

Understanding and Improving Human Data Relations

Alex Bowyer

Contents

1	Mapping the Human Data Relations Landscape	1
1.1	Obstacles to the HDR Objective of Data Awareness & Understanding	2
1.1.1	Invisible, Inaccessible or Unrelatable Data	2
1.1.2	The Personal Data Diaspora	4
1.2	Obstacles to the HDR Objective of Data Useability	10
1.2.1	Immobile, Inaccessible or Unmalleable Data	6
1.3	Obstacles to the HDR Objective of Ecosystem Awareness & Understanding	10
1.3.1	Complex and Invisible Personal Data Ecosystems	10
1.3.2	A Lack of Metadata	12
1.4	Obstacles to the HDR Objective of Ecosystem Negotiability	15
1.4.1	Hegemony through Data Holding	15
1.4.2	The Active Diminishing of User Agency	19
1.4.3	The Intractable Data Self	21
1.5	Obstacles to the HDR Objective of Effective, Commercially-Viable and Desirable Systems	21
1.5.1	A Lack of Individual Demand	21
1.5.2	Closed, Insular and Introspective Practices	24
1.5.3	A Lack of Organisational Investment in HDR	25
1.5.4	A Lack of Interoperability	26
1.5.5	Insufficient Machine Understanding of Human Information	26
1.6	Summation of Chapter 8: From Obstacles to Opportunities . . .	28
	Bibliography	29

1 Mapping the Human Data Relations Landscape

“There are certain things you do not in good conscience do to humans. To data, you can do whatever you like.”—Nikhil Sonnad, data

journalist and technology commentator

In this chapter, we begin to engage with the expanded research question laid out in 7.1. Considering how better HDR might be achieved in practice generates further questions. Like SI's *barriers cascade* (Li, Forlizzi and Dey, 2010), what barriers exist that inhibit the building or adoption of human-centric technologies? What opportunities might overcome these barriers? How can we catalyse progress toward *MyData*'s human-centric agenda (*MyData*, 2017)? What challenges are faced when attempting to build human-centric technologies for today's world? Building on an understanding of human experience of the data-centric world, can we more provide an outlook for PDE design & development and define a *research agenda* for the next step of tackling the PDE challenge?

Focusing on the four objectives [7.7], and informed by the peripheral work [7.2], I can identify specific *obstacles*. Analogous to Li's *barriers cascade* [2.2.3; Li, Forlizzi and Dey (2010)], these are the challenges that individuals or system designers must be empowered to overcome. These obstacles are documented in the following sections, accompanied by *insights* that might help adversarial designers or strategists to tackle them. Figure 8.1 shows an HDR-specific barriers cascade: a route of overcoming obstacles through which individuals might be empowered and by which organisations might become more HDR-friendly. The concepts on this diagram will be refined and explained across this chapter and the next. The last of these (corresponding to the 'solution space' box) covers some of more pervasive obstacles that apply to all four HDR objectives.

1.1 Obstacles to the HDR Objective of Data Awareness & Understanding

1.1.1 Invisible, Inaccessible or Unrelatable Data

In pursuit of visible, understandable data [6.1.1; 6.1.2], the first obstacle encountered is that most personal data is **invisible**, **inaccessible** or **unrelatable**. It is trapped in service providers' databases, or on different devices or hard drives, or by software limitations and proprietary file formats (Abiteboul, André and Kaplan, 2015; Bowyer, 2018). My research participants spoke of 'not knowing' what data exists and of being 'in the dark'. Case Study Two showed that even where data is accessible, it is not **relatable** (*legible* (*Mortier et al.*, 2014); 2.3.2). The objective here, addressed in Insight 1, is to ensure that people not only have awareness of their data, but can understand ('make sense' [*Gurstein* (2011); 2.1.4]) of what it means.

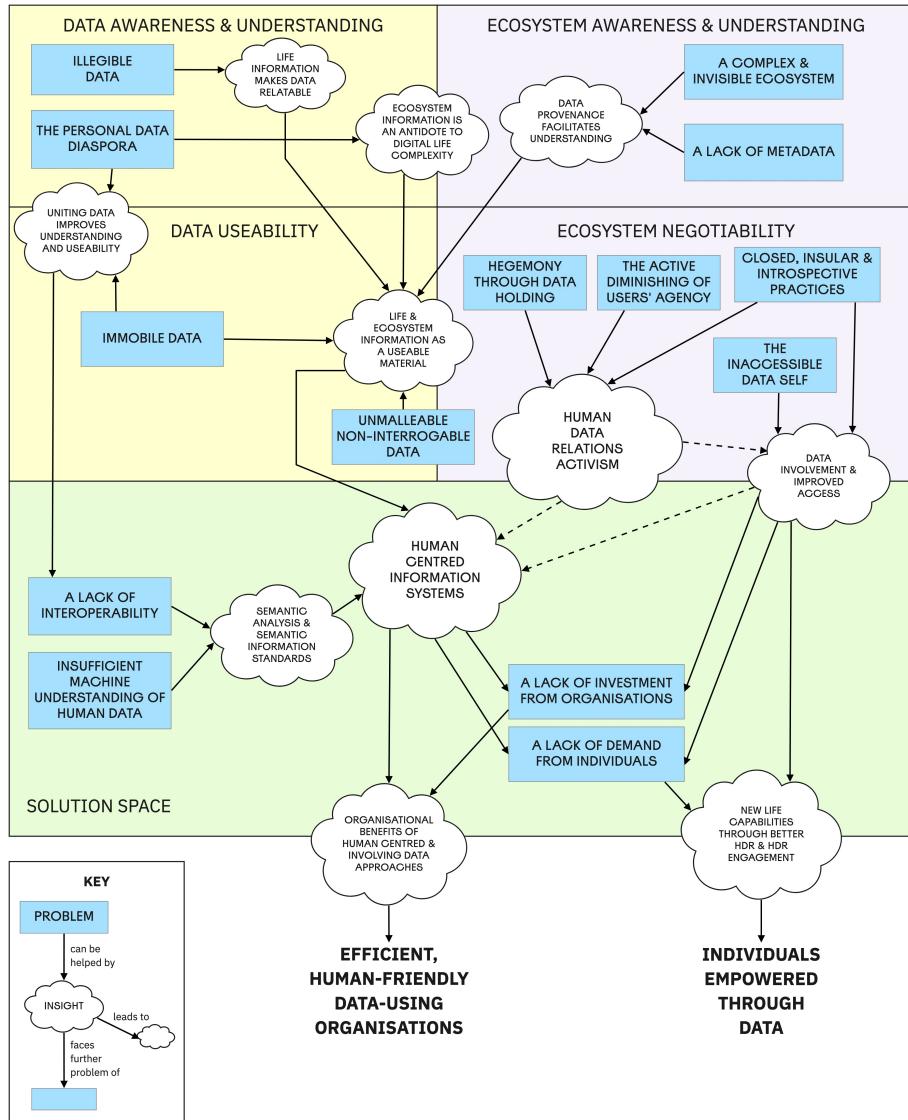


Figure 1: Figure 8.1: Obstacles and Resulting Insights in the HDR Opportunity Landscape

INSIGHT 1: Life Information makes Data Relatable

In the pilot study and Case Study One, data cards' were used to represent civic data [Figure 3.6]. In Case Study Two [Figure 3.7] and in Hestia.ai's digipower investigation [ARI7.2], a categorisation of provider-held data was displayed. In my BBC research report (Bowyer, 2020), the use of **relatable examples** was identified as an important way to help people understand what a piece of data represents. Recalling that to make data meaningful, we must be able to interpret it as information [2.1.1], this can be refined further:

To make data meaningful, it needs to be expressed as information about your life.

(continues...)

Spreadsheets and 'big data' sound dry and (to many) dauntingly technical, but once those same datapoints are expressed as 'facts about your life', the hurdle of relatability is overcome [4.2.1]. The effectiveness of applying this principle is evident in successful online services like Netflix, Spotify and Strava, and in social media platforms like Facebook: these interfaces show understandable everyday concepts like Friends, Events, Movies, Playlists, not files, records, folders or database rows. They have successfully '*pushed the technology into the background*', in line with Weiser's vision (Weiser, 1991), Rogers' calm computing, and the quote that opens this chapter. While exploring this idea of representing **life concepts** further at BBC R&D, I produced Figure 8.2, which shows a near-exhaustive overview of the many different informational concepts in an individual's life that providers might hold as data:

(continues...)

This diagram shows how most common personal data types handled today can be mapped to more relatable life information concepts. These life concepts (exemplified where possible) make data meaningful to individuals, and can help people find value in their data [5.4.3].

1.1.2 The Personal Data Diaspora

Another important obstacle to consider here is what I call the **Personal Data Diaspora**¹. As illustrated by Imogen Heap's quote opening Chapter 1, an individual's personal data is typically very widely dispersed, and there is no central, holistic view of one's data. For example, if I consider just my movement tracking data, I have over time accumulated activity logs from walking, running,

¹The word '*diaspora*' is typically used with reference to populations, but is an apt term, derived from the Greek 'diaspeirein' meaning 'scattered about' or 'dispersed'.

