

Understanding and Improving Human Data Relations

Alex Bowyer

Contents

1 Practical Approaches to Improve HDR	1
1.1 Additional Background: Theories of Change (ToC)	2
1.2 Approach 1 to Improving HDR: Discovery-Driven Activism	4
1.2.1 Obtaining Your Personal Data	5
1.2.2 Data Flow Auditing	8
1.2.3 How Activists Pursue Change	9
1.2.4 Data Access & Understanding Services	11
1.3 Approach 2 to Improving HDR: Building the Human-centric Future	11
1.3.1 Life Interfaces	12
1.3.2 Life Partitioning	14
1.3.3 Entity Extraction From Pooled Life Data	18
1.3.4 Life Information Manipulation Capabilities	21
1.3.5 Digital Self Curation & Inclusive Data Flows	26
1.4 Approach 3 to Improving HDR: Defending User Autonomy and Hacking the Information Landscape	28
1.4.1 A Deceptive Landscape	28
1.4.2 The Battle at The Seams	30
1.4.3 Regulating to Free the Information Landscape	33
1.5 Approach 4 to Improving HDR: Winning Hearts and Minds: Teaching, Championing and Selling the Vision	34
1.5.1 HDR Education & HDR Literacy	36
1.5.2 Demonstrating the Business Value of Better HDR	37
1.5.3 Proving the Viability of Human-centric Approaches	39
1.6 Summation of Chapter 9, and of Part One	41
Bibliography	43

1 Practical Approaches to Improve HDR

“The world is working exactly as designed. And it’s not working very well. Which means we need to do a better job of designing it.”—Mike

Monteiro, author of *Ruined by Design: How Designers Destroyed the World, and What We Can Do to Fix It*

In this chapter, I take on the roles of adversarial designer, social activist and technical strategist, drawing upon:

- the findings of Part One [Chapter 6];
- my practical experiences from peripheral projects 7.2];
- the theoretical framing and research agenda of HDR laid out in Chapter 7; and
- the established understanding of key obstacles and proposed strategies to tackle them from Chapter 8.

The objective here is to expand further on the map of the HDR landscape in Figure 8.1 so that we might chart a course towards the desired change. In particular, the focus now is solution-focused, considering the nuts and bolts of how we might begin to tackle those obstacles in pursuit of the HDR objectives [7.7] of data and ecosystem awareness, understanding and negotiability. In the following sections, I will present four different ‘flavours’ of HDR reform. Each section begins with a diagrammatic representation of that approach as a trajectories of change. These diagrams use a model known as *Theories of Change (ToC)* which is explained in [9.1] below.

1.1 Additional Background: Theories of Change (ToC)

To provide a structure for cataloguing the insights conveyed by this section, I use a *Theory of Change (ToC)* framing. ToC is a set of methodologies commonly used by philanthropists, educators and those trying to improve the lives of disadvantaged populations (Brest, 2010); the theories can be used in different ways including planning, participatory design and field evaluation of the effectiveness of new initiatives. There are many different implementations, but common to most of them is a focus on explicitly mapping out desired outcomes (Taplin and Clark, 2012) with a clear focus on who is acting and whether the change being brought about is a change in action, or a change in thinking (Es, Guijt and Vogel, 2015). In this section, ToC theory will be used in a very limited way, not as a methodology but simply to provide a structural frame for proposed changes, as described below. Using ToC to perform evaluation of the effectiveness of proposed change approaches in action in society would be well beyond the scope of this thesis. Nonetheless, this frame is a useful way to map out the different approaches to changing the world in pursuit of the ideal of better HDR.

Figure 9.1 illustrates the aspects of ToC thinking that are important when using this frame. Specifically, desired changes can be broken down into:

²Diagram used here unchanged from *Hivos ToC Guidelines* (Es, Guijt and Vogel, 2015, p. 90) under a CC-BY-NC-SA 3.0 license, whose authors state that this diagram was adapted from earlier work by Wilber (1996), Keystone (2008) and Retolaza (2010, 2012).

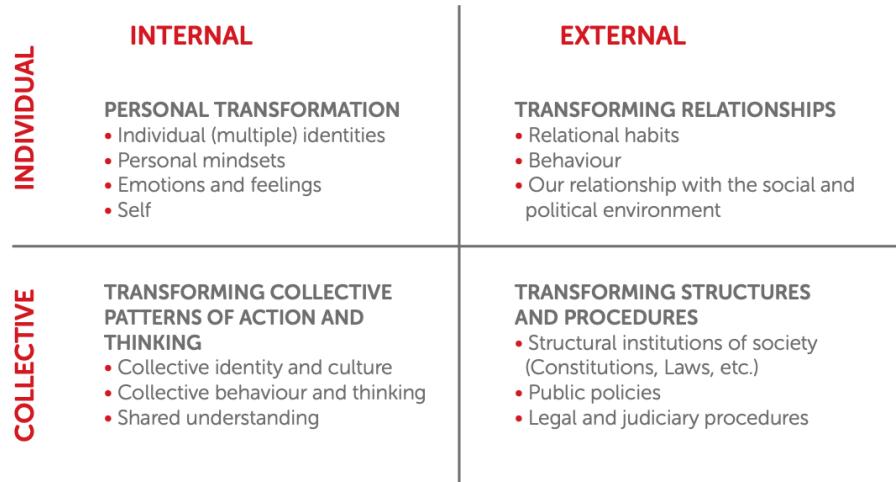


Figure 1: Figure 9.1: Theory of Change [ToC]: The Four Dimensions of Change²

- **Internal changes:** changes in thinking, feeling, reasoning, understanding, attitudes or identity.
- **External changes:** changes in actions, behaviour, interactions, structure, policy, technological capability, processes and the external environment.

At the same time, desired changes can be broken down into:

- **Individual changes:** changes to individual thought or actions
- **Collective changes:** changes to the thoughts or actions of groups of people together, or to the systems, practices and norms of society at large.

These two splits produce four dimensions of change, and form four quadrants representing different types of change, which are shown in Figure 9.1 and described here:

- **Individual/Internal (II):** This top-left quadrant represents changes to what individuals know and understand, and to how they think, feel and plan to act.
- **Individual/External (IE):** This top-right quadrant represents changes to how individuals' relationships with others; acting (or being enabled to act) differently in their daily lives and when interacting within society.
- **Collective/Internal (CI):** This bottom-left quadrant represents changes in the shared knowledge of groups of people or to the collective identity or values of social groups.
- **Collective/External (CE):** This bottom-right quadrant represents changes to the structures and procedures within which people operate, including technology, law, societal norms and communications.

Key to ToC thinking is the idea that making changes in one quadrant can

stimulate change in others; for example, collective learning about data attitudes and practices, such as the research conducted in this PhD, (lower left quadrant) could inform the design of new technologies, interfaces or processes (lower right quadrant), which if built could make new structures available to have an impact on improving individual-provider relationships (upper-right quadrant). The changes to those relationships could then in turn lead to individuals thinking and feeling differently (upper left quadrant), for example feeling more empowered or having greater awareness of data practices. The diagram at the start of each of the following four subsections illustrate that approach's trajectory of change through the ToC quadrants.

1.2 Approach 1 to Improving HDR: Discovery-Driven Activism

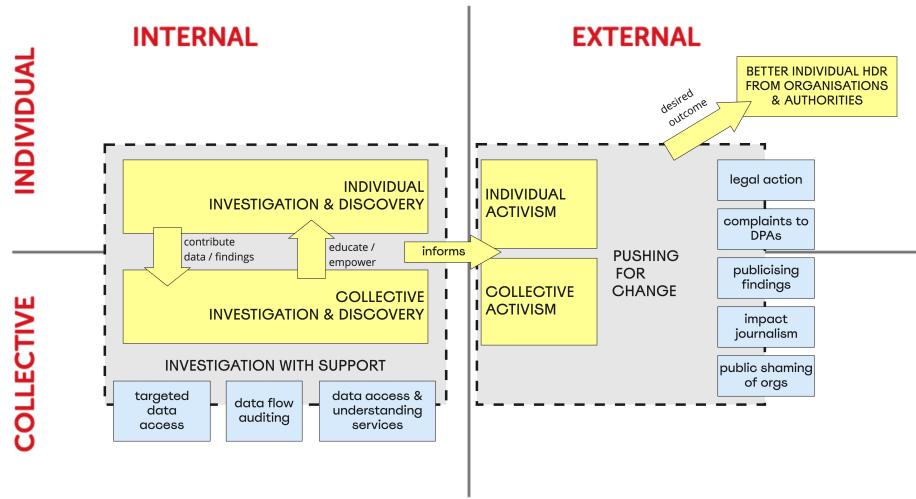


Figure 2: Figure 9.2: HDR Approach 1: Discovery-Driven Activism

The approach to HDR reform presented in this section, and illustrated in Figure 9.2 above, embraces the *activist* aspect of being a recursive public [7.8], as well as the idea of reconfiguring one's world. This approach focuses on the realities of the current data-centric provider ecosystem, and is focused upon deeply understanding it so that it can be challenged from a grounded position of strength. The approach, which