Understanding and Improving Human Data Relations

Alex Bowyer

Contents

Appendices

Appendix A: The Pilot Study: CHI 2018 Paper

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

Appendix B: Case Study One Methodology: Additional Details (Storyboarding Cards)

Appendix C: Case Study Two Methodology: Additional Details

In this section, the methodology used for the analysis of data from Case Study Two is explained. The content of this appendix is identical to Appendix 3 in the Supplemental Materials of the CHI 2022 paper from this study (Bowyer, Holt, *et al.*, 2022). Case Study Two was written first as a paper and then expanded to produce Chapter 5. While the paper was co-written, Chapter 5 was written entirely by Alex Bowyer.

All coding was carried out by Alex Bowyer and Jack Holt, who followed the following process over a nine-month period, comprising at least 200 person-hours:

1. **EXTRACTION AND ANALYSIS OF SEMI-QUANTITATIVE DATA**: Identifying closed question (or brief) responses that might be processable quantitatively.
2. **TEXT FILE PROCESSING**: Splitting, organising, anonymising and some cleaning of auto-transcribed and time-coded text files.
3. **CATEGORISATION INTO CSVs**: Categorised extraction of time-coded text sections from text files into cells of 6-topic spreadsheet, then generation of CSV files for importing into Quirkos Cloud (Daniel Turner, 2014)
4. **INDUCTIVE CODING**: Importing of CSVs into Quirkos Cloud and labelling by Participant, Company, and Topic. Inductive coding of source texts, ensuring good coverage per topic and per participant.
5. **REDUCTIVE CYCLES**: Reductive cycles of merging, renaming and reorganising the codes hierarchy, resulting in 10 top-level codes with hierarchies of coded texts underneath them.
6. **THEME IDENTIFICATION & QUOTE EXTRACTION**: Construction of 3 paper-focussed themes using Workflowy (Turitzin and Patel, 2010) and quote gathering using the organised codes hierarchy.

Some additional detail on the stages:

1. Semi-Quantitative Data Extraction & Analysis

Prior to beginning coding the data, responses to some key closed questions from

the transcripts were combined with field notes, response emails from companies forwarded by participants, sketches and tables from Interview 1/2, data from the interview 2/3 spreadsheet cells, and other data collected, and used to populate a spreadsheet that featured summaries of those responses. For example, where participants had been asked to outline their hopes for the outcomes of their GDPR data requests, these responses were recorded on the spreadsheet to be used as a resource for summarising participant hopes in a manner that could be easily quantified and referred back to. In some cases this data was analysed within the spreadsheet to produce insights, graphs and percentages. Such data was later used to support and illustrate findings from the coding process. This spreadsheet also included important information relating to each participant's GDPR process experience, such as the timeliness and completeness of their data returns, which could serve as a reference point when analysing the transcripts.

The semi-quantitative data areas captured or derived from captured data were:

- Company Response Timelines
- Power Scores
- Trust Scores
- Hopes, Goals and Imagined Uses
- Term Definitions
- For each participant + target company + data type (+ subtype in some cases):
- Provided or Not?
- Perceived Value
- Completeness
- Understandability
- Accuracy
- Useability
- Usefulness
- Meaningfulness
- Feelings about data (general, and company-specific)
- General questions (general, and company-specific)
- Best and Worst Companies (taking into account provided, completeness, understandability, accuracy, usability, usefulness)
- Sankey analysis of participant journeys

2. Text File Processing (Splitting & Recombination)

The researchers then moved on to prepare for the fully qualitative analysis. All interview audio was auto-transcribed using Zoom and Google Recorder, and then the generated text files were cleaned. Cleaning consisted of listening to sections of audio where transcription seemed inaccurate and correcting the transcripts. Due to the volume of data this cleaning was not done for all texts, only where ambiguity or typos meant it was needed for accurate coding and for quotes. Some anonymisation of source texts was also carried out at this stage and later, with a particular focus on quotes included in the chapter. The researchers used this data preparation stage as an initial means of (re)familiarising with the

dataset. With reference to the structured interview schedules, the initial 33 text transcripts were split up by participant, company and topic using the labelling scheme outlined in 'Text File Labelling Strategy' below.

At the end of this process, roughly 100 'pieces' had been identified for each participant (slightly more for P11 whose interview 1 covered a broader scope and considerably less for P9 who only did interview 1).

3. Categorisation into CSVs

The pieces from stage 1 were then recombined, across all participants, into 233 source files. These 233 source files were then further grouped into 6 topics areas. (The aim of the analysis was to identify common opinions and ideas around different topics, not to explore individual participant journeys end-to-end). The six topic areas were:

1. **POWER** – discussions and scoring around the power of data holding companies
2. **TRUST** – discussions and scoring around participants' subjective trust in data holding companies
3. **LIFE** – life sketching and annotation discussions, and 'digital life' questioning
4. **HOPES & USES** – discussions around motivations, expectations, goals and hopes, and imagined uses of data
5. **COMPANY-SPECIFIC** – (repeated once per target company per participant) – all discussions around the data return from a particular company
6. **GENERAL** – all non-company specific discussions not captured elsewhere

This produced too many files for import into Quirkos Cloud, so once organised by topic, these six groups of files were further combined into 11 General files and 46 Company-Specific, files (with **Life** and **General** going into the General files and everything else going into **Company-Specific**). This gave 57 organised CSV files ready for use in the first coding phase.