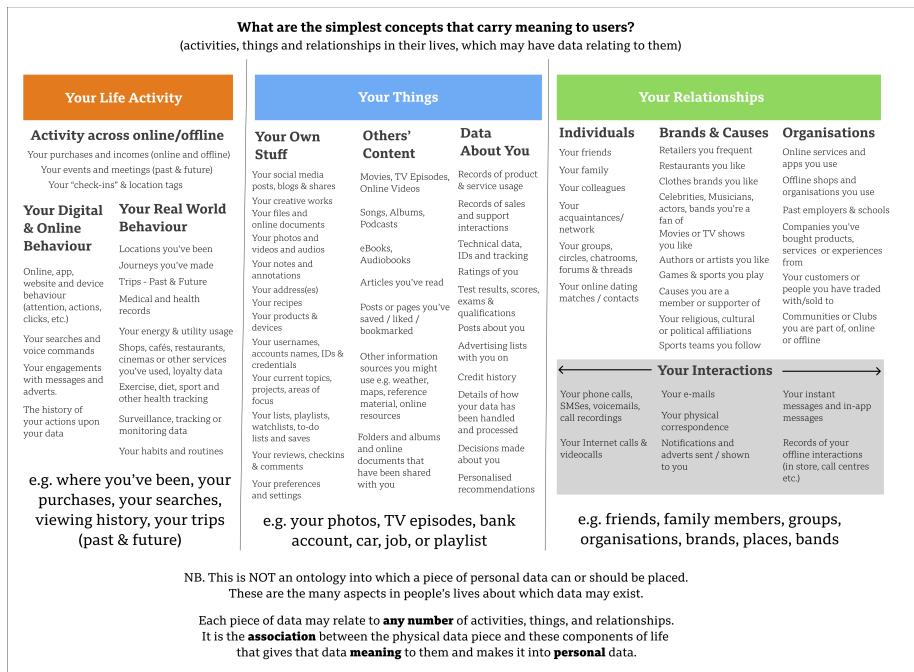


Figure 2: Figure 8.2: Life Concept Modelling

cycling, and driving which are stored by Nike+, MyFitnessPal, Strava, Google Fit, Fitbit, Apple Health and Google Maps, not to mention the records remaining on my different smart watches, smartphones and hard drives. This is the SI problem of *Integration* (Li, Forlizzi and Dey, 2010) [2.2.3]. EAs well as the challenge of managing one's data ecosystem [2.2.4], this makes impossible to view physical activity history in one place, to spot patterns over time or make comparisons. To overcome this, we need interfaces that recognise the scattered reality of each individual's personal data, and begin to make that ecosystem visible and understandable [see 8.3 and 8.4 below].

Data awareness and understanding is a problem of representation. Invisible data should be visibly represented. All data should be represented as contextually relatable life information.

1.2 Obstacles to the HDR Objective of Data Useability10

1.2.1 Immobile, Inaccessible or Unmalleable Data

To improve the useability of data, let us consider what properties of data make it hard to use. Most personal data is immobile, inaccessible, unmalleable and not interrogable.

It is **immobile** in that is very difficult to move a dataset out of its environment. Most data exists in organisations' internal databases, where it is tightly coupled to technology stacks, interfaces and business processes. Separating one's data from the service that holds it is difficult and often impossible. It is **inaccessible** to individuals (in the sense of *effective access* (Gurstein, 2011)). Data access requests such as GDPR are typically satisfied by creating a copy of the data, creating problems of delay, divergence and understanding. Even then, returned data is incomplete [5.4.2]. Its accessibility is also hindered by the technical nature of data. Data is often be stored in complex proprietary structures which are designed for the algorithmic efficiency of the specific business operations rather than for general-purpose re-use.

People need to be able to ask questions of their data [Table 5.4; 4.3.2]. But data is **not interrogable**. It must stand for itself, yet there is no obvious way to ask a question about the meaning of the data or its ability to answer a particular question. To ask questions of data requires either the co-operation of the data holder or advanced technical skills in data querying and analysis. Data needs to be **malleable**—capable of being broken down, looked at from different perspectives, and reconstituted in different ways. This goes beyond visually representing the data, and implies an ability to interact with the data to produce new interpretations and insights to investigate specific questions.

To overcome these obstacles, data must be freed from its current constraints and moved into environments where it can be freely examined and reconstituted without restriction. This leads to Insight 2:

INSIGHT 2: Data Needs to be United and Unified

It is clear that better HDR involves recognising this splintered reality (Lemley, 2021) and moving beyond it. To make data useable for individuals, the diaspora must be united. This means that data from different sources must first be **united**—brought together—and then **unified**, which means making it into a collection of data about the individual and their life, rather than scattered slices of company data that may have secondary value to the individual. This is a multi-faceted sociotechnical problem of access, interpretation and *integration* [Li, Forlizzi and Dey (2010); [2.2.3]]. Negotiability remains important; we can only unite data that we can access, and only data holders can fully explain it [see 8.3 and 8.4]. Setting that aspect aside, the pragmatic way forward begins with creating a space where data can be held, combined, controlled and **owned** by the individual - '*a place for your personal data*' (Jones, 2011, p. [2.2.4]). This can form the seed of their new human-centric personal data ecosystem. This follows Bergman's *subjective classification principle*:

'All related items should be classified together regardless of technological format'—Bergman, Beyth-Marom and Nachmias (2003)

(continues...)

We could add: '*regardless of where they are held*'. This vision is embodied in the **Personal Data Stores** (PDS) concept [2.3.4]. The BBC R&D Cornmarket project [ARI7.1] examines how to build PDSs. In Chapter 9.2 I explore possible design approaches. At this stage, only the *concept* is important. Once data is united and unified, PDSs enable the creation of new views of data that were not previously possible, because code can execute across data that was previously dispersed. For example, today each separate TV app, device or streaming service maintains separate records of what you have watched. Once unified in a PDS, it would be possible to present you with a unified view of all the past content you had viewed, across all channels, as this mock-up I made during my BBC internship shows:

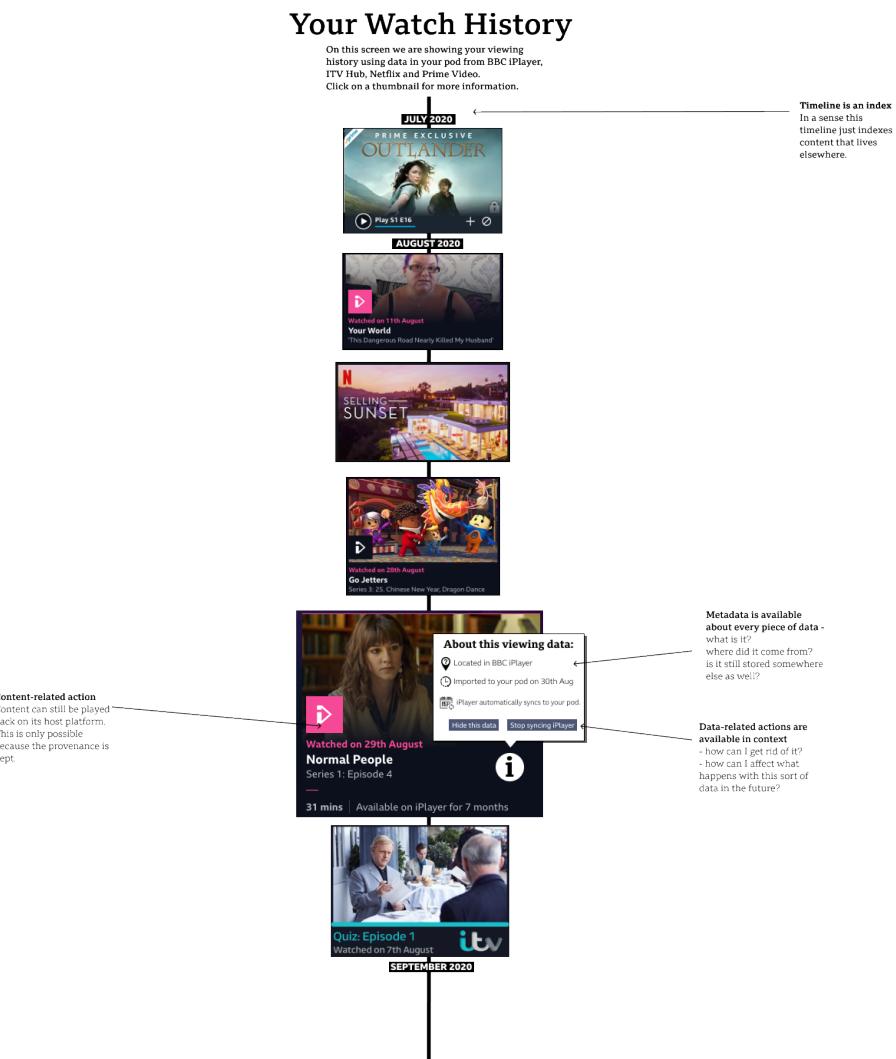


Figure 3: Figure 8.3: Mock-up of a Unified TV Viewing History Interface

INSIGHT 3: Data Must Be Transformed into a Versatile Material

In Case Study Two (see Table 5.4 and the additional detail in the Supplemental Materials of Bowyer, Holt, *et al.*, 2022), participants expressed diverse goals for personal data, including reflection, pattern-finding, goal-tracking, and creative use. In the PIM space [2.2.2] relevant innovations include associative exploration, spatial arrangement, and embodied interaction for different contexts) Drawing on all of these, allows me to infer that unified data must be transformed into a **versatile material**. Individuals need to be able to use data—represented as facts or assertions about one’s life by performing manipulations such as:

- creating,
 - deleting,
 - moving,
 - grouping,
 - annotating,
 - copying,
 - sharing,
 - modifying,
 - labelling,
 - organising, and
 - separating.
-

(continues...)

