Understanding and Improving Human Data Relations

Alex Bowyer

Contents

1 Defining a New Field: Human Data Relations

> *"What drives and drags the world are not machines, but ideas."*—Victor Hugo

1.1 Introduction to Part Two

Chapter 6 concluded the academic inquiry part of this thesis. We now know, backed up by the insights from the participants of the Case Studies, what people want from direct [RQ1] and indirect [RQ2] data relations. That is, however, not the end of the story. Bringing to bear my experience as a designer and software developer, I can advance the exploration of this problem space further, turning from theory to practice, from what is needed to what is possible. It has been my good fortune to have been involved throughout the duration of this PhD with peripheral activities that also can be seen, in the context of this thesis' findings, to relate very much to the pursuit of personal empowerment and human-centric

data use. For Part Two, therefore, I expand the original research question, going beyond the initial RQ's inquiry into what relationship people need with their personal data, and explore how those needs could be met in practice:

> **"Having understood what relationship people need with their personal data, how might these better data relations be achieved?"**

Methodologically, Part Two is set apart from the main academic enquiry. The Case Studies prioritised a participatory and investigatory approach. But, there is a need for specialist design innovation that cannot always arise from working with everyday users. The approach now, therefore, is more UCD than PD [3.2.1]. Thesis findings are now considered as material to inform myself as an adversarial designer, proposing technical and societal changes that can bring about better HDR. 7.2 describes the peripheral R&D activities I undertook, which inform the remainder of the thesis.

The wide-reaching objective of better HDR in practice has many facets: technical, design, commercial, legal, moral, social and political. These will not all be covered. Collectively, Chapter 7, 8 and 9 present an understanding of the multi-faceted realities of today's PDE landscape *sufficient to inform the design* of PDE processes and systems in pursuit of better relationships with data. This understanding is synthesised from my real world practical designs and insights as well as from the work of other innovators and activists, and is contextualised relative to existing literature and the thesis's earlier contributions.

In this chapter, I position the topic of this thesis, as a field of study in its own right, *Human Data Relations (HDR)*, formally defined in [7.3]. Additional insights into how people relate to data are identified [7.4], as well an important dichotomy of people's needs for better relations with their data [7.6]. The six wants [Chapter 6] are repurposed as four core objectives for a landscape of better HDR [7.7]. I conceptualise those who pursue these objectives as *HDR reformers* and reflect on the researcher-turned-activist stance that drives this chapter, recognising a nascent *recursive public* [7.8].

1.2 Peripheral Research & Design Settings

As established in 3.2.2 and 3.6, this second part of the thesis explores the wider action research [3.2.2] cycle that has contributed to my evolving learning about HDR, looking beyond direct academic investigation and drawing upon both self-experimentation and my embedded work in in the personal data space as both developer and researcher. Through field experience, I have understood constraints and opportunities that affect data interaction system and process design. Concurrently I have fed research learnings back into those projects, creating practical impact. Instead of conducting formal studies, I have undergone a process of acculturation to the world of practical system building and project operation in the PDE. Through design, technical prototyping and pushing boundaries of existing systems, I have developed knowledge and gained expertise

which allows me to draw conclusions with confidence about how the discipline of HDR should proceed in its future R&D to best serve individual and societal interests.

Concurrent to this PhD, I took a major role in two industrial research projects (1 & 2), and two academic research projects (3 & 4):

1. **BBC R&D's Cornmarket Project** (Sharp, 2021), which explored through user experience design, technical prototyping and participatory research, how individuals might interact with data through a Personal Data Store interface [see ARI7.1];
2. **Sitra/Hestia.ai's digipower Investigation** (Härkönen and Vänskä, 2021), a successor to Case Study Two, in which European politicians examined companies' data practices through exercising data rights and conducting technical audits [see ARI7.2];
3. **Connected Health Cities (CHC)'s SILVER Project** (Connected Health Cities, 2017), where I, along with a backend developer and a team of researchers, developed a prototype health data viewing interface for Early Help support worker [see 3.4.1]; and
4. **Digital Economy Research Centre (DERC)'s Healthy Eating Web Augmentation Project**, which explored the use of web augmentation techniques to modify the user interface of takeaway service Just Eat to insert health information, in support of healthy eating [see ARI7.3].