4. Inductive Coding

The majority of the analysis took place with the use of Quirkos Cloud (Daniel Turner, 2014), a computer-assisted qualitative data analysis software (CAQDAS) package that allows for collaborative analysis by more than one researcher. The 57 files from stage 3 were imported into Quirkos Cloud, with each having a unique number. The sources in Quirkos were labelled by Participant, Company and Topic for easy search and retrieval. The researchers then collaboratively coded sections of the interview transcripts to develop and ensure a consistent approach, based on established techniques (Huberman and Miles, 2002; Braun and Clarke, 2006). Codes were identified inductively and not according to a fixed or predetermined set. Once a baseline codeset and strategy had been established, they each coded sections of interviews in parallel, regularly regrouping to discuss generated codes and any new questions or challenges arising. At first, these codes were created in an unstructured/flat state with only occasional clustering

on the Quirkos interface. Due to the volume of data, not every piece of every transcript was coded, however care was taken to ensure a representative sample of views from across the participant pool was included. These were clustered into loose code-topic areas, an example is shown in the following screenshot taken approximately 6 weeks into coding:
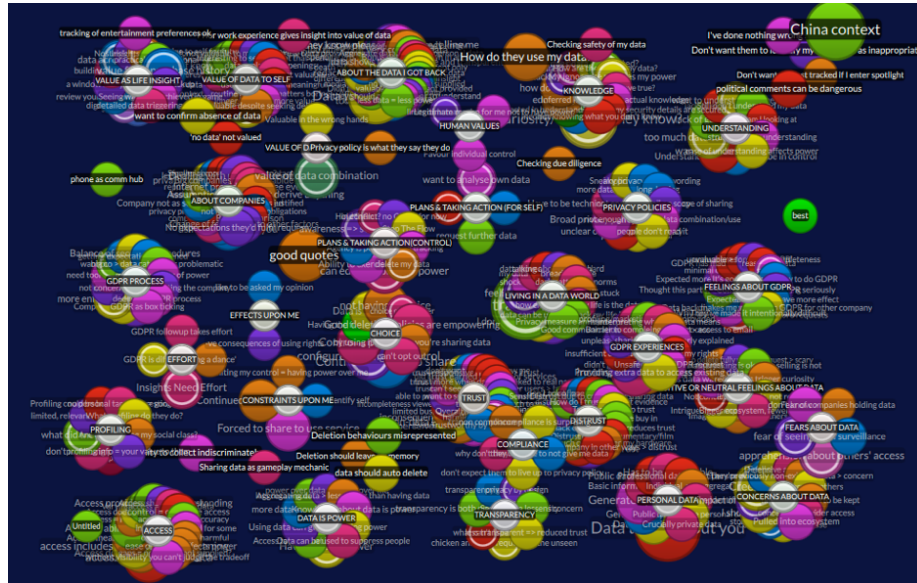

Figure 1: Figure C.1: Screenshot from Quirkos During Coding Process

5. Reductive Cycles

As more codes were identified and structures and commonalities between them were formed, existing codes were merged or absorbed into one another and grouped together in small clusters. The researchers regularly met to discuss each other's codes according to their context and occasionally amended wording or merged concepts that were labelled differently but semantically equivalent. All codes were checked and agreed between these two researchers. Over time, the codes were iteratively structured and restructured, creating top-level thematic clusters around different research questions that held multiple layers of related codes. These clusters were then summarised with a short sentence or paragraph of text, allowing summaries to be produced at different levels of hierarchy. These summaries were kept in the Description fields of codes in Quirkos and also in external structured text-based documents. These can be seen in the following screenshot, taken 5 months into coding:

The above-pictured structure of the coded corpus at the end of the Quirkos Cloud phase was as follows:

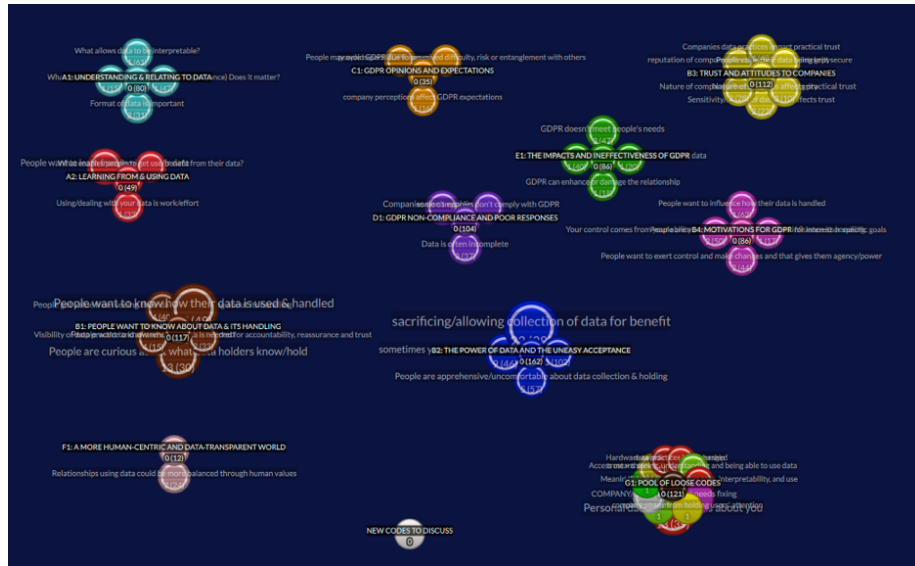- A (129 codes): What do people/need want from their data and how do

Figure 2: Figure C.2: Screenshot from Quirkos at End of Coding Process

they feel about data?

- A1 (80 codes): **Understanding and Relating to Data**: People want to understand and need to be able to relate to data.
- A2 (49 codes): **Learning from and Using Data**: People want to learn more from and do more with their data.
- B (279 codes): What do people feel about the data-centric world?
  - B1 (117 codes): **People want to know about data and its handling**: People want to know what data exists and how it is handled, so they can understand what companies do to hold them to account, and inform their choices/trust.
  - B2 (162 codes): **The Power of Data and the Uneasy Acceptance**: People accept a certain amount of data collection and use but are apprehensive and sometimes feel they have no choice, because data holding is a form of power.
  - B3 (112 codes): **Trust and Attitudes to Companies**: Trust placed in companies is influenced by both the nature and reputation of the company, as well as what data they hold and how that data is kept and handled.
  - B4 (86 codes): **Motivations for GDPR**: People want to make use of their data and influence how it is handled and see GDPR as having the potential to help them achieve this.
- C (35 codes): What do people think about GDPR?
- C1 (35 codes) **GDPR Opinions and Expectations**: People's expectations for GDPR are affected by their perception of the company and its

perceived difficulty, risk and entanglement; people expect non-compliance.