Data as material will be new to most except data scientists. This is novel not just for end users but for designers too. Eva Deckers, in her work on *data-enabled design*, an approach to design which also calls for data to become a material, notes (and we could expand this to laypeople too):

“Designers are in general not trained and prepared to work with data. They’re not equipped with the right tools. Data manipulation is not part of the schools’ curriculum and designers are rarely interested in understanding data.”—(Deckers, 2018).

(continues...)

Her work with colleagues on the ‘connected baby bottle’ illustrates how treating data as a design material creates a novel iterative user-centred development of new capabilities (Bogers *et al.*, 2016). In HDR terms, I theorise that what this material should be is *human information* - life information and ecosystem information [7.2]. Data usability therefore calls for the creation of systems that enable **human information to be treated as a material**.

As Insight 3 show, data will only become useable once we change its nature.

Since the 1970s, drawing on the then-common metaphor of a filing cabinet, computers have considered *files* as the basic material that users will interact with. Where we do interact with data as information instead of files, that information is typically presented in limited contexts within certain products or apps [Insight 1]. To move up the DIKW pyramid [2.1], we need smarter computer systems, that move beyond files (Bowyer, 2011). We need systems whose basic material is not files, but pieces of human information.

We need a human information operating system.

1.3 Obstacles to the HDR Objective of Ecosystem Awareness & Understanding

1.3.1 Complex and Invisible Personal Data Ecosystems

As established [2.2.5; 2.3; 6.2; 7.2], HDR cannot be made effective without a sea change in the way that individuals are able to interact with the complex ecosystem of personal data they each inhabit. Our personal data ecosystems are incredibly complex and largely invisible. For example, it is easy to allow a handful of messaging and social media apps to access your contact list. Before you know it, you have created a complex and unmanageable network of connections that silently sync and propagate your addresses and phone numbers across the Internet. And there are deeper layers which are not even slightly visible to users: networks of data brokers, advertisers and digital cookie companies exchange user identifiers, activity data and personal information about you while you browse or use apps (Pidoux *et al.*, 2022). The ability to build up a meaningful picture of your personal data ecosystem is completely absent [4.3.4] or severely limited. People remain ‘in the dark’, leading to fear (Bowyer *et al.*, 2018), overload [2.2.4] and resignation [5.4.4]. Managing one’s personal data ecosystem is an **overwhelming, unmanageable task** that even personal data experts are not fully able to get a handle on. We do not feel ‘in control’ [Teevan (2001); 2.2.2]. The ability to provide a user with ecosystem transparency is hindered by the complexity and multiplicity of data relationships they have been encouraged to set up. People lack tools to provide a meaningful, or indeed any, view of those relationships. In both Case Study contexts, we saw that no one individual or organisation has the ability to see the whole of a user’s data ecosystem [4.3.4; Cornford, Baines and Wilson (2013)]. There is little commercial motive to try and solve this problem, as each provider focuses on their own apps, websites and services. Making one’s ecosystem visible, transparent and understandable is therefore an essential objective for better HDR, as Insight 4 shows.

INSIGHT 4: Ecosystem Information is an Antidote to Digital Life Complexity

Acquiring ecosystem information and understanding is a key motivator for many people—encompassing 74% of participant goals in Case Study Two [Table 5.4]—and is essential for better HDR. This suggests two distinct goals for system builders: **ecosystem detection** and **ecosystem information display** as ingredients to help overcome the obstacle. As a representative example let us examine a recent app called SubsCrab [Figure 8.4]:

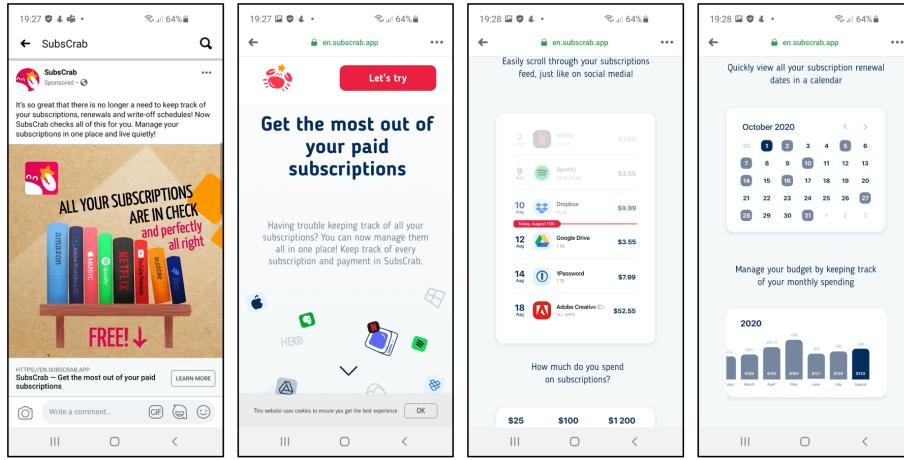


Figure 4: Figure 8.4: SubsCrab: An Example Application for Ecosystem Detection and Visualisation

(continues...)

This app connects to the user's e-mail account, and searches it and monitors it for e-mails from service providers such as Netflix, Spotify, Dropbox, or Google with which the user has monthly or annual subscriptions. In doing so, it is detecting part of the user's ecosystem. It is identifying which companies they have a payment relationship with. It parses found e-mails to identify billing dates and payment amounts. It then provides additional representations of that ecosystem information to the user, so that they might get on top of their subscriptions, see what they need to pay (or cancel), and feel more 'in control' [Teevan (2001); 2.2.2] of this aspect of their digital life. From this example, it is easy to imagine other types of ecosystem detectors, which could detect relationships with free services and websites, identify account numbers and e-mail addresses, password resets, address book syncs, OAuth logins, family identities and more. Alistair Croll and I explored possibilities for *inbox scanning* in 2009 (Croll and Bowyer, 2009), and while there has been some innovation in this space, it has largely been for commercial reasons (Braun, 2018). New ecosystem detectors could power new interfaces, contributing to the simplification of the user's digital life. This would give people more visibility and control over their previously unmanageable data ecosystem.

A secondary consideration in achieving the required 'sea change' in approaches HDR, is that current PDS and SI approaches are very life-information-centric. It is implicitly assumed that the only way to unite data is to collect it. The difficulty in such an approach is that you can only collect that which you can extract. To address this, I draw inspiration from a computer programming concept known as *pass by reference* (as opposed to *pass by value*) (Ananya, 2020) where data is 'pointed to' rather than moved. Productivity guru David Allen recommends the use of '*placeholders*' (Allen, 2015) to keep track of tasks you cannot otherwise bring into your planning. To build a complete map of a user's ecosystem we must be able to keep track of accounts and data that are remote, much like a search engine points to information on different pages around the web. We can create **proxy representations** of service-provider-held or otherwise immobile data (e.g. data which is offline or restricted). These representations become part of the manipulable material in the user interface, and could be augmented with links to visit the remote source.