For additional details about these projects and my involvement in them, see the linked sections. See also ARI9 for a note about the attribution and origin of the ideas presented within this chapter.

1.3 'Human Data Relations': A Definition

Chapter 6 established six 'wants' that people have in their relationships with data: **visible**, **understandable** and **useable data**; **process transparency**, **individual oversight** and **decision-making involvement**.

The major contribution of this thesis, beyond evidencing these wants in chapters 4 to 6, is to transform these desires into a clearly defined field for future research and innovation. Repurposing concepts of 'human-technology relations' and later 'human-data relations' which have been the subject of some study in the contexts of philosophy, embodied interaction and the performing arts (Ihde, 1990; Hogan, 2012; Windeyer, 2021), I have chosen to name this field **Human Data Relations**, or **HDR** for short. I propose this field as a successor to Mortier *et al.*'s Human Data Interaction (HDI) (Mortier *et al.*, 2014).

HDR builds upon HDI but takes a broader sociotechnical stance. HDR encompasses all the ways people and organisations can and should relate to data, not just direct data interaction. Through its greater focus on relationships and ecosystems, and approaches that address today's data-centric power-imbalanced reality, it offers a more actionable research agenda for the world of the 2020s.

HDR's definition draws upon three distinct connotations or readings of its name:

Human Data Relations - A Definition
The field of HDR encompasses all the ways in which humans and human organisations relate to, and with, data, specifically: 1. *Human-Data Relations*: users' direct interaction with data to understand and use it, similar to HDI, in service of the direct data wants [6.1] of visible, understandable and useable data. 2. *Human "Data Relations"*: individuals' relationships with organisations that hold data about them, in service of the indirect data wants [6.2] of transparency, individual oversight and involvement. 3. *Human/Data Relations*: how organisations manage their customers with respect to personal data. Similar to *public relations* or *customer relations*, organisations choose how present their data practices (so as to build trust), and whether they will involve users with data, and provide support to understand data to their users. Organisations can empower individuals and build more effective customer relationships through HDR [4.4.1; 5.5.2; 6.1.2].

Having scoped HDR, we see that 'better' HDR can be achieved by working to improve upon the identified six wants for data relations outlined in Chapter 6. However, as this section will explain, HDR is motivated in two distinct ways, to which those wants apply differently. As background understanding, it is first necessary to examine more closely what role data plays in people's lives.

1.4 The Role of Personal Data

Today, almost anything can be encoded as data. Many previously analogue objects and activities now have digital equivalents, so the concept of data has become broad and hard to pin down. Underlying HDR is a need to recognise what roles data can play in people's lives—what it **is** to people. I have so far identified eight distinct lenses to explain how people might relate to data—including as property, as memory and as creative work. These are modelled in Table ARI5.2.

People may think of their personal data through any or all of these *lenses* [Karger *et al.* (2005); 2.2.2] at any given time. Any data interaction process or interface design should take these into account. Different informational representations might be needed at different times (Lindley *et al.*, 2018), bringing different aspects of the data to the forefront. Looking across these lenses, I identify four specific *roles* that data can serve:

1. Data has a role as an **artefact of value** to your life;
2. Data has a role in **informing** you about yourself, the world, and the prior or recent actions of others that may affect you;
3. Data has a role as **a useable10 material with which to effect change** in your life;

4. Data has a role as **a means to monitor changes** in data holders' behaviours, in digital influences upon you, or within your life.