- D: What is the experience of GDPR as a means to gain awareness of and access to usable and understandable data?
- D1 (104 codes) **GDPR Non-Compliance and Poor Responses**: The data returned from GDPR is often incomplete, hard to deal with, lacking explanation, or poorly formatted. Many companies are not complying.
- E: What is the experience of GDPR as a means to gain influence and achieve goals with data/What is the practical impact of GDPR?
- E1 (86 codes) **The Impacts and Ineffectiveness of GDPR**: People's interest in GDPR comes from curiosity to exert their rights or from specific questions about data handling or data use goals. GDPR rarely delivers upon on any of their goals but it does change people's outlook and affects the relationship with the data holder.
- F: How should the world change or be different?
- F1 (12 codes) **A more human-centric and data-transparent world**: People want companies to provide greater transparency and data control/agency and act in a more human manner so they can trust them.
- G: Loose/ungrouped codes (121 codes)

Total codes = 645.

6. Theme Identification & Quote Extraction

Having produced the structure above as a reduced representation of *'what the codes say'* that the participants think, the researchers used outlining tool Workflowy (Turitzin and Patel, 2010) to develop the arguments and primary narrative of the chapter into a structured three-theme-based summary of the most important items from these findings. The code hierarchy was used as source material to populate the three key themes with illustrative quotes and observed findings. An example from later in this process (around 8-9 months since Stage 1 began) is shown in the screenshot below:

The themes are broken down in detail in 5.4 and can be summarised as:

1. **Insufficient Transparency**: Organisations appear evasive over data when responding to GDPR, leaving people "in the dark" even after making GDPR requests.
2. **Confusing Data**: When presented with their data, people struggle to understand it and relate it to their lives and are not able to make use of it.
3. **Fragile Relationships**: Companies' data practices, and in particular their privacy policies and GDPR response handling, can be impactful to customer relationships, carrying a risk of damaging trust but also the potential to improve relations.

In all, the process from commencing data analysis to writing up thematic findings in the chapter took over 200 person-hours over a 9-month period from January to September 2020.

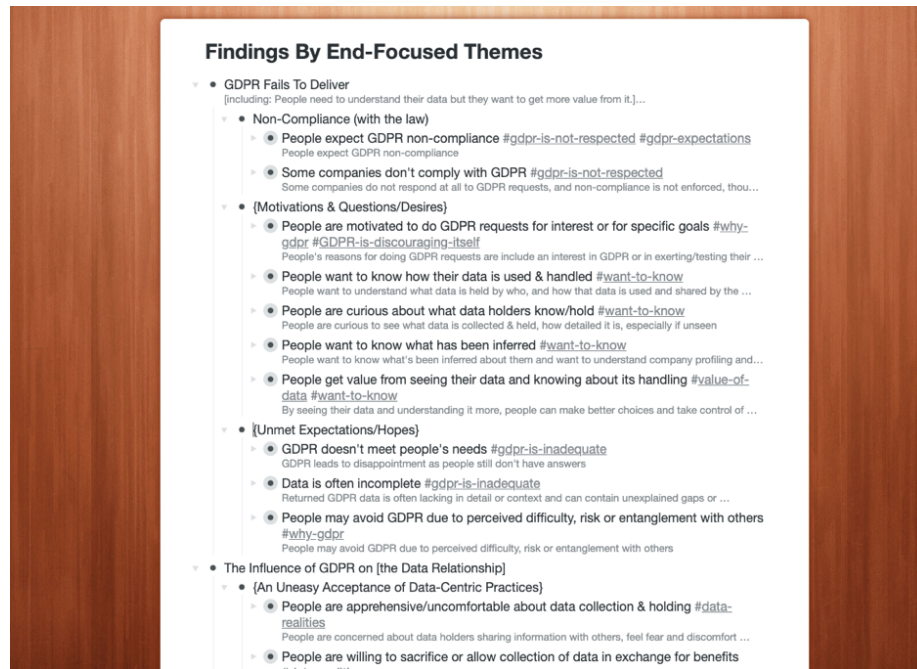**Text File Labelling Strategy used in Stage 2**

Figure 3: Figure C.3: Screenshot from Workflowy During Theme Construction

In stage 2, text files were initially broken down into small pieces and labelled according to the following strategy:

Interview 1 (Sensitisation / Poster Display Chat)

Break into 5 parts:

- `Comp` - list of companies
- `Type` - types of data
- `DoWt` - potential uses of data ['what would you do with the data?']
- `GDPR` - GDPR
- `Motv` - motivation for taking part

Interview 1 (Main Sketch Interview)

Break down as follows:

- `SktR` - review of previous sketch interview from prior study [p11 only]
- `DPer` - definition of personal data
- `DAcc` - definition of access to data
- `DCon` - definition of control of data
- `DPow` - definition of power
- `Sket` - sketching
- `Anno` - annotation

8

- `SelC` - company selection
- `XXXX` - per company [use first four letters of company]
- `Powr` - power
- `Hope` – hopes
- `Uses` – uses
- `Wrap` - [Wrap up]/What happens next

Format: `NN-pXX-iX-[Comp/Type/Uses/GDPR/Motv]-[company first three letters].txt`

e.g. `01-p01-i1-Comp.txt` or `02-p01-i1-Powr-Face.txt`

Interview 2

Break down as follows:

- `XXXX` - per company [use first four letters of company name]
  - `Priv` - viewing privacy policy
  - `Powr` - power
  - `HopU` - hopes & uses
  - `Trst` - trust [p10 & p11]
  - `Pow2` - end power
  - `Trs2` - end trust
  - `Hop2` - end hopes and uses

Format: `NN-pXX-iX-[....]-[company first three letters].txt`

e.g. `01-p01-i2-Priv-Goog.txt`

Interview 3

Break down as follows:

- [intro & consent] - no need to transcribe/code
- `XXXX` - per company [use first four letters of company name]
  - `Powr` - power rating
  - `Trst` - trust rating
  - `RPow` - retro power
  - `RTrs` - retro trust
  - `Hope` - hope (for company) and uses (how well have hopes been met / how practical are the envisaged data uses
  - `Data` - Overall data overview
  - `Prov` - Data provided by you
  - `Indr` - Data indirectly / automatically collected
  - `Derv` - Data derived about you
  - `Othr` - Data from other sources
  - `Meta` – Metadata
  - `GenQ` - general questions about this company
  - `Pow2` - end power
  - `Trs2` - end trust
  - `Next` - what next for this company specifically

- **Genr** - General topics
- **Hope** - Hope (general)
- **Wrap** - Wrap up questions / the future

Format: `NN-pXX-iX-[....]-[company first three letters].txt`

e.g. `01-p01-i3-Cred-Indr.txt` or `02-p01-i3-Genr-Wrap.txt`

Appendix D: Peripheral Research Settings: Additional Details

BBC R&D's Cornmarket Project

I took a three-month sabbatical from my PhD in the summer of 2020. I was remotely embedded within a full-time research internship at **BBC R&D** - the British Broadcasting Corporation (BBC)'s Research and Development (R&D) department (British Broadcasting Corporation, 1997), working with specialists, designers, researchers and developers on an exploratory research project codenamed *Cornmarket*. I continued this involvement as a part-time research consultant and critical friend for a further 5 months after the conclusion of the initial three-month placement.

As part of its Royal Charter, one of the BBC's lesser known obligations is to maintain a centre of excellence for research and development in broadcasting and electronic media, and to this end it employs over 200 researchers in its R&D department looking at everything from AV engineering and production tools to new forms of media, virtual reality, digital wellbeing and human data interaction (British Broadcasting Corporation, 1997). The Cornmarket project, launched in 2019, is a BBC-internal human-data interaction research project which explores a possible role for the BBC as it moves beyond broadcast television, using its public service responsibility to guide citizens to a position of empowerment within today's digital landscape - encompassing not just entertainment but health, finance and self-identity. Due to its unique funding from UK-wide TV licensing and its duties to not only entertain but to inform and educate the general public, the BBC is uniquely placed to take a more human-centred approach than commercial innovators in this space as it needs only to deliver value, not profit. The project is exploring the use of Solid (Berners-Lee, 2022) technology to build a working Personal Data Store (PDS) prototype [2.3.4] while also developing, iterating and trialling user interface designs and conducting participatory research interviews and activities all to explore what for a BBC PDS might take and what features its potential users might value.

The proposed BBC Cornmarket product, internally called *My PDS*, would allow people to populate a PDS with personal data from APIs and data downloads from a variety of services including BBC iPlayer, Netflix, All4, Spotify, Instagram, Strava, Apple Health, banks and finance companies, as well as social media companies such as Facebook, LinkedIn and Twitter, and then to use these

combined data sources to create personal *profiles* for Health, Finance, Media (i.e. entertainment) and Core, within which various data insights, visualisations, capabilities would be delivered. One feature the work explores in depth as potentially valuable to users is the ability to include and exclude certain datapoints from the imported viewing history data in order to present a more accurate, curated view of oneself that could then be fed back to other applications such as BBC Sounds to give better content recommendations.

With a cross-disciplinary team of around 20 people including architects, developers, user experience designers, product designers, innovators, participatory researchers and marketers, and funding to outsource public engagement research to agencies, this project represents a significant player in the emerging personal data economy [2.3.4]. As such the Cornmarket project is a fertile ground in which to learn more from practitioners in the PDE space and to test the learnings of this thesis in practice while also finding deeper insights in response to my research questions - in particular RQ3 which is concerned with the building of more human-centric personal data interfaces in practice.

Much of the work I did during this extended internship can be seen in the designs within 7.4.3, as well as the research report I wrote (Bowyer, 2020a) and internship writeup (Bowyer, 2020b). My work with the Cornmarket project can be seen as the concluding part of one of several action research cycles within my PhD 3.2.2.

A number of articles relating to the Cornmarket project have been published:

- https://www.bbc.co.uk/rd/blog/2021-09-personal-data-store-research
- https://paper.dropbox.com/doc/Building-trusted-data-services-and-capabilities-Us49Ek0nex7yClKughPN4
- https://www.wired.co.uk/article/bbc-data-personalisation
- https://www.theregister.com/2021/10/04/column_data_privacy/
- https://www.thetimes.co.uk/article/bbc-and-sir-tim-berners-lee-app-mines-netflix-data-to-find-shows-viewers-like-lxp002gg8
- https://www.ibc.org/download?ac=18659
- https://www.telegraph.co.uk/business/2022/06/09/bbc-wages-war-online-echo-chambers-unbiased-tech/
- https://parliamentlive.tv/event/index/7d249bcf-78e9-447b-907c-81df72b87542?in=15:01:35

Hestia.ai, and Sitra's *digipower* Project

Following the conclusion of the funded period of my PhD, I took up a near-full-time position as Project Leader and Personal Data Coach at **Hestia.ai** (Dehaye, 2019)), a startup based in Geneva, Switzerland. Hestia.ai is a company conducting research, developing technologies, and delivering training, in the emergent MyData/PDE space [2.3.4]. In essence, the company's mission is to help individuals and especially collectives to more easily obtain and understand data held about them, and to help them visualise, aggregate and make use of

that data. It is an example of a **data access and understanding services** company as described in 7.4.5.3.