1.3.2 A Lack of Metadata

As we start to consider *what the data is about*, new possibilities are unlocked. A PDS-type system could built that is not only a repository of personal data, but (using proxy representations), a collection of ecosystem information and *contextually-situated* life information too. This could include information about relationships with data holders or other entities. Builders of such a system would face a further challenge—a **lack of metadata** [2.2.2]. Typically, most data on our hard drives lacks context about its origin, and how it relates to the individual in a holistic life/ecosystem sense. Where data access rights are executed (or

data is personally shared [4.3.2]), the attention is on the data itself: what it says. But as Case Study Two showed, some of the most desired information was not the data itself, but handling information and inferences—information that can only come from metadata, which was rarely forthcoming [Table 5.3]. Metadata could include many facets that could be quantified and recorded, as illustrated in Figure 8.5, which I created at BBC R&D:

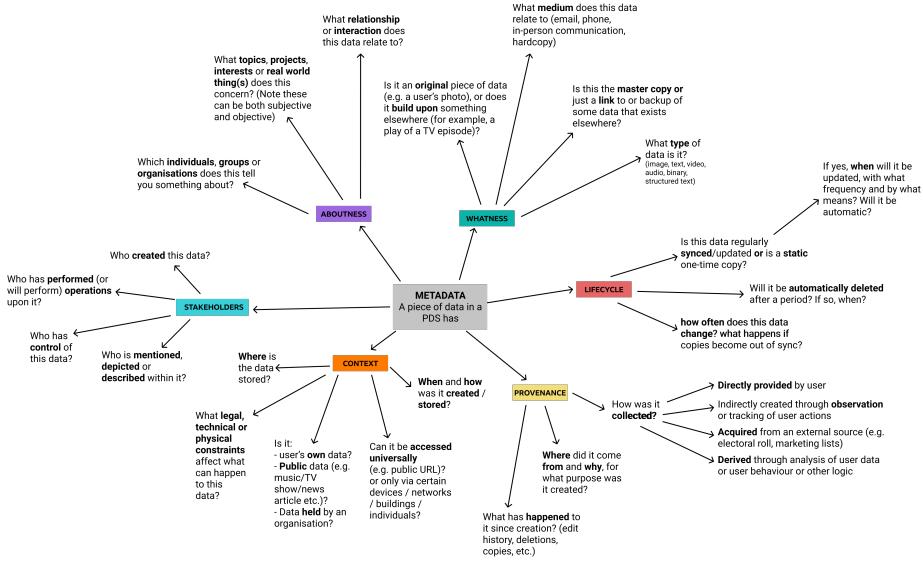


Figure 5: Figure 8.5: Some of the Many Aspects of Metadata that Might Exist About a Datapoint or Dataset

These facets can be mapped back to the 5 W's that collectively make up the user's *context* [Abowd and Mynatt (2000); 2.2.2]. Many of these facets are not explicitly recorded today, or would take significant work to capture. Nonetheless, this exploration shows how data can be better contextualised, supporting contextual and associative approaches [2.2.2]. This leads to Insight 5:

INSIGHT 5: We Must Know Data's Provenance

Metadata is what gives information *context*. Context is critical to *sensemaking* [2.2.3] and enables good experience-centred design [2.3.2; 2.3.3]. Without context, data loses meaning [5.4.3]. Collecting historical data about the individual is important for reflection [2.2.3] and considered valuable [4.3.3], but knowing the **history of a piece of data** allows its context to be understood. Data is not neutral, and is inherently biased, since it was created for a specific purpose with a specific agenda in mind (Gitelman, 2013; Neff, 2013). To combat this bias, more context is needed. Significant research in this space has been undertaken by Professors Mike Martin and Rob Wilson at Northumbria University, formerly Newcastle University, who promote the idea of **data with provenance**; in other words:

Data must carry with it the details of why it exists, how it came to be, and what has happened to it since its inception.

(continues...)

Provenance should be communicated alongside any visualisation of the data, in order for it to be fairly assessed in context. Provenance is essential for data to be trusted, argues Martin, and should be quite granular: a piece of data should be attributed not just to an individual or organisation, but to the relationship between role-holding individuals in a specific context. Greater insights can be gained when considering all actions upon data as motivated communications from one party to another; only by capturing this information in-situ can the data's meaning be fully appreciated (Martin, 2022). This framing essentially advances the concept of history tracking [2.2.3] into the sociotechnical, ecosystem-aware problem space. While everyday system designs have not approached this level of granularity, the importance of data provenance has been recognised in the PIM space. Temporal PIM systems [2.2.2] from Lifestreams (Freeman and Gelernter, 1996) to *activity streams* (Hart-Davidson, Zachry and Spinuzzi, 2012) rely upon data provenance in some form. A study by Jensen *et al.* concluded that provenance tracking can be valuable for identifying related documents, a critical part of knowledge work today (Jensen *et al.*, 2010). Lindley *et al.* proposed the idea of *file biographies*, which view the lifetime of a file as something that should remain connected, so it could be traversed in order to understand the context of the file different moments of interaction (Lindley *et al.*, 2018). This comes close to Martin's vision but does not capture the motivation for each interaction. While provenance capture is not a solution in its own right to the understanding of data and of ecosystems, it is clear that data with provenance is very likely to be a valuable part of any design that aims to provide understanding of complex and invisible personal data ecosystems.

Paying attention to ecosystem information, metadata and provenance, facilitates

a new space that, at the time of writing in 2022, almost no-one is building for. For people to manage their digital world, they need a map. This is the first step on the road to giving individuals oversight of their personal data ecosystem.

1.4 Obstacles to the HDR Objective of Ecosystem Negotiability

There are three distinct obstacles to ecosystem negotiability:

- the intrinsic structures that give data holders power 8.4.1,
- the trend of actively diminishing user agency 8.4.2, and
- the intractable data self 8.4.3.

1.4.1 Hegemony through Data Holding

It is in the pursuit of oversight [6.2.2] and involvement [6.2.3] that the impact of the power imbalance [2.1.2] becomes most clear; unlike the other HDR objectives, individuals cannot act to claim ecosystem negotiability for themselves. Negotiability means having the power to act, and in the context of systems and interfaces owned and designed by service providers, **that power can only be given**. The hegemony of data holders is therefore the greatest obstacle to this objective, so it is vital to examine the nature of that power. Where does it come from?

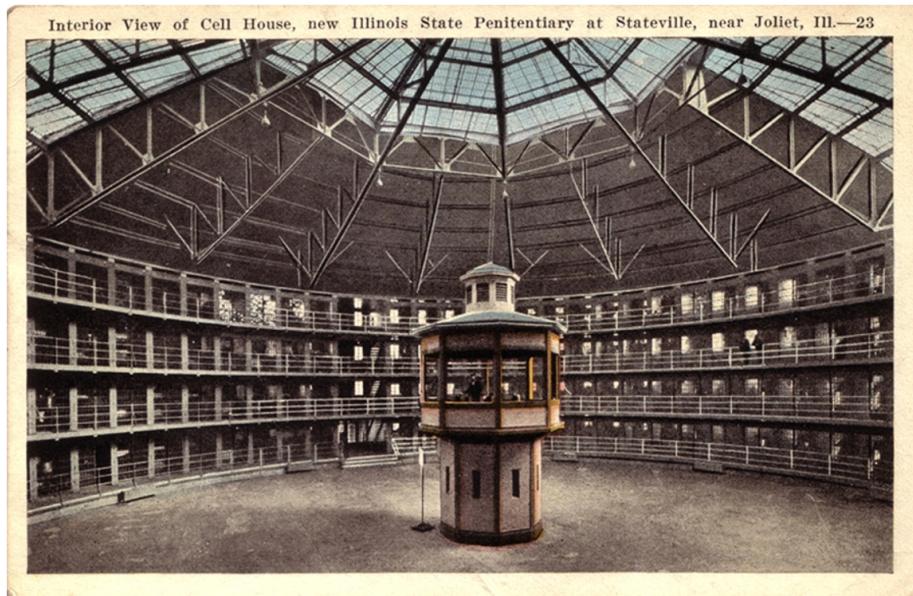


Figure 6: Figure 8.6: The Panopticon Structure of the Illinois State Penitentiary

A helpful analogy for the relationship between provider and user can be seen

in the design of Jeremy Bentham's *panopticon* (Bentham and Bozovic, 1791), a real-world version of which is pictured in Figure 8.6. The panopticon is an 18th century prison architecture that elevates the power of the (hidden) prison guards to observe all the prisoners easily at any time while removing prisoners' privacy and providing them no ability to observe those in power. As in Orwell's *Nineteen Eighty-Four*, individuals are unable to know when they are being watched (in this case, because the guards are hidden from view by one-way screens). This enforces compliance. Structuralist philosopher Foucault interpreted the panopticon as a political design, recognising that human environments can be configured to influence or regulate behaviour, in order to defend the power of the ruling class (Foucault, 1975). Such designs embody his four principles:

- **Pervasive Power:** the guards see everything all the prisoners do, all the time
- **Obscure Power:** the guards can see into any cell at any time, but the prisoners can't know when, how or why they are being observed
- **Direct Violence Made Structural:** the structure motivates the prisoners to self-regulate their behaviour without being coerced (through beating or punishment)
- **Structural Violence Made Profitable:** having been made compliant by the structure, the prisoners can be put to work for the benefit of those in power, as it is the only option available to them.

We can see at least three of these traits in modern Internet platforms such as Facebook today. These platforms monitor users' behaviour (pervasive power) without their knowledge and without accountability (obscure power). Interfaces are designed to offer only those actions that benefit the platforms (for example, clicking ads, sharing content or spending more time on site (structural violence made profitable)). This has happened through the processes of *platformisation* and *infrastructurisation* (Helmond, 2015; Plantin *et al.*, 2018), which have supplanted the Web 2.0-era promise of a free, open Internet that could have been more empowering to individuals.

