Understanding and Improving Human Data Relations

Alex Bowyer

Table of Contents

# Appendices

## A The Pilot Study: CHI 2018 Paper

For additional context, see [1.3.1](#Xcefc2f8651dab1dca3f5569d5c8495d75f8956a). 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.

### HDR Wants

Direct Data Relations:

1. Data Visibility - [6.1.1](#want-d1)
2. Data Understanding - [6.1.2](#want-d2)
3. Data Useability[10](#fn10) - [6.1.3](#want-d3)

Indirect Data Relations:

1. Process Transparency - [6.2.1](#want-i1)
2. Individual Oversight - [6.2.2](#want-i2)
3. Involvement in Decision Making - [6.2.3](#want-i3)

### HDR Objectives

The objectives are introduced in [7.2.4](#X3985f3ebbcec0cf2e6c7ba8fe728e40700e225f). They are explored in 7.3:

Objective 1. Data Awareness & Understanding - [7.3.1](#X993934f31a10ddda5a7fb5e979600aa8bc7d69d) Objective 2. Data Useability[10](#fn10) - [7.3.2](#X449b014b7ebab4885168edd640cc6aa39b7762a) Objective 3. Data Ecosystem Awareness & Understanding - [7.3.3](#Xead27062f705c3ed0f2b1ac2ad2a7f47418176c) Objective 4. Data Ecosystem Negotiability - [7.3.4](#X140dbb5fccd129ea40bd0f3125865b5adce3acb) Objective 5. Effective, Commercially Viable and Desirable HDR Systems - [7.3.5](#Xf3a1125e1b784cbe6cb78356d9b9ca5b7741fcf)

### HDR Obstacles

The obstacles are collectively shown in [Figure 7.3](#figure-7.3) and are explained in the following subsections of 7.3:

* Closed, Insular and Introspective Practices - [7.3.5](#obstacle-self-interest-interop)
* Complex and Invisible Personal Data Ecosystems - [7.3.3](#obstacle-complex-invisible-eco)
* Diminishing Individual Agency - [7.3.4](#X61c02b3f97760ee3c8583db1659a97df9e7b532)
* Data Holder Hegemony - [7.3.4](#Xeb34d101d5161623ed77334cea0f47e8a0c0af7)
* Immobile Data - [7.3.2](#X253ca7a0ae189192d482fce5658d985559bcf22)
* Inaccessible Data - [7.3.1](#X76859212715c5665bfc5f937ffe5f4ec7d39143), [7.3.2](#X253ca7a0ae189192d482fce5658d985559bcf22)
* Insufficient Machine Understanding of Human Data - [7.3.5](#obstacle-lack-machine-understanding)
* Intractable Data Self, the - [7.3.4](#X84a3a078bae4004a69bca503806ceb7ed0684b4)
* Invisible Data - [7.3.1](#X76859212715c5665bfc5f937ffe5f4ec7d39143)
* Lack of Metadata - [7.3.3](#obstacle-lack-metadata)
* Lack of Individual Demand - [7.3.5](#obstacle-demand)
* Lack of Interoperability - [7.3.5](#obstacle-self-interest-interop)
* Lack of Provider Investment - [7.3.5](#obstacle-self-interest-interop)
* Non-Interrogable Data - [7.3.2](#X253ca7a0ae189192d482fce5658d985559bcf22)
* Personal Data Diaspora, the - [7.3.1](#obstacle-diaspora)
* Unmalleable Data - [7.3.2](#X253ca7a0ae189192d482fce5658d985559bcf22)
* Unrelatable Data - [7.3.1](#X76859212715c5665bfc5f937ffe5f4ec7d39143)

### HDR Insights

Insights in [7.3](#X96c51c3d98f021d42ee8c458ed421add6b4adde):

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

Insights in [7.4](#X40141584308035bb03b454584dbe23925c8bab3):