1.5 Human Data Interaction or Human Information Interaction?

To unpack HDR further, we must differentiate between humans relating to data, and humans relating to information. HDI concerns the way people interact with data. Mortier *et al.* (Mortier *et al.*, 2013, 2014) defined the field of HDI without making the important distinction between data (the digital artefact stored on computer) and information (the facts or assertions available from that data). This is an important distinction. *Human Information Interaction (HII)* originated in library sciences to consider how humans relate to information without regard to the technologies involved (Marchionini, 2008). Jones *et al.* called for a new sub-field of HII in an HCI context[1], highlighting the need to focus on information interaction:

> "[HCI can] *unduly focus attention on the computer when, for most people, the computer is a means to an end—the effective use of information.*"—Jones *et al.* (2006)

DIKW theory [2.1.1] highlights that interpretation of data to obtain information is a discrete activity. This was borne out in the findings of Case Study Two, where it became clear that participants have distinct needs from data, and from information [5.4.3]. Access to data *and* information is critical to both understanding and useability [6.1.2; 6.1.3].

Drawing on DIKW theory, allows the identification of three distinct artefacts people can have relations with:

1. **data** - the stored digital artefacts held by organisations for algorithmic processing and human reference, copies of some of which can be obtained using data rights;
2. **information about individuals** (a.k.a. *life information*) - facts and assertions about the individual and their life, obtained through human or algorithmic interpretation of stored data or analytical inference; and
3. **information about data_** (a.k.a. *metadata* [Table 5.2; 5.3.1] or *ecosystem information*) - stored facts about data, such as storage location, access history, means of collection, contextual meaning, or sharing records.

1.6 The Two Distinct Motivations for Better Data Relations

Considering these two types of information in the context of the six wants [Chapter 6] reveals two very different reasons why people might want better data

[1]The HCI panelists involved (excepting Fidel) were seemingly unaware of the existing HII field in library sciences, as they positioned the publication as a call for a 'new field'.

relations:

(i) to acquire *information about your data*, so that you might exert control over where the data is held and how it is used, in order to be treated fairly and make informed choices about personal data. This is **Personal Data Ecosystem Control (PDEC)**.

(ii) to acquire *information about yourself*, so that you might gain insights into your own behaviour, and gain personal benefits from those insights or make changes in your life. This is **Life Information Utilisation (LIU)**.

Figure 7.1 shows processes individuals might go through in pursuit of these motives. PDEC is a process of holding organisations to account and managing *what happens* to personal data, often regardless of what it means. LIU is more concerned with *what the data means* and its inherent personal value, regardless of where it is stored and how it is used[2]. This novel motivational model was first proposed in (Bowyer, 2021).

1.6.1 Life Information Utilisation (LIU)

Life Information Utilisation is a superset of *Self Informatics (SI)* [2.2.3], including all purposes relating to self-monitoring and self-improvement through personal data, but also other uses including creative expression, evidence gathering, nostalgia, keeping, and sharing. Many such desires were expressed in Case Study Two [Table 5.4], and also hinted at in the Early Help context [4.4.1]. While the existence of digitally-encoded information clearly unlocks new possibilities, LIU has existed in some form throughout human civilisation, as seen through analogue processes such as storytelling, journaling and scrapbooking.

The most relevant of the six wants to LIU are *data understandability* [6.1.2] and *data useability10* [6.1.3], which relate closely to the HDI concepts of *legibility* and *agency* respectively.

1.6.2 Personal Data Ecosystem Control (PDEC)

Unlike LIU, *Personal Data Ecosystem Control* is a *new* individual need, arising as a result of the emergence of the data-centric world [2.1; 2.2.4]. Only when organisations began to collect and store facts about people as a substitute for direct communication and involvement did it become necessary. The more data is collected about individuals, and the more parties doing so, the greater the

[2]There is some overlap. Organisations hold data to enable interpretation (usually algorithmic) to inform decision making. In this way, organisations are doing LIU for *their* benefit. This grey area is situated as part of PDEC, because from the individual's perspective, how organisations understand you through information informs decisions that affect your life. As such, it is more likely to enable you to exert control over use of your data than to pursue personal LIU goals.