Through the control of data and of interface design—the only channels through which they can be observed—**service providers and platforms assert a structural power over the digital landscape**. Just as the design of the panopticon regulates the behaviour of the prisoners, so the configuration of platforms, apps and service interfaces we use regulate and limit our behaviour as users. As Lessig wrote, '*code is law.*' (Lessig, 2000). This infrastructural power is explained further in [Insight 6] below.

Structural power is not the only form of power which modern-day data-centric service providers hold. Jasperson *et al.*'s extensive review of types of power in the context of technology organisations (Jasperson *et al.*, 2002) identifies 23 different power paradigms, of which at least 13 can be, and are, asserted by data-centric organisations today:

- **authority:** ownership of technology or infrastructure (for example of

websites, servers and code)

- **resource control**: controlling the flow of resources (in this case of information/data)
- **systems/structural power** structural manipulation of others (as detailed above)
- **rational power**: controlling decision-making processes
- **disciplinary power**: using an influential position to affect others' mental models (for example, positioning location tracking as theft resilience)
- **zero sum power**: winning a battle for ownership/resource control at the other party's expense (e.g. losing control of your sacrificed data)
- **behavioural influence**: persuading others to carry out the desired behaviour (e.g. restricting features to motivate subscription payments, or promoting certain content or actions)
- **interpretative influence**: determining how reality is externally represented (e.g. Facebook determining the way in which your social network is represented to you)
- **network centrality**: becoming an indispensable hub of a wider ecosystem (for example, Facebook/Google dominance in online ad-brokering)
- **processual power**: changing processes for competitive advantage (for example, platforms offering preferential APIs or rates to compliant partners)
- **socially shaped power**: influencing a wide audience to settle upon a preferred interpretation (e.g. using dominant market position to dominate debates e.g. about privacy norms)
- **interpretive power**: creating the internal representations of reality within an organisation (for example, presenting unpopular attitudes to data privacy to staff as normal/acceptable/beneficial for business)

INSIGHT 6: Data Holders use Four Levers of Infrastructural Power

Hestia.ai [ARI7.2] have produced a model to explain the mechanisms by which technology companies gain power and use it to shape today's digital landscape. In this model, *infrastructural power* comes from three things:

-
- *technical ability*,
 - *organisational ability*, and
 - *the acquisition of data about individuals and populations*.

(continues...)

As organisations (especially platforms) collect more data, and grow in market influence or technical capability, they gain power over individuals and over other organisations. They exert this power using four 'levers'. Simplified and expressed in the terms of this thesis, these are:

1. **Collect & Interpret Data to Acquire Knowledge**: Data and signals are collected from individuals and interpreted in order to infer their intents

and interests. For example, Google collects raw GPS and wi-fi hotspot data from mobile phones, which it then statistically analyses to infer which shops or venues you visited and what forms of transport you used, increasing Google's knowledge about individuals and populations.

2. **Present Content and Configure Structures to Influence Individual Behaviour:** Knowledge of individual intents and interests is exploited within user interfaces to influence desired individual actions. For example, Facebook or presents a user with a product relevant to their interests, which they are motivated to click upon, generating ad revenue. Another example would be Twitter manipulating the content of the user's feed to show more tweets from conversation topics where they can show promoted tweets, increasing ad revenue.
3. **Configure Structures to Improve Knowledge Acquisition:** A provider uses its dominant position to force other organisations to improve the provider's ability to acquire knowledge. For example, Google provides free analytics tools to web developers, but requires the end users of those client websites to supply visitor data back to Google, increasing their ability to acquire knowledge about individuals and populations.
4. **Configure Structures to Disadvantage Others:** Certain providers (typically of operating systems or popular devices) can configure the structural relationships between other parties. For example, a smartphone manufacturer could limit data exchange between other apps, while still extensively collecting data signals themselves, such as when Google was found to be collecting call history from Android's dialer app.

(continues...)

The precise mechanisms and techniques employed when exerting infrastructural power, as well as the social and market consequences of these practices, are explored in detail in Hestia.ai's digipower technical reports, of which I was a co-author (Bowyer, Pidoux, *et al.*, 2022; Pidoux *et al.*, 2022).

The research highlights that providers' power is far greater than many realise. Unlike in the physical realm, providers of popular online platforms can **reconfigure the landscape to change the way that individuals perceive reality**, in line with the powers of interpretative influence, behavioural influence and socially shaped power described above (Bowyer, Pidoux, *et al.*, 2022).

Providers control the extent to which (if at all) data stored behind the scenes, and internal processes that use that data, are visible, and how data and processes are represented.

The model shows that the accumulation of data (and hence, information) is implicitly and objectively a form of power, consistent with participants' observations in 5.4.4. As long as current service providers are free to collect so much personal information, the information landscape is likely to remain imbalanced and individuals will not be able to acquire ecosystem negotiability.

(continues...)

This insights shows that the most powerful data holders exert huge influence over the digital landscape, in terms of what is *knowable* and what is *doable*. HDR reformers' abilities to balance the landscape are hindered by the fact that they are operating in a landscape that the incumbent platform and service providers effectively control.

1.4.2 The Active Diminishing of User Agency

The second major obstacle to ecosystem negotiability is that platformisation and power exertion are not a one-off transition, but rather an ongoing process. Today's platforms exhibit a **continuing trend of actively diminishing individuals' agency**, especially in the last decade. When software was sold in a box, manufacturers competed based upon which product would let the user take home the greatest range of features and capabilities. New releases with new features drove new product sales. But in the cloud computing era, a smaller set of core features done well is sufficient to guarantee an ongoing subscription revenue from a user. Cost savings in development and support costs can be made by reducing feature sets. Constrained, compliant users are easier to manage. The relentless pursuit of increased profits and further cost saving sees products lose, not gain, features. Interfaces are reshaped to serve businesses' interests first and foremost. Providers' focus on making user behaviours constrained, predictable and profitable, more than meeting their needs or providing maximal value [2.3.5]. Plantin *et al.* describe the particular harmful influence on the ecosystem of Facebook's power exertions:

"Facebook is a formidable force in a profit-motivated platformisation which is beginning to eat away at the Open Web. This entails moving away from published URIs and open HTTP transactions in favour of closed apps that undertake hidden transactions with Facebook through a Facebook-controlled API."—Plantin *et al.* (2018)

Here are just a few examples of the ways in which users' agency has been, and continues to be, diminished:

- Facebook closed their RSS feeds, and later parts of their APIs, meaning that users could no longer consume their friends' posts in any other environment than the ad-filled and manipulated Facebook main feed. Later, they eliminated feeds of friends' posts and favourite pages, removing users' ability to compartmentalise their content viewing to certain friends groups. The 'Friends' page on Facebook currently shows a list of recommended new friends. To access your current friend list requires an extra click. Encouraging users to grow their networks is prioritised over user convenience.
- Twitter closed the parts of its APIs that allowed real-time notifications and access to one's home feed, killing off primary functionality for a healthy ecosystem of third-party Twitter clients that increased user choice (Newton,

2018). TweetDeck, a major third-party Twitter client was acquired, and later shut down, as was Twitter's own desktop client. Eventually, the only option left to users was to use the web interface. (Gayomali, 2015; Hatmaker, 2018; Siegal, 2022)

- Apple has been diminishing users' agency for a long time. Users cannot open up iPhones even to change the battery without invalidating their warranty. Apple have removed disk drives, headphone ports, SD card slots and other ports. Certain parts of the hard drive on macOS devices are now read-only and non-writeable by users.
- Facebook recently announced they will no longer store users' historical location data (though they will still use location information) (Pegoraro, 2022). This means users will lose the capability to access historical location records. But this makes it harder for users to see how their location data will be used in future, as there will be no historical log to examine. Data-centric companies can change their practices to limit agency and reduce accountability.
- Online news and discussion site Reddit has removed content access for non-logged in users, and uses deceptive techniques to present advertisements that look like posts from users, and to discourage users from appearing offline. These patterns have been described as *disrespectful design* (Regoje, 2021).
- In an example from the public sector, through my work on the SILVER project [3.4.1] just prior to the introduction of the GDPR in 2018, I heard whispers in at least one local authority of plans to 'shift from getting data collection consent from supported families towards simply informing them of our practices' (in other words, removing their choice). The instinct to further organisational interests over those of the individual appears not to be limited to commercial data holders.
- In a similar vein, TikTok recently announced that it would rely on *legitimate interest* rather than consent when it comes to using users' activity data to personalise the app experience. This removing users' ability to withdraw consent to such use. This plan has subsequently been paused after warnings that this might breach GDPR (Lomas, 2022).

Unchecked, trends to reduce users' agency and further providers' interests at the expense of human autonomy are likely to continue. Today's data-centric systems suffer from a lack of consideration to individual welfare. Data centricity encourages neglect of the human end user perspective, creating potential for harm:

*"There are certain things you do not in good conscience do to humans.
To data, you can do whatever you like."*—Sonnad (2022)

The trend to diminish users' agency is needs explicit targeting if data interfaces are to become more free-flowing (Bowyer, 2018), and if ecosystem negotiability is to be realised. Somehow, the trend needs to be halted, before it can be reversed. The TikTok example suggests this can only be achieved through regulatory

changes.