I was specifically hired to co-lead the *digipower* project (Härkönen and Vänskä, 2021), for Hestia.ai's client, **Sitra** (Sitra, 1967). Sitra is a non-profit organisation in Finland, funded by the Finnish Parliament and accountable to the Finnish people. The goal of the digipower project was to guide 15 European politicians, civil servants and journalists, through the process of obtaining and exploring their own data. The participants were high-profile VIPs, including the former Prime Minister of Finland and former European Commission Vice President, Jyrki Katainen. The goal was to empower those individuals to better understand the workings of the data economy, so that they might be able to influence others and effect change. One of Sitra's goals is to establish a fairer data economy (Sitra, 2018). Methodologically, the project drew heavily on my own Case Study Two [Chapter 5], adopting a similar method of guiding individuals through the process of making GDPR requests and scrutinising the returned data; I was employed on the project for this expertise. Where it differs from my own Case Study is that the focus of the research was outward, on the data economy and the practices of service providers, rather than inward, on the lived experience of the participants. Other differences included the building and use of software interfaces to provide participants with data visualisations, the use of TrackerControl software to audit mobile phone apps [Insight 12], and the direct analysis of participants' retrieved personal data by the Hestia.ai research team (whereas my Case Study explicitly avoided handling participants' personal data). The project resulted in three reports:

- Sitra's official project report (Härkönen *et al.*, 2022); and
- Two technical research reports by Hestia.ai:
  - A high-level interpretation of models of power and influence in the data economy (Pidoux *et al.*, 2022); and
  - A detailed auditing of provider practices, evidenced by examples from participants' data (Bowyer, Pidoux, *et al.*, 2022).

At the time of publication of this thesis (August 2022), I continue to be employed by Hestia.ai, working on the research, design and development of tools to help collectives [Insight 10] with data, make data easier to understand [6.1.2; 7.2.4], and exploring methods to help people 'hack the seams' of digital platforms and services [7.4.4].

Where the BBC internship [D.1] has helped me to understand the practicalities of connecting people with their personal data in pursuit of Life Information Utilisation [7.2.3.1], my work with Hestia.ai has helped me understand the practicalities of how people might acquire greater Personal Data Ecosystem Control [7.2.3.1]. In this sense, both peripheral activities have been highly complementary to developing an overview of the pursuit of HDR in practice.

Digital Economy Research Centre (DERC)'s Healthy Eating Web Augmentation Project

As a software developer I have been aware for a long time that one of the biggest challenges in building new data interfaces is to gain programmatic access to the necessary data. As part of the trend towards cloud-based services and data-centric business practices, it has become increasingly difficult to access all of the data held about users by service providers. Application Programming Interfaces (APIs) are a technical means for programmers to access a user's data so that third-party applications may be built using that data. Unfortunately, as a result of commercial incentives to lock users in and keep data trapped (Abiteboul, André and Kaplan, 2015; Bowyer, 2018), much of users' data can no longer be accessed via APIs. While GDPR data portability requests do open up a new option for the use of one's provider-collected data in third-party applications, this is an awkward and time-consuming route for both users and developers. **Web augmentation** provides a third possible technical avenue for obtaining data from online service providers. It relies on the fact that a user's data is loaded to the user's local machine and displayed within their web browser every time a website is used, and therefore it is possible to extract that data from the browser using a browser extension; this as another **seam** that can be hacked - see 7.4.4 and Insight 12. Similarly, once loaded into the browser, a provider's webpage can be modified to display additional data or useful human-centric functionality that the provider failed to provide.

In order to better understand what is and is not possible using this technique, I participated part-time from 2018 to 2020 as the sole software engineer in a project which was using the web augmentation technique to improve the information given to users of Just Eat, a takeaway food ordering platform in the UK, the theoretical basis for which was published in (Goffe *et al.*, 2021, 2022). While this particular use case does not concern personal data, the technology and techniques being used by the project to exploit the browser seam were considered highly relevant, and the goals of the research project were also human-centric, and consistent with this thesis's research goals - tackling the hegemony of service providers in order to better serve individual needs.

Appendix E: Ethics Approvals

Appendix F: The Private Data Viewing Monitor

By removing the filter layer on an old monitor and modifying cinema IMAX glasses, a monitor was created that only allowed viewing by the holder of the viewing glasses, which would be ideal for interviewing someone about their data while respecting privacy. Face to face interviewing had to be abandoned due to COVID-19, so this technique was sadly never used in practice.

Figure 4: Figure F.1: Private Data Viewing Monitor with Viewing Glasses

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 within this thesis.

HDR Wants

- **Direct Data Want 1: Data Visibility** - [ADD SECTIONREF]
- **Direct Data Want 2: Data Understanding** - [ADD SECTIONREF]
- **Direct Data Want 3: Data Useability**13 - [ADD SECTIONREF]
- **Indirect Data Want 1: Process Transparency** - [ADD SECTIONREF]
- **Indirect Data Want 2: Individual Oversight** - [ADD SECTIONREF]
- **Indirect Data Want 3: Involvement in Decision Making** - [ADD SECTIONREF]

HDR Objectives

- **HDR Objective 1: Data Awareness & Understanding** - [ADD SECTIONREF]
- **HDR Objective 2: Data Useability**13 - [ADD SECTIONREF]
- **HDR Objective 3: Ecosystem Awareness & Understanding** - [ADD SECTIONREF]
- **HDR Objective 4: Ecosystem Negotiability** - [ADD SECTIONREF]
- **HDR Objective 5: Effective, Commercially Viable and Desirable HDR Systems** - [ADD SECTIONREF]

HDR Obstacles

- **Closed, Insular and Introspective Practices** - [ADD SECTIONREF]
- **Diminishing Individual Agency** - [ADD SECTIONREF]
- **Immobile Data** - [ADD SECTIONREF]
- **Inaccessible Data** - [ADD SECTIONREF]
- **Increasing Data Holder Hegemony** - [ADD SECTIONREF]
- **Insufficient Machine Understanding of Human Data** - [ADD SECTIONREF]
- **Intractable Data Self, the** - [ADD SECTIONREF]
- **Invisible Data** - [ADD SECTIONREF]
- **Lack of Individual Demand** - [ADD SECTIONREF]
- **Lack of Interoperability** - [ADD SECTIONREF]
- **Lack of Provider Investment** - [ADD SECTIONREF]
- **Non-Interrogable Data** - [ADD SECTIONREF]
- **Unmalleable Data** - [ADD SECTIONREF]
- **Unrelatable Data** - [ADD SECTIONREF]