[4]The illustrated processes incorporate existing data access processes such as GDPR, where the only access is through provision of a copy of one's data. This is *not* ideal, as it creates divergent versions and will quickly become out-of-sync, however for the sake of simplicity that inefficiency is ignored [see 5.5.1 for improvements to copy-based access].

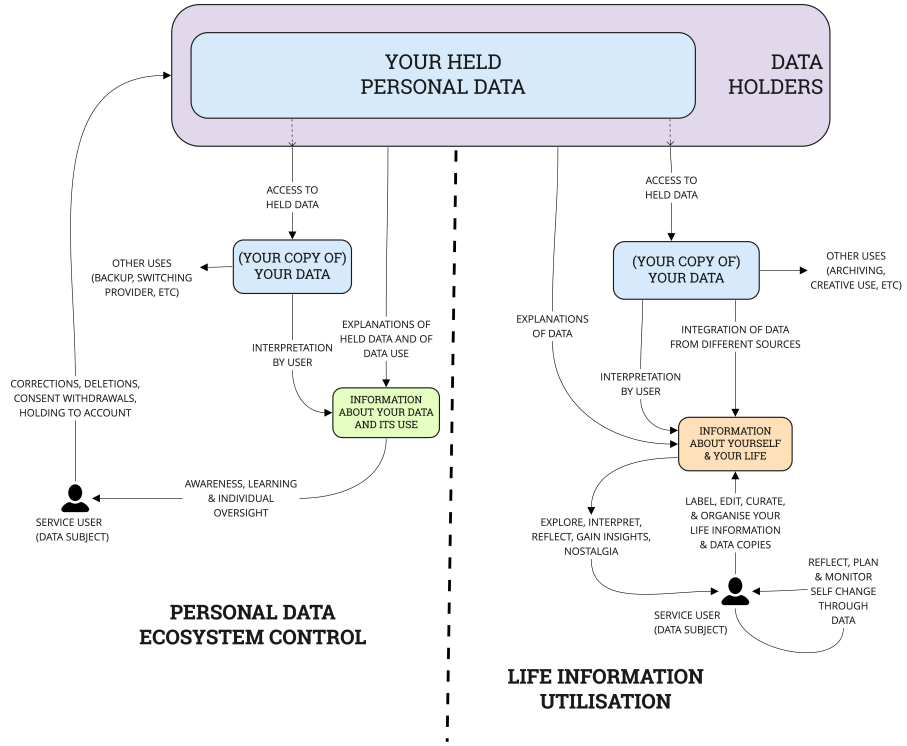

Figure 1: Figure 7.1: The Two Motivations for HDR: Controlling Your Personal Data Ecosystem and Utilising Your Information About Your Life(with 'idealised'[4] processes illustrated)

need for individuals to understand these acts so that they might influence them (or risk their lives being affected in unexpected or unfair ways). PDEC is a direct response to the power imbalance between data holders and individuals [Hoffman (2014); 2.1.2].

Several of the six wants are important to PDEC: visible data and transparent processes (referred to collectively as *data ecosystem transparency*), and individual oversight and involvement (referred to collectively as *data ecosystem negotiability*, drawing on the HDI concept of *negotiability*). These grouped terms are used below.