1.4.3 The Intractable Data Self

The third obstacle to ecosystem negotiability is *the intractable data self*. Data about individuals serves as their proxy [[Bowyer *et al.* (2018); 5.4.4]. This is their *data self* [4.4.1]. If it is incomplete, inaccurate or unfair—highly likely given the difficulties of representing people in data (Martin, 2007; Cornford, Baines and Wilson, 2013)—which can cause harm (Bowyer *et al.*, 2018; Crossley, 2022). Yet currently, although some legal rights to data correction exist (Information Commissioner’s Office, 2018), people cannot modify or assert control over this most important version of themselves—the version of them that exists in data. Even when data can be seen, people lack the ability to **exert influence over their data self** [5.5.2; Cornford, Baines and Wilson (2013)]. To address this obstacle, HDR reformers should explore giving people a role in the curation of their data self [4.4.3; 5.5.2] and 6.3].

To date, research and innovation on ecosystem negotiability has been very limited. It is easier to find business models and research funding for narrow and well-defined contexts. Without a business motive, only non-profit socially-focussed research organisations such as BBC R&D and Sitra have found themselves well-equipped to explore this problem space. Nonetheless, there is an urgent societal need. People need to reclaim their data selves, and be given control over their digital lives at the broadest level.

1.5 Obstacles to the HDR Objective of Effective, Commercially-Viable and Desirable Systems

The previous four subsections considered the obstacles to the HDR objectives [7.7]. However, through pursuit of these objectives, and through observation of public and business responses to human-centricity, I observed additional obstacles that affect *all* efforts to make progress towards improving HDR. The main challenge is around building such disruptive systems that are so different from the status quo:

Businesses and individuals will not readily invest time and money in HDR, because it is unfamiliar.

1.5.1 A Lack of Individual Demand

Customers are not demanding HDR capabilities in their lives, and, all but the most socially-responsible businesses do not see value in an approach that runs so contrary to current business models, based as they are on data accumulation and the constraining of customer experiences.

Data is overwhelming, complex, and ‘sounds boring’. Engaging with your personal data economy to any degree more than that of passive consumer is hard work. People routinely accept data sacrifice, click through T&Cs and cookie

banners and are unwilling (or in some cases lack sufficient technical literacy, comprehension or skill) to do the work of asserting control over their digital lives. There is not a clear demand for holistic digital life management and control. Research in this this and at Cornmarket suggests that even if human-centric information systems and more inclusive service interaction practices emerged, people would not be inclined to use them in great numbers. It could seem like hard work or not worthwhile. This obstacle that affects all HDR improvement approaches. Indeed is why many companies in the emergent PDE economy [2.3.4] struggle to find a business model. There are clear benefits, but better HDR does not appear to something a mainstream audience will pay for. This should not deter disruptive innovation nor diminish the potential value for such tools. As automobile pioneer Henry Ford famously said, *“If I had asked people what they wanted, they would have said faster horses.”* Nonetheless, it is a clear overarching obstacle, which Insight 7 attempts to address.

INSIGHT 7: Human-centred Information Systems must serve Human Values, Relieve Pain and Deliver New Life Capabilities

Through work at BBC R&D exploring how to better connect people with their data, it became clear that there is a way to combat such indifference and apathy of users. It emerges from the realisation that the way people find value in data is to connect it their lives. The more that people see relatable life information and can imagine ways to harness that information in their everyday life, the more motivated they will be. BBC R&D conducted some research (Forrester, 2021) that identified fourteen specific Human Values that people seek to satisfy in their lives, which are shown in in Figure 8.7. These are, at the most abstract, goals that people care about in their daily existence.



Figure 7: Figure 8.7: Human Values, as Identified in BBC R&D Research Funded by Nesta

(continues...)

Given these and the earlier observation that life information is what makes data relatable, the insight I offer here is that the way to make people care about their data is to use it to help them in their life. By starting with a focus on a user's world, one can then focus in on their life, and then the data that represents elements of that life. Then, the individual has a vested interest.

Systems and features should be designed from this life-centric perspective. This is known as *value-centred design* (Reber and Duffy, 2005) and it has been argued that this should become the guiding design philosophy in HCI (Cockton, 2004). And to offer true individual value, all human-centric system designs must also consider *context* [2.3.2], *environment* (Abowd, 2012) and *experience* [3.2.1]. In business modelling, there is a tool called the *value proposition canvas*, which identifies three ways of conceptualising value: *gain creators*, pain relievers and *jobs-to-be-done*. If we can use those concepts to inform our designs, we can produce better human-centric functionality - relieve an individual's pain points, help them complete their tasks, or offer them some gain over the status quo. In the HDR space, given the lack of existing tools for digital life management, we have the opportunity to create quite a unique type of gain: new capabilities over your digital life** that you have never had before. This ability to do new things has been identified as key ingredient of user empowerment (Meschtscherjakov, Wilfinger and Tscheligi, 2014; Schneider *et al.*, 2018).

Here is an example of what this value-centric approach might look like in the HDR space: Myself and BBC R&D colleague Jasmine Cox imagined focusing on address books and contact lists as a strong relatable starting point to generate demand for a human-centric interface. This could provide people with new life capabilities while also relieving pains. Many people have address and contact information scattered far and wide, and face a complexity they cannot easily manage when it comes to the automated syncing and sharing of potentially sensitive contact information between devices, apps and providers. Developing human-centric personal information management capabilities to bring that messy situation under control would offer a clear and tangible benefit to users. In Figure 8.8, we show how there could be a strategic path, beginning with detecting ecosystem and life information from the individual's calendar and e-mail inbox, through to building up to more holistic life-level PDS capabilities.

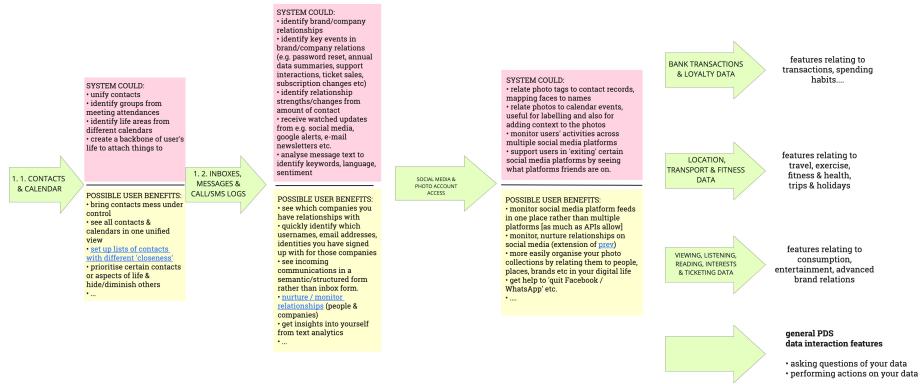


Figure 8: Figure 8.8: A Contact-and-Calendar-centric PDS Approach

(continues...)

A helpful example is that of a vacation from my 2011 article (Bowyer, 2011) and shown in Figure 8.9. Today, all the information around such a holiday is scattered into multiple systems - emails, online provider bookings, chat logs, cloud synced photos, web browser bookmarks, smartphone location logs, etc. It is not hard to imagine that a system that was able to bring all related information about that vacation together in one central interface (mock-up in Figure 8.10) could deliver huge value to users and be very compelling. Such context-targeted human-centric offerings can have a much greater chance of generating interest and impact than offerings that merely allow you to ‘organise your data’ or some other abstract phrasing.



Figure 9: Figure 8.9: The Scattered Data Relating to a Vacation

1.5.2 Closed, Insular and Introspective Practices

The kind of life-spanning, unifying interfaces described in the insight above are nothing like the interfaces that are built today, as they span across different providers’ data and services. This highlights the secondary obstacle that all HDR system builders will face, whichever objective they wish to target: **closed, insular and introspective practices**.



Figure 10: Mock-up of a Unified Interface for a Vacation

(the so-called ‘*big five*’) build **proprietary, incompatible silos** or ‘*walled gardens*’—sub-Internets that pretend that the alternatives do not even exist, in order to encourage a flow of money and attention to their own products and services.

Commercial motives encourage them to get users to spend time in their own proprietary spaces (so that resultant ad revenue can be captured) and in order to maintain subscription revenues it is in providers’ interests to make it hard for individuals to leave or switch providers. In effect, providers build for a world that does not exist, where every individual is imagined to only interact with that single company’s interfaces. I would argue, for example, that Google’s venture into social networking with Google+ did not succeed because it failed to build for a reality where most people and their friends were already on Facebook.