HDR Insights & Approaches

- **HDR Insight 1: Life Information Makes Data Relatable** - [ADD SECTIONREF]
- **HDR Insight 2: Data Needs to be United and Unified** - [ADD SECTIONREF]
- **HDR Insight 3: Data Must Be Transformed into a Versatile Material.** - [ADD SECTIONREF]
- **HDR Insight 4: Ecosystem Information Is an Antidote to Digital Life Complexity]** - [ADD SECTIONREF]
- **HDR Insight 5: We Must Know Data's Provenance.** - [ADD SECTIONREF]
- **HDR Insight 6: Data Holders use Four Levers of Infrastructural Power.]** - [ADD REF]
- **HDR Insight 7: Human-centred Information Systems Must Serve Human Values, Relieve Pain and Deliver New Life Capabilities.** - [ADD REF]
- **HDR Insight 8: We Need to Teach Computers To Understand Human Information.** - [ADD REF]
- **HDR Insight 9: Individual GDPR requests can compel companies to change data practices.** - [ADD REF]
- **HDR Insight 10: Collectives can compare and unify their data and use it to demand change.** - [ADD REF]
- **HDR Insight 11: Automating the Identification of Entities can enhance Machine Understanding and Unburden Life Interface Users.**
- **HDR Insight 12: The 'Seams' of Digital Services need to be identified, exploited and protected.** - [ADD REF]
- **HDR Insight 13: It is Possible (and Necessary) to Demonstrate Business Benefits of Transparency and Human-centricity.** - [ADD REF]

HDR Approaches

- **HDR Approach 1: Discovery-Driven Activism** - [ADD SECTIONREF]
- **HDR Approach 2: Building the Human-centric Future** - [ADD SECTIONREF]
- **HDR Approach 3: Defending User Autonomy and Hacking the Information Landscape** - [ADD SECTIONREF]
- **HDR Approach 4: Teaching, Championing and Selling the HDR Vision** - [ADD SECTIONREF]

Contributions and Major Concepts of This Thesis

- **Auditing Data Holders** - [ADD SECTIONREF]
- **Categories of Family Civic Data** - [ADD SECTIONREF]

Glossary of Pre-Existing Abbreviations, Names and Terms

- **Action (stage of Personal Informatics)** - see SI
- **Action Research** - [ADD SECTIONREF] [ADD DEFINITION]
- **Accessibility Tags (ARIA)** - [ADD SECTIONREF] [ADD DEFINITION]
- **Activism** - [ADD SECTIONREF] [ADD DEFINITION]
- **Agency** - [ADD SECTIONREF] [ADD DEFINITION]
- **Barriers Cascade** - [ADD SECTIONREF] [ADD DEFINITION]
- **BBC R&D** - [ADD SECTIONREF] [ADD DEFINITION]
- **Boundary Objects** - [ADD SECTIONREF] [ADD DEFINITION]
- **Card Sorting** - [ADD SECTIONREF] [ADD DEFINITION]
- **CHC - Connected Health Cities** - [ADD SECTIONREF] [ADD DEFINITION]
- **Civic Hacking** - [ADD SECTIONREF] [ADD DEFINITION]
- **Civil Libertarianism** - [ADD SECTIONREF] [ADD DEFINITION]
- **Collection (stage of Personal Informatics)** - see SI
- **Consent, Dynamic** - [ADD SECTIONREF] [ADD DEFINITION]
- **Consent, Informed** - [ADD SECTIONREF] [ADD DEFINITION]
- **Constructivism** - [ADD SECTIONREF] [ADD DEFINITION]
- **Conceptual Anchors** - [ADD SECTIONREF] [ADD DEFINITION]
- **Context-aware Computing** - [ADD SECTIONREF] [ADD DEFINITION]
- **Co-experience** - [ADD SECTIONREF] [ADD DEFINITION]
- **Cornmarket** - [ADD SECTIONREF] [ADD DEFINITION]
- **Critical Algorithm Studies** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data Access and Understanding Services** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data Access Request** - see Subject Access Request
- **Data Brokers** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data Controller** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data Download Portal** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data Flow Auditing** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data Justice** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data Portability Request** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data Processor** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data Provenance** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data Self** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data Subject** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data Transcendence** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data (general)** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data, Civic / Family Civic** - [ADD SECTIONREF] [ADD DEFINITION]