1.7 Four Objectives for Human Data Relations

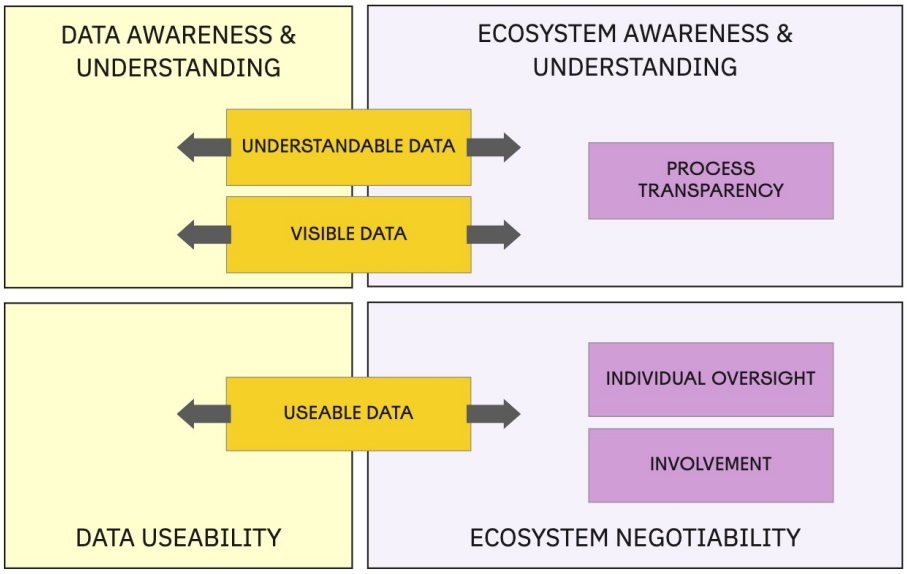

Figure 2: Figure 7.2: Mapping the Six Wants into Objectives for the HDR Opportunity Landscape

To offer future value to future researchers, activists and innovators, this chapter contributes a map of the HDR opportunity landscape. This map is expressed in abstract here, and explored in more depth in Chapters 8 and 9. First, the six wants [Chapter 6] are transformed into four simple *landscape objectives* which shape our ultimate goals for effective HDR in this landscape of opportunity:

1. Data Awareness & Understanding;
2. Data Useability10;
3. Data Ecosystem Awareness & Understanding[5] and

[5]To avoid overly cumbersome wording, subsequent sections will drop the 'Data' prefix from 'Data Ecosystem Awareness & Understanding' and 'Data Ecosystem Negotiability'.

4. Data Ecosystem Negotiability15.

As Figure 7.2 shows, the need for data to be understandable, visible and useable applies to all types of data, whether that data is interpretable as *life information* (information within the data, that says something about the individual) or *ecosystem information* (information *about* the data, where it is held and how it is used). These two types of information will collectively be referred to as **human information**. These terms are used in subsequent sections.

1.8 Better Human Data Relations as a Recursive Public

Let us revisit the stance from which we approach this change. This PhD has been grounded in participatory action research and experience-centred design [3.2]. Using a *Digital Civics* (Vlachokyriakos *et al.*, 2016) approach to understand people's unmet needs, we can model how the world should change. Such research is political [3.2.1], seeking to correct an imbalance in the world through *adversarial design* (DiSalvo, 2012). This chapter steps forward in the role of activist researcher, not only understanding what needs to change, but exploring how individuals and groups can actually change their world.

In this, we can consider ourselves (those who pursue better HDR, or *HDR reformers* as a shorthand) as a nascent *recursive public* ('Recursive Public (Discussion Page)', no date). This term originates in the free software movement to describe:

> *'a collective, independent of other forms of constituted power, capable of speaking to existing forms of power through the production of actually existing alternatives'*-Kelty (2008)

Being a recursive public means using various means at our disposal to seek to modify the systems and practices we live within in pursuit of our goals. These methods might include participatory research, experience-centred design, software prototyping, rights exertion and campaigning.

