural Power **Power, Zero Sum** - winning a battle for ownership/resource control at the other party's expense **Power Imbalance (over**

Personal Data) - the established fact that data holders have more power in service relationships than data subjects. **Pragmatism** - an epistemology that believes knowledge is constantly renegotiated by individuals **QSM - Quantified Self Movement** - see SI **R&D** - Research & Development **Reflection (stage of Personal Informatics)** - see SI **Recursive Public** - a community of people who are attempting to reconfigure society for the better **SAR - Subject Access Request** - a request to a DPO of an organisation for a copy of held personal data **SI - Self Informatics** - an umbrella term for Personal Informatics and the Quantified Self Movement, where people track their activity in data and reflect upon it, setting goals and tracking progress **SILVER** - the project working in the Early Help space that I worked with for Case Study One **Sitra** - Finnish non-profit research organisation for which the digipower investigation was conducted **Scraping** - the process of programmatically extracting information from interfaces such as websites that were intended for human browsing. **Seams** - the 'edges' of products and services, at which service providers can exert restrictions and at which users can find new ways to adapt their product usage and data access **Support Worker** - a specialist social worker who helps a family in an Early Help context **Supported Family** - a family participating in an Early Help social care programme **Text Mining** - the process of programmatically examining textual data to infer new facts and assertions from the data **ToC - Theories of Change** - a model for thinking about how to achieve change in society **Things to Think With** - the idea that tangible representations can be useful to aid discussions **Timelines** - visual representations of information anchored against points in time **TrackerControl** - see Data Flow Auditing. **Troubled Families** - historic term for those families targeted for help by programmes such as Early Help **VRM - Vendor Relationship Management** - a model where vendors are selected by customers in response to their published needs, instead of relying on broadcast advertising to find customers **Web Augmentation** - the process of modifying a web page to provide new functionality or access data after it has been downloaded to a user's we browser **Web Extensions** - pieces of user code that are loaded into a web browser to modify or programmatically interrogate web pages **Wisdom Curve** - the process of converting data, to information, to knowledge, to wisdom **world2vec** - a system in Facebook that attempts to understand the world through analysis of social media content

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

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