- **Data, Acquired** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data, Derived** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data, Metadata** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data, Observed** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data, Personal** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data, Volunteered** - [ADD SECTIONREF] [ADD DEFINITION]
- **Data, Trapped** - [ADD SECTIONREF] [ADD DEFINITION]
- **DERC** - [ADD SECTIONREF] [ADD DEFINITION]
- **Design, Co-** - [ADD SECTIONREF] [ADD DEFINITION]
- **Design, Adversarial** - [ADD SECTIONREF] [ADD DEFINITION]
- **Design, Design After** - [ADD SECTIONREF] [ADD DEFINITION]
- **Design, Disrespectful** - [ADD SECTIONREF] [ADD DEFINITION]
- **Design, Experience-centred** - [ADD SECTIONREF] [ADD DEFINITION]
- **Design, Magical** - [ADD SECTIONREF] [ADD DEFINITION]
- **Design, User-centred Design** - [ADD SECTIONREF] [ADD DEFINITION]
- **Design, Value-centred Design** - [ADD SECTIONREF] [ADD DEFINITION]
- **Device Tenancy** - [ADD SECTIONREF] [ADD DEFINITION]
- **Digital Civics** - [ADD SECTIONREF] [ADD DEFINITION]
- **Digital Self** - see Data Self
- **DIKW pyramid** - see Wisdom Curve
- **Double Diamond** - [ADD SECTIONREF] [ADD DEFINITION]
- **DPA - Data Protection Authority** - [ADD SECTIONREF] [ADD DEFINITION]
- **DPO - Data Protection Officer** - [ADD SECTIONREF] [ADD DEFINITION]
- **EPSRC** - [ADD SECTIONREF] [ADD DEFINITION]
- **Early Help** - [ADD SECTIONREF] [ADD DEFINITION]
- **Effective Access** - [ADD SECTIONREF] [ADD DEFINITION]
- **Embodied Interaction** - [ADD SECTIONREF] [ADD DEFINITION]
- **Empowerment in Use** - [ADD SECTIONREF] [ADD DEFINITION]
- **Entities** - [ADD SECTIONREF] [ADD DEFINITION]
- **Entity Extraction** - [ADD SECTIONREF] [ADD DEFINITION]
- **Explainable AI** - [ADD SECTIONREF] [ADD DEFINITION]
- **Faceted Search** - [ADD SECTIONREF] [ADD DEFINITION]
- **Entity Extraction** - [ADD SECTIONREF] [ADD DEFINITION]
- **File Biography** - [ADD SECTIONREF] [ADD DEFINITION]
- **Files, why they need to die** - [ADD SECTIONREF] [ADD DEFINITION]
- **GDPR - General Data Protection Regulation** - [ADD SECTIONREF] [ADD DEFINITION]
- **Gatekeeper**- [ADD SECTIONREF] [ADD DEFINITION]
- **HCI - Human Computer Interaction** - [ADD SECTIONREF] [ADD DEFINITION]

- **HDI - Human Data Interaction** - [ADD SECTIONREF] [ADD DEFINITION]
- **Hestia.ai** - [ADD SECTIONREF] [ADD DEFINITION]
- **HestiaLabs** - [ADD SECTIONREF] [ADD DEFINITION]
- **HII - Human Information Interaction** - [ADD SECTIONREF] [ADD DEFINITION]
- **Humane Technology** - [ADD SECTIONREF] [ADD DEFINITION]
- **ICO - Information Commissioner's Office** - [ADD SECTIONREF] [ADD DEFINITION]
- **Ideation Grids** - [ADD SECTIONREF] [ADD DEFINITION]
- **Individualism** - [ADD SECTIONREF] [ADD DEFINITION]
- **Information (general)** - [ADD SECTIONREF] [ADD DEFINITION]
- **Information, Human** - [ADD SECTIONREF] [ADD DEFINITION]
- **Information, Life** - [ADD SECTIONREF] [ADD DEFINITION]
- **Information, Ecosystem** - [ADD SECTIONREF] [ADD DEFINITION]
- **Information Landscape** - [ADD SECTIONREF] [ADD DEFINITION]
- **Infrastructural Power, and its Four Levers** - [ADD SECTIONREF] [ADD DEFINITION]
- **Integration (stage of Personal Informatics)** - see Self Informatics
- **Interoperability** -
- **Interoperability, Adversarial** - [ADD SECTIONREF] [ADD DEFINITION]
- **Keeping** - [ADD SECTIONREF] [ADD DEFINITION]
- **Legibility** - [ADD SECTIONREF] [ADD DEFINITION]
- **Lifelogging** - [ADD SECTIONREF] [ADD DEFINITION]
- **Lifestreams** - [ADD SECTIONREF] [ADD DEFINITION]
- **Lenses** - [ADD SECTIONREF] [ADD DEFINITION]
- **Life Sketching** - [ADD SECTIONREF] [ADD DEFINITION]
- **MyData** - [ADD SECTIONREF] [ADD DEFINITION]
- **Negotiability** - [ADD SECTIONREF] [ADD DEFINITION]
- **NER - Named Entity Recognition** - see Entity Extraction
- **Open Lab** - [ADD SECTIONREF] [ADD DEFINITION]
- **Orienteering** - [ADD SECTIONREF] [ADD DEFINITION]
- **PDS - Personal Data Store** - See Personal Data Lockers
- **PIM - Personal Information Management** - [ADD SECTIONREF] [ADD DEFINITION]
- **PIM systems, contextual** - [ADD SECTIONREF] [ADD DEFINITION]
- **PIM systems, networked** - [ADD SECTIONREF] [ADD DEFINITION]
- **PIM systems, semantic** - [ADD SECTIONREF] [ADD DEFINITION]
- **PIM systems, spatial** - [ADD SECTIONREF] [ADD DEFINITION]
- **PIM systems, subjective** - [ADD SECTIONREF] [ADD DEFINITION]
- **PIM systems, temporal** - [ADD SECTIONREF] [ADD DEFINITION]
- **PIMS** - Personal Information Management Services - See Personal Data Lockers
- **Participatory Action Research** - see Action Research
- **PDE - Personal Data Economy** - [ADD SECTIONREF] [ADD DEFI-