This idea of reconfiguring society in this way has been conceived as *civic hacking* (Crabtree, 2007; Levitas, 2013; Tauberer, 2014). The collective around HDR reform does not yet exist as a named and identifiable *public* (Le Dantec, 2016) but its members congregate around interconnected and overlapping movements such as:

- the *MyData* community [MyData (2017); 2.3.4];
- personal data lockers (CitizenMe, 2021; Sharp, 2021; 'Digi.me', no date);
- digital rights ('Open rights group: Who we are', no date);
- gig economy worker rights (Kirven, 2018; 'Worker info exchange', 2022);
- privacy by design (Cavoukian, 2010);
- privacy activism (Davies, 1990; 'Bits of freedom: Our focus', 2000);
- data justice (Taylor, 2017; Crivellaro *et al.*, 2019);
- critical algorithm studies (Gillespie and Seaver, 2016);
- adversarial interoperability (Doctorow, 2019);

- 'makers' (Altsitsiadis, 2021);
- humane technology (Harris, 2013); and
- explainable AI ('Explainable AI: Making machines understandable for humans', no date).

1.9 Summation: HDR—A Landscape to Explore

The commonality to so many groups 7.8 suggests HDR reform is an emergent cultural phenomenon, whether or not a single identifiable public coalesces. Time will tell whether *Human Data Relations* as laid out in this thesis is sufficient to give form to that phenomenon. At the least, HDR offers as descriptive umbrella term. The breadth of research, innovation and activism validates the need *and* the desire for such a recursive public around HDR reform to exist. In fact, it already does, whether named or not. Therefore, this chapter takes an unashamedly critical view of the status quo, favouring disruptive societal change that would further HDR reform and provide actionable approaches for members of this public. This chapter asks,

> *"How can we change the world into the one we want?"*

[TODO ADD TEXT HERE - now we will dive into this field, start out on the road]

Bibliography

Altsitsiadis, E. (2021) 'The maker movement – the quiet, game-changing revolution near you'. available at: http://www.bos-cbscsr.dk/2021/01/05/the-maker-movement/ (accessed: 21 August 2022).

'Bits of freedom: Our focus' (2000). available at: https://www.bitsoffreedom.nl/english/.

Bowyer, A. (2021) 'Human-Data Interaction has two purposes: Personal Data Control and Life Information Exploration'. available at: https://eprints.ncl.ac.uk/273832#.

Cavoukian, A. (2010) 'Privacy by design: the definitive workshop. A foreword by Ann Cavoukian, Ph.D', *Identity in the Information Society*, 3(2), pp. 247–251. doi: 10.1007/s12394-010-0062-y.

CitizenMe (2021) 'Become a Citizen and unlock the value of your data'. available at: https://www.citizenme.com/for-citizens/ (accessed: 23 August 2021).

Connected Health Cities (2017) 'SILVER Project: Smart Interventions for Local Residents'. available at: https://web.archive.org/web/20210308040602/https://www.connectedhealthcities.org/research-projects/troubled-families/ (accessed: 8 March 2021).

Crabtree, J. (2007) 'Civic hacking: A new agenda for e-democracy', *openDemocracy*. available at: https://www.opendemocracy.net/en/civic_hacking_a_new_agenda_for_e_democracy/.