1.5.3 A Lack of Organisational Investment in HDR

But one can understand their motives; there is little incentive to open up the ecosystem when the free flow of information and of users might result in loss of income for the company in question. Users with negotiability would be more able to leave. And this also encourages keeping users in the dark [5.4.2]. The less agency and negotiability that users have, the more freedom the provider has to do exactly what they want with their data. In this context, users are ‘*pathetic dots*’ (Lessig, 2000) or ‘*docile bodies*’ (Foucault, 1975).

The tendency of organisations to work in closed, introspective practices and to be resistant to opening up data or services is not solely motivated by commercial reasons: the public sector has a vastly complex, closed and fragmented ecosystem [Pollock (2011); Copeland (2015); 4.1.2]. Efforts to build a system to share

health data with support workers for the SILVER project [3.4.1] proved hugely challenging. Sometimes the challenge was a more technical one - incompatible data formats that are hard to reconcile, or data being stored in legacy systems with no public API that would allow programmatic access to that data, or issues around licensing. Data sharing agreements must be established, especially in the public sector which is by its nature more liable to scrutiny and accountability. But more than these technical or procedural issues there was resistance to change in data processes and an unwillingness to share data between agencies, often motivated by a fear of legal repercussions. **Data-centrism encourages insular thinking:** it encourages organisations to codify the world into their own systems, processes and formats for their own use.

1.5.4 A Lack of Interoperability

And yet, for effective HDR, **data needs to be separable from services**. The more users data is tightly coupled to specific services, the less agency users have and the harder it is to build life-centric systems. On BBC R&D's Cornmarket project, attempts to build an interface for users to import data from multiple popular Internet services proved to be a hugely complicated endeavour, requiring access to many different APIs or manual exports and imports of data by users. There needs to be greater interoperability and greater establishment and adoption of **standard formats for exchanging human information** (as distinct from establishing standards for data or service-specific APIs). As mentioned above, platformisation breaks the Open Web (Plantin *et al.*, 2018). To overcome this, companies must be persuaded that human-centric thinking, interoperability and transparency has not just social benefits, but business benefits too.

1.5.5 Insufficient Machine Understanding of Human Information

But at an abstract level the technical obstacle, the problem is one that has always faced the tech industry, which is that there often is no universally agreed way to represent important concepts - in this case human-centric information concepts such as events, social media posts, website visits, location history information, app activity, etc. And any entity that does create a standard then faces the challenge of trying to persuade others that their standard is the best one to use. In general, standards work best when established by non-commercial industrial standards bodies (for example the World Wide Web Consortium (W3C) or International Organisation for Standardization (ISO) and then mandated through policy such as European Union law. Such standards much be established with input from industry experts.

INSIGHT 8: We Need to Teach Computers To Understand Human Information

In order to move towards standardised ways to store and unify personal data from multiple sources, computer systems must be taught to understand the information within the data, and how it relates to an individual and the world. This moves beyond just capturing data provenance: put simply, **computers need to understand human information**. They need to move beyond files (Bowyer, 2011) and databases, and begin to perform operations on human informational concepts, and to associate those concepts according to what they mean - i.e. ***semantically***. This is a preliminary step that will enable the building of systems and interfaces that are able to deal in human concepts and represent the elements of everyday life.

We need to store **semantic context and semantic associations**, i.e. the meaning of things, not just raw bundles of data. This is advocated by the Web's inventor Tim Berners-Lee in his vision of a Semantic Web (Berners-Lee, Hendler and Lassila, 2001) and by proponents of *networked* and *semantic* PIM systems, as detailed in 2.2.2. There is a need to develop standard ways to digitally model facts and assertions about users' lives, so that those disparate pieces of data can be unified, connected, correlated and compared. Some standards are already developing, such as *data shapes* ('ShapeRepo: Make your apps interoperable', 2022). And the extraction of meaning from data is a problem domain all of its own. Sizable industries have built up around Content Analytics and Enterprise Content Management. But to consider the problem at its simplest level, I offer this insight: Through **the capture of metadata** at the point of data recording, and through **subsequent programmatic analysis** of stored data, as illustrated in Figure 8.11, we can begin to teach computers what the data we store represent.

Machine learning technologies and Artificial Intelligence have pushed machine understanding of human words, images and content to impressive levels in recent years and such technologies can certainly be helpful, but in fact at the core what we are talking about here is something much simpler than AI; It is simply about labelling datapoints in as many different ways as possible so that those datapoints can be associatively retrieved from many different angles, and providing humans with ways to amend incorrect labels and to reclassify data or apply new semantic associations. Issues of interoperability for PDS systems are being actively explored and developed in the '*Solid*' community (Bansal, 2018; Berners-Lee, 2022) in pursuit of a decentralised web (Verborgh, 2017).

Such approaches are in their infancy, and have not yet been adopted extensively in commercial settings. Even after addressing the obstacles of end-user buy-in and the technical complexities of building human-centric systems, data-driven corporations, motivated as they are by profit and business success (and smaller online organisations too) need to be persuaded of the business value of transparency, interoperability and human-centricity. This is explored further in

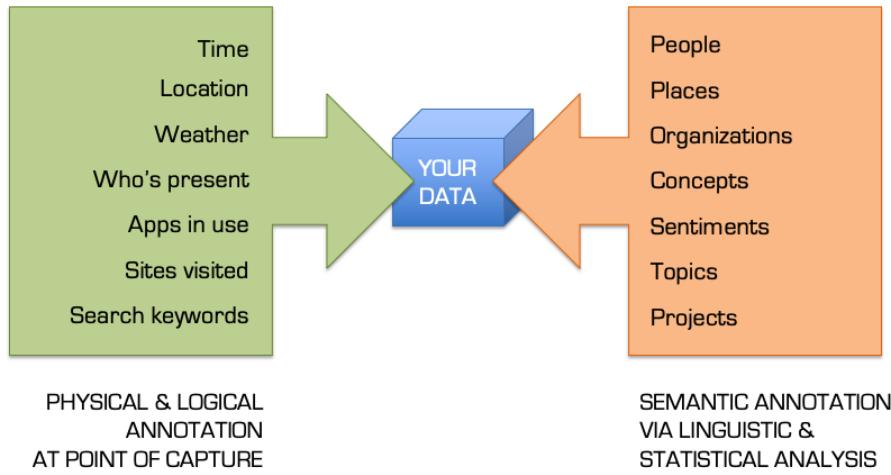


Figure 11: Figure 8.11: Annotating Data with Semantic Context

9.5.

In summary, whichever of the above four HDR objectives are targeted, all HDR reformers involved in building HDR systems must:

1. create, adopt and co-ordinate around **new standards** for human information storage and management
2. invest in systems that elevate computers from data-processing machines to **human-information-processing machines**, and
3. make a persuasive case to both businesses and individuals that the new approach offers **tangible, previously unavailable value**.

1.6 Summation of Chapter 8: From Obstacles to Opportunities

This chapter has presented, in effect, a map of the HDR landscape. It has described the major obstacles to better HDR including invisible, inaccessible, scattered, immobile, unmalleable, or unrelatable data; the complexity of current personal data ecosystems; a lack of metadata and machine understanding; the ongoing exertion of power by introspective data holders to diminish user agency; and a lack of demand and investment in HDR. Figure 8.1] showed an overview of how these different obstacles relate.

From the details above, HDR reformers can ‘hit the ground running’ with a clear understanding of some of the challenges that exist, as well as having been introduced to some insights that may suggest possible strategies to tackle them. A good high-level understanding of the landscape combined with some specific

ideas should be valuable for anyone working in the HDR space. Chapter 9 expands further on these understandings and insights, presenting four specific and detailed strategic approaches to tackling the obstacles in this chapter.