NITION]
- **Personal Data Ecosystem** - [ADD SECTIONREF] [ADD DEFINITION]
- **Personal Data Lockers** - [ADD SECTIONREF] [ADD DEFINITION]
- **Personal Data Vault** - See Personal Data Lockers
- **Personal Informatics** - see SI
- **Preparation (stage of Personal Informatics)** - see SI
- **Perspectives** - [ADD SECTIONREF] [ADD DEFINITION]
- **Point of Severance** - [ADD SECTIONREF] [ADD DEFINITION]
- **Power - Behavioural Influence** - [ADD SECTIONREF] [ADD DEFINITION]
- **Power - Interpretative Influence** - [ADD SECTIONREF] [ADD DEFINITION]
- **Power - Network Centrality** - [ADD SECTIONREF] [ADD DEFINITION]
- **Power, Authority** - [ADD SECTIONREF] [ADD DEFINITION]
- **Power, Disciplinary** - [ADD SECTIONREF] [ADD DEFINITION]
- **Power, Infrastructural** - [ADD SECTIONREF] [ADD DEFINITION]
- **Power, Interpretive** - [ADD SECTIONREF] [ADD DEFINITION]
- **Power, (power to)** - [ADD SECTIONREF] [ADD DEFINITION]
- **Power, (power over)** - [ADD SECTIONREF] [ADD DEFINITION]
- **Power, Processual** - [ADD SECTIONREF] [ADD DEFINITION]
- **Power, Rational** - [ADD SECTIONREF] [ADD DEFINITION]
- **Power, Resource Control** - [ADD SECTIONREF] [ADD DEFINITION]
- **Power, Social** - [ADD SECTIONREF]
- **Power, Socially-shaped** - [ADD SECTIONREF] [ADD DEFINITION]
- **Power, Systems/Structural** - see Infrastructural Power
- **Power, Zero Sum** - [ADD SECTIONREF] [ADD DEFINITION]
- **Power Imbalance (over Personal Data)** - [ADD SECTIONREF] [ADD DEFINITION]
- **Pragmatism** - [ADD SECTIONREF] [ADD DEFINITION]
- **Priming a Discussion** - [ADD SECTIONREF] [ADD DEFINITION]
- **QSM - Quantified Self Movement** - see SI
- **R&D - Research & Development** - [ADD SECTIONREF] [ADD DEFINITION]
- **Reflection (stage of Personal Informatics)** - see SI
- **Recursive Public** - [ADD SECTIONREF] [ADD DEFINITION]
- **Reminding** - [ADD SECTIONREF] [ADD DEFINITION]
- **SAR - Subject Access Request** - [ADD SECTIONREF] [ADD DEFINITION]
- **SI - Self Informatics** - Self Informatics [[ADD REF] ()]
- **SILVER** - [ADD SECTIONREF] [ADD DEFINITION]
- **SITRA** - [ADD SECTIONREF] [ADD DEFINITION]
- **Scraping** - [ADD SECTIONREF] [ADD DEFINITION]
- **Seams** - [ADD SECTIONREF] [ADD DEFINITION]
- **Subject Access Request** - [ADD SECTIONREF] [ADD DEFINITION]
- **Subjective Classification Principle** - [ADD SECTIONREF] [ADD

DEFINITION]
- **Subjective Importance Principle** - [ADD SECTIONREF] [ADD DEFINITION]
- **Subjective Context Principle** - [ADD SECTIONREF] [ADD DEFINITION]
- **Support Worker** - [ADD SECTIONREF] [ADD DEFINITION]
- **Supported Family** - [ADD SECTIONREF] [ADD DEFINITION]
- **Text Mining** - [ADD SECTIONREF] [ADD DEFINITION]
- **ToC - Theories of Change** - [ADD SECTIONREF] [ADD DEFINITION]
- **Things to Think With** - [ADD SECTIONREF] [ADD DEFINITION]
- **Timelines** - [ADD SECTIONREF] [ADD DEFINITION]
- **TrackerControl** - see Data Flow Auditing.
- **Troubled Families** - [ADD SECTIONREF] [ADD DEFINITION]
- **VRM - Vendor Relationship Management** - [ADD SECTIONREF] [ADD DEFINITION]
- **Web Augmentation** - [ADD SECTIONREF] [ADD DEFINITION]
- **Web Extensions** - [ADD SECTIONREF] [ADD DEFINITION]
- **Wisdom Curve** - [ADD SECTIONREF] [ADD DEFINITION]
- **world2vec** - [ADD SECTIONREF] [ADD DEFINITION]

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.

Berners-Lee, T. (2022) 'Solid: Sir tim berners-lee's vision of a vibrant web for all'. Inrupt. Available at: https://inrupt.com/solid/.

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. (2020a) 'Design research for cornmarket PDS, recommender & associated permissions: Report by alex bowyer (BBC research intern/open lab PhD)'. Available at: https://bit.ly/bbc-pds-research-bowyer.

Bowyer, A. (2020b) 'Designing personal data interfaces - a multi-disciplinary challenge'. Available at: https://bit.ly/bbc-internship-alex-bowyer (Accessed: 18 August 2022).

Bowyer, A., Pidoux, J., *et al.* (2022) *Digipower technical reports: Auditing the data economy through personal data access.* doi: 10.5281/zenodo.6554177.

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

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology.* Taylor & Francis, 3(2), pp. 77–101. doi: 10.1191/1478088706qp063oa.

British Broadcasting Corporation (1997) 'Our purpose'. Available at: https:

//www.bbc.co.uk/rd/about/our-purpose (Accessed: 18 August 2022).

Daniel Turner (2014) 'Quirkos cloud'. Available at: https://www.quirkos.com/learn-qualitative/features.html.

Dehaye, P.-O. (2019) 'Hestia.ai: About us'. Available at: https://hestia.ai/en/about/.

Goffe, L. *et al.* (2021) 'Appetite for disruption: Designing human-centred augmentations to an online food ordering platform', in *34th british HCI conference*, 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*.

Härkönen, T. *et al.* (2022) *Tracking digipower: How data can be used for influencing decision-makers and steering the world.* Sitra. Available at: https://www.sitra.fi/en/publications/tracking-digipower/.

Härkönen, T. and Vänskä, R. (2021). Sitra. Available at: https://www.sitra.fi/en/projects/digipower-investigation/#what-is-it-about.

Huberman, M. and Miles, M. B. (2002) *The qualitative researcher's companion.* Sage.

Pidoux, J. *et al.* (2022) *Digipower technical reports: Understanding influence and power in the data economy.* doi: 10.5281/zenodo.6554155.

Sitra (1967). Available at: https://www.sitra.fi/en/topics/strategy-2/#what-is-sitra (Accessed: 18 August 2022).

Sitra (2018) 'Sitra's fair data economy theme: What is it about?' Available at: https://www.sitra.fi/en/themes/fair-data-economy/#what-is-it-about (Accessed: 18 August 2022).

Turitzin, M. and Patel, J. (2010) 'Workflowy'. Available at: https://www.workflowy.com/features/.