* [9](#insight-9) - Individual GDPR Requests can Compel Companies to Change Data Practices.
* [10](#insight-10) - Collectives can Compare and Unify their Data and Use it to Demand Change.
* [11](#insight-11) - Automating the Identification of Entities can enhance Machine Understanding and Unburden Life Interface Users.
* [12](#insight-12) - The ‘Seams’ of Digital Services need to be identified, exploited and protected.
* [13](#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 [7.5](#Xd8b45c5920a4ae6b8956c42dcd24c7e655d0317), and illustrated and explained individually in subsections of [7.4](#X40141584308035bb03b454584dbe23925c8bab3):

* Approach 1. Discovery-Driven Activism - [7.4.2](#X4e136873d85cb62901bb699285d9280e05828df)
* Approach 2. Building the Human-centric Future - [7.4.3](#X414dd37628fe75b8fd90976be90a8508ebd49ff)
* Approach 3. Defending User Autonomy and Hacking the Information Landscape - [7.4.4](#Xfcbca708d249a5264bfb58f121b983b931f9729)
* Approach 4. Teaching, Championing and Selling the HDR Vision - [7.4.5](#X56be413628d0da2da37a214dd577f73f291b0fe)

### Contributions and Major Concepts of This Thesis

* Auditing Data Holders - [7.4.2](#flow-audit)
* Categories of Family Civic Data - [ARI4.1](#ari-fcd-types)
* Categories of Personal Data - [Figure 3.6](#figure-3.6)
* Data Access & Understanding Services - [7.4.2](#daus), [7.4.4](#Xfcbca708d249a5264bfb58f121b983b931f9729)
* Data Literacy in an HDR Context - see HDR Literacy
* Data Cards - [Figure 3.5](#figure-3.5), [3.5.2](#X464e9f2e5e0cf52e9c26d3864e71249d73ea33a), [4.2.1](#Xc2fcfc265ab40ba82ed064a42903a2f576bb84b); [4.4.2](#X2ad50ca4cbd63a3f83ddbd28315b55b52f600fb), Bowyer *et al.* ([2018](#ref-bowyer2018family))
* Data Wants - [Chapter 6](#chapter-6)
* Data Relations, Direct - [6.1](#X8258c05de79d267cff0777b650c09dd0e24396f)
* Data Relations, Indirect - [6.2](#X86a6a7bbe6a4ae5c1502558f013f6ed6b9ae6f2)
* Digital Self Curation - [4.4.3](#X8a50e98458a9c28886ed15ffb2cc666b2d3d49b), [5.5.2](#X279d3e70c4a8279cdfb499a60bef2c4f405d995), [6.3](#X3c10c50990743199cc887aaacd3f88a0a0a026e)
* Ecosystem Detection - [Insight 4](#insight-4)
* Ecosystem Information - [7.2.4](#X3985f3ebbcec0cf2e6c7ba8fe728e40700e225f), [Insight 4](#insight-4)
* Ecosystem Information Display - [Insight 4](#insight-4)
* Ecosystem Negotiability - [7.2.4](#X3985f3ebbcec0cf2e6c7ba8fe728e40700e225f), [7.3.4](#X140dbb5fccd129ea40bd0f3125865b5adce3acb)
* Empowerment (in the context of data wants)\*\* - [6.3](#X3c10c50990743199cc887aaacd3f88a0a0a026e)
* Family Design Games - see Bowyer *et al.* ([2018](#ref-bowyer2018family))
* Family Facts - see Bowyer *et al.* ([2018](#ref-bowyer2018family))
* Free Data Interfaces - Bowyer ([2018](#ref-bowyer2018freedata))
* Human Data Relations (HDR) - [7.2](#Xd90f00e19f5543904caf9ab2abd5b800e0613c0)
* HDR Literacy - [7.4.5](#X33b4cf25384363ac565ef3c7838bb451c8a241b)
* HDR Wants - see Data Wants
* Human Information Operating System - [7.3.2](#info-os)
* Human Information - see Life Information and Ecosystem Information
* Inclusive Data Flows - [7.4.3](#X8406c5db4ee9e20d3a43eab68ad54939a4d0647)