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.
- Abowd, G. D. (2012) ‘What next, ubicomp?: celebrating an intellectual disappearing act’, in *Proceedings of the 2012 ACM conference on ubiquitous computing*. New York, New York, USA: ACM Press, pp. 31–40. doi: <http://dx.doi.org/10.1145/2370216.2370222>.
- Abowd, G. D. and Mynatt, E. D. (2000) *Charting Past, Present, and Future Research in Ubiquitous Computing*. 1, pp. 29–58. available at: <https://www.cc.gatech.edu/fce/pubs/abowd-mynatt-tochi-millenium.pdf>.
- Allen, D. (2015) *Getting things done: The art of stress-free productivity*. Penguin.
- Ananya (2020) ‘Java: Pass by value or pass by reference’, *Medium*. available at: <https://medium.com/swlh/java-passing-by-value-or-passing-by-reference-c75e312069ed>.
- Bansal, A. (2018) ‘An introduction to SOLID, tim bernes-lee’s new, re-decentralized web’, *FreeCodeCamp*. available at: <https://www.freecodecamp.org/news/an-introduction-to-solid-tim-berners-lees-new-re-decentralized-web-25d6b78c523b/>.
- Bentham, J. and Bozovic, M. (1791) *The panopticon writings (1791/2011)*. Verso Books (Radical thinkers). available at: <https://books.google.co.uk/books?id=VbpvDwAAQBAJ>.
- Bergman, O., Beyth-Marom, R. and Nachmias, R. (2003) ‘The user-subjective approach to personal information management systems’, *Journal of the American Society for Information Science and Technology*, 54(9), pp. 872–878. doi: 10.1002/asi.10283.
- Berners-Lee, T. (2022) ‘Solid: Sir tim bernes-lee’s vision of a vibrant web for all’. Inrupt. available at: <https://inrupt.com/solid/>.
- Berners-Lee, T., Hendler, J. and Lassila, O. (2001) ‘The Semantic Web’, *Scientific American*, 284(5), pp. 34–43. available at: <https://jstor.org/stable/10.2307/26059207>.
- Bogers, S. et al. (2016) ‘Connected baby bottle’, pp. 301–311. doi: 10.1145/2901790.2901855.
- 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.
- Bowyer, A. (2020) ‘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., 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, A. (2018) ‘Which email providers are scanning your emails?’, *maketecheasier*. available at: <https://www.maketecheasier.com/which-email-providers-scanning-emails/> (accessed: 21 August 2022).
- Cockton, G. (2004) ‘Value-centred HCI’, *NordiCHI*.
- Copeland, E. (2015) *Small Pieces Loosely Joined: How smarter use of technology and data can deliver real reform of local government*. available at: www.policyexchange.org.uk https://policyexchange.org.uk/publication/small-pieces-loosely-joined-how-smarter-use-of-technology-and-data-can-deliver-real-reform-of-local-government/.
- Cornford, J., Baines, S. and Wilson, R. (2013) ‘Representing the family: how does the state ‘think family?’’, *Policy & Politics*, 41(1), pp. 1–19. doi: 10.1332/030557312X645838.
- Croll, A. and Bowyer, A. (2009) ‘Why email clients need to change’. available at: <https://web.archive.org/web/20120128050842/http://gigaom.com/2009/04/24/why-email-clients-need-to-change/>.
- Crossley, S. (2022) ‘Misrecognising ‘complex’ families: A social harm perspective’, in *Negotiating families and personal lives in the 21st century*. Routledge, pp. 30–45.
- Deckers, E. (2018) ‘Data-enabled design’, *UXDX*. available at: <https://uxdx.com/blog/data-enabled-design/>.
- Forrester, I. (2021) ‘Talking about human values and design’, *BBC Research & Development*. available at: <https://www.bbc.co.uk/rd/blog/2021-07-talking-about-human-values-and-design>.
- Foucault, M. (1975) ‘Discipline and punish: The birth of the prison’, *New York*. Pantheon Books.
- Freeman, E. and Gelernter, D. (1996) ‘Lifestreams: A Storage Model for Personal Data’, *SIGMOD Record (ACM Special Interest Group on Management of Data)*. Association for Computing Machinery (ACM), 25(1), pp. 80–86. doi: 10.1145/381854.381893.
- Gayomali, C. (2015) ‘Why twitter is killing TweetDeck’. The Week. available at: <https://theweek.com/articles/467040/why-twitter-killing-tweetdeck>.
- Gitelman, L. (2013) *Raw data is an oxymoron*. edited by Lisa Gitelman. MIT Press, p. 182. available at: <https://mitpress.mit.edu/books/raw-data-oxymoron>.
- Gurstein, M. B. (2011) ‘Open data: Empowering the empowered or effective data use for everyone?’, *First Monday*. First Monday, 16(2). doi: 10.5210/fm.v16i2.3316.
- Hart-Davidson, W., Zachry, M. and Spinuzzi, C. (2012) ‘Activity streams:

- Building context to coordinate writing activity in collaborative teams', in *SIGDOC'12 - proceedings of the 30th ACM international conference on design of communication*. New York, New York, USA: ACM Press, pp. 279–287. doi: 10.1145/2379057.2379109.
- Hatmaker, T. (2018) ‘Twitter is killing its twitter for mac desktop client’. TechCrunch.
- Helmond, A. (2015) ‘The platformization of the web: Making web data platform ready’, *Social media + society*. Sage Publications Sage UK: London, England, 1(2), p. 2056305115603080.
- Information Commissioner’s Office (2018) ‘Your data matters - Your rights’. available at: <https://ico.org.uk/your-data-matters/>.
- Jasperson, J. (Sean). et al. (2002) ‘Review: Power and Information Technology Research: A Metatriangulation Review’. Society for Information Management; The Management Information Systems Research Center. doi: 10.2307/4132315.
- Jensen, C. et al. (2010) ‘The life and times of files and information: A study of desktop provenance’.
- Jones, W. (2011) ‘The Future of Personal Information Management Part I: Our Information, Always and Forever’.
- Lemley, M. A. (2021) ‘The splinternet’, *Duke Law Journal*, pp. 1397–1428. available at: <https://perma.cc/92LZ-B8DN%5D>.
- Lessig, L. (2000) ‘Code is law: On liberty in cyberspace’, *Harvard Magazine*. available at: <https://www.harvardmagazine.com/2000/01/code-is-law-html>.
- Li, I., Forlizzi, J. and Dey, A. (2010) ‘Know thyself: Monitoring and reflecting on facets of one’s life’, *Conference on Human Factors in Computing Systems - Proceedings*, pp. 4489–4492. doi: 10.1145/1753846.1754181.
- 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.
- Lomas, N. (2022) ‘TikTok ‘pauses’ privacy policy switch in europe after regulatory scrutiny’. TechCrunch. available at: <https://techcrunch.com/2022/07/12/tiktok-pauses-privacy-policy-switch/>.
- Martin, M. (2007) *Research note: Representing identity and relationships in information systems*, *International Journal of Business Science & Applied Management (IJBSAM)* Suggested Citation: Martin, pp. 47–51. available at: <http://hdl.handle.net/10419/190583><https://creativecommons.org/licenses/by/2.0/uk/>.
- Martin, M. (2022) ‘The trustworthy, governable platform: A concept of safe social spaces in the public network’. in prep.
- Meschtscherjakov, A., Wilfinger, D. and Tscheligi, M. (2014) ‘Mobile attachment-Causes and consequences for emotional bonding with mobile phones’, *Conference on Human Factors in Computing Systems - Proceedings*, pp. 2317–2326. doi: 10.1145/2556288.2557295.
- 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).

- Neff, G. (2013) ‘Why Big Data Won’t Cure Us’, *Big Data*, 1(3), pp. 117–123. doi: 10.1089/big.2013.0029.
- Newton, C. (2018) ‘Twitter officially kills off key features in third-party apps’. The Verge. available at: <https://www.theverge.com/2018/8/16/17699626/twitter-third-party-apps-streaming-api-deprecation>.
- Pegoraro, R. (2022) ‘Facebook will soon stop tracking your location and delete your location history’, *Fast Company*. available at: <https://www.fastcompany.com/90750241/facebook-will-soon-stop-tracking-your-location-and-delete-your-location-history>.
- Pidoux, J. et al. (2022) *Digipower technical reports: Understanding influence and power in the data economy*. doi: 10.5281/zenodo.6554155.
- Plantin, J. C. et al. (2018) ‘Infrastructure studies meet platform studies in the age of google and facebook’, *New Media and Society*. SAGE Publications Ltd, 20, pp. 293–310. doi: 10.1177/1461444816661553.
- Pollock, R. (2011) ‘Building the (Open) Data Ecosystem – Open Knowledge Foundation Blog’. available at: <https://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/> (accessed: 23 July 2019).
- Reber, M. and Duffy, A. (2005) ‘Value centred design: Understanding the nature of value’, *International Conference on Engineering Design*.
- Regoje, O. (2021) ‘Reddit’s disrespectful design’. available at: <https://ognjen.io/reddits-disrespectful-design/>.
- Schneider, H. et al. (2018) ‘Empowerment in HCI - A survey and framework’, in *Conference on human factors in computing systems - proceedings*. Association for Computing Machinery. doi: 10.1145/3173574.3173818.
- ‘ShapeRepo: Make your apps interoperable’ (2022). available at: <https://shaperepo.com/>.
- Siegal, J. (2022) ‘Twitter is killing TweetDeck for mac on july 1st and everyone’s angry’. BGR. available at: <https://bgr.com/tech/twitter-is-killing-tweetdeck-for-mac-on-july-1st-and-everyones-angry/>.
- Sonnad, N. (2022) ‘Everything bad about facebook is bad for the same reason’. available at: <https://qz.com/1342757/everything-bad-about-facebook-is-bad-for-the-same-reason/>.
- Teevan, J. B. (2001) ‘Displaying dynamic information’, in *Conference on human factors in computing systems - proceedings*, pp. 417–418. doi: 10.1145/634067.634311.
- Verborgh, R. (2017) ‘Paradigm shifts for the decentralized web’. available at: <https://ruben.verborgh.org/blog/2017/12/20/paradigm-shifts-for-the-decentralized-web/>.
- Weiser, M. (1991) ‘The computer for the 21st century’, *Scientific American*, 265(3), pp. 94–105. doi: 10.1145/329124.329126.