Crivellaro, C. *et al.* (2019) 'Not-equal: Democratizing research in digital innovation for social justice', *Interactions*, 26(2), pp. 70–73. doi: 10.1145/3301655.
Davies, S. (1990) 'Privacy international: About us'. available at: https://privacyinternational.org/about.
'Digi.me' (no date). available at: https://digi.me/ (accessed: 23 August 2021).
DiSalvo, C. (2012) *Adversarial Design*. MIT Press (Design thinking, design theory). doi: 10.7551/mitpress/8732.003.0007.
Doctorow, C. (2019) 'Adversarial interoperability', *Electronic Frontier Foundation*. available at: https://www.eff.org/deeplinks/2019/10/adversarial-interoperability.
'Explainable AI: Making machines understandable for humans' (no date). available at: https://explainableai.com/ (accessed: 16 June 2022).
Gillespie, T. and Seaver, N. (2016) 'Critical Algorithm Studies - A Reading List'. available at: https://socialmediacollective.org/reading-lists/critical-algorithm-studies/.
Härkönen, T. and Vänskä, R. (2021). Sitra. available at: https://www.sitra.fi/en/projects/digipower-investigation/#what-is-it-about.
Harris, T. (2013) 'A Call to Minimize Distraction Respect Users' Attention'. available at: http://www.minimizedistraction.com/.
Hoffman, W. (2014) *Rethinking Personal Data : A New Lens for Strengthening Trust*. May. World Economic Forum, p. 35. available at: http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf.
Hogan, T. (2012) *Toward a phenomenology of human-data relations*. available at: http://www.manovich.net/DOCS/data_art.doc,.
Ihde, D. (1990) *Technology and the lifeworld: From garden to earth*. Indiana University Press.
Jones, W. *et al.* (2006) '"It's about the information stupid!": Why we need a separate field of human-information interaction', *Conference on Human Factors in Computing Systems - Proceedings*, pp. 65–68. doi: 10.1145/1125451.1125469.
Karger, D. R. *et al.* (2005) 'Haystack: A customizable general-purpose information management tool for end users of semistructured data', in *2nd biennial conference on innovative data systems research, CIDR 2005*, pp. 13–27. available at: https://s3.amazonaws.com/academia.edu.documents/46870765/haystack.pdf.
Kelty, C. M. (2008) *Geeks and Recursive Publics*. Duke University Press, pp. 27–63.
Kirven, A. (2018) 'Whose gig is it anyway: Technological change, workplace control and supervision, and workers' rights in the gig economy', *U. Colo. L. Rev.* HeinOnline, 89, p. 249.
Le Dantec, C. A. (2016) *Designing publics*. MIT Press.
Levitas, J. (2013) 'Defining civic hacking', *Medium*. available at: https://medium.com/civic-innovation/defining-civic-hacking-16844fc161cd.
Lindley, S. E. *et al.* (2018) 'Exploring new metaphors for a networked world through the file biography', *Conference on Human Factors in Computing Systems - Proceedings*, 2018-April, pp. 1–12. doi: 10.1145/3173574.3173692.
Marchionini, G. (2008) 'Human-information interaction research and development', *Library and Information Science Research*, 30(3), pp. 165–174. doi:

10.1016/j.lisr.2008.07.001.
Mortier, R. *et al.* (2013) 'Challenges & opportunities in human-data interaction', *University of Cambridge, Computer Laboratory*. Citeseer. doi: 10.5210/fm.v17i5.4013.
Mortier, R. *et al.* (2014) 'Human-data interaction: The human face of the data-driven society', *Available at SSRN 2508051*. doi: 10.2139/ssrn.2508051.
MyData (2017) 'Declaration - MyData.org'. available at: https://web.archive.org/web/20210325143142/https://www.mydata.org/declaration/ (accessed: 25 March 2021).
'Open rights group: Who we are' (no date). available at: https://www.openrightsgroup.org/who-we-are/ (accessed: 16 June 2022).
'Recursive Public (Discussion Page)' (no date). available at: https://wiki.p2pfoundation.net/Recursive_Public (accessed: 16 June 2022).
Sharp, E. (2021) 'Personal data stores: Building and trialling trusted data services - BBC r&d', *BBC R&D Blog*. available at: https://www.bbc.co.uk/rd/blog/2021-09-personal-data-store-research.
Tauberer, J. (2014) 'Civic hacking', *Open Government Data: The Book*. available at: https://opengovdata.io/2014/civic-hacking/.
Taylor, L. (2017) 'What is data justice? The case for connecting digital rights and freedoms globally', *Big Data and Society*, 4(2). doi: 10.1177/2053951717736335.
Vlachokyriakos, V. *et al.* (2016) 'Digital civics: Citizen empowerment with and through technology', *Conference on Human Factors in Computing Systems - Proceedings*, 07-12-May-, pp. 1096–1099. doi: 10.1145/2851581.2886436.
Windeyer, R. C. (2021) *Black box exposures: Enriching public engagement with human-data relations through intermedial performance strategies*.
'Worker info exchange' (2022). available at: https://www.workerinfoexchange.org/.