* Information Standards - [5.5.1](#X18b35402f3fbcb74b8e5bbc8aa2c0e3e784e473), [7.3.5](#Xf3a1125e1b784cbe6cb78356d9b9ca5b7741fcf)
* Landscape of HDR Opportunity - [Figure 7.3](#figure-7.3), [7.4](#X40141584308035bb03b454584dbe23925c8bab3), [Figure 7.36](#figure-7.36)
* Life Concepts - [7.3.1](#X993934f31a10ddda5a7fb5e979600aa8bc7d69d)
* Life Information Utilisation - [7.2.3](#X7027e4bdfd232de99e156aba3e4d77d931c643b), [Figure 7.1](#figure-7.1)
* Life Information - [Insight 1](#insight-1)
* Life Interface Design - [7.4.3](#Xc997691e73c532cdcc9a952ed6fc5ed481fa2ee)
* Life Partitioning - [7.4.3](#Xc381dfe67e88860a35f9a3a53f1be150e366992)
* Locus of Decision Making (LDM) - [4.4.3](#X8a50e98458a9c28886ed15ffb2cc666b2d3d49b)
* Meaning in Data - [Figure 2.1](#figure-2.1), [4.3.2](#X4bf33bfd9c3793655f2a19fca0ee1ca41e62e6e), [5.4.2](#Xd9a9eb3723d36c07bbcef114c168744dedfa05e), [6.1.2](#want-d2), [Insight 1](#insight-1)
* Perceived Individual Power - [5.3.4](#X17fee6fbf9ee82826dccbd5118d485029a82ebe)
* Personal Data Diaspora, the - [7.3.1](#obstacle-diaspora)
* Personal Data Ecosystem Control - [7.2.3](#X218bee7cd760788b3c5ee66a16fde8ab56b5209), [Figure 7.1](#figure-7.1)
* Personal Data as a Proxy for Involvement - [5.4.4](#X38f625be2c0c748970e0e254c020e238dedd97d), Bowyer *et al.* ([2018](#ref-bowyer2018family))
* Personal Data Stewardship - [4.3.3](4.3.3.4), [5.6](#X2246273db175b4d951c8ae09a9b87d55b422463)
* Pushing the Seams - [Insight 12](#insight-12)
* Proxy Representations of Immobile Data - [Insight 4](#insight-4)
* Shared Data Interaction - [4.2.4](#Xdecc0e8ff84b4e8fdb98d4490fcb5b15d4e95fa)
* Surface Information Injustices - [7.4.5](#X33b4cf25384363ac565ef3c7838bb451c8a241b), [Insight 12](#insight-12)
* Storyboarding Action Cards - [ARI4.3](#ari-storyboarding)
* Trust - [4.3.4](#Xbab51b354b67876c6284de28df0e549940fb873), [4.4.1](#X1cbcb0941f4181aab74224b7105fd44cb905d42), [5.3.4](#X17fee6fbf9ee82826dccbd5118d485029a82ebe), [5.4.4](#X38f625be2c0c748970e0e254c020e238dedd97d), [6.2.1](#want-i1), [Insight 13](#insight-13)
* Types of Personal Data (by origin) - [Table 5.2](#table-5.2)
* Useability[10](#fn10) (as distinct from Usability) - [6.1.3](#want-d3)]

# 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 **thority** 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 **countable** 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 **rson 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](#ref-bowyer2011filesdie)) **GDPR** - General Data Protection Regulation - the EU’s 2018 regulations that give users rights over the collection and **e 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 **formation 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 **e 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 **stem 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** 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 **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 **n find new ways to adapt their product usage and data access** Subject Access Request\*\* - a request to a DPO of an organisation for a copy of held personal data **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](https://doi.org/10.1145/3173574.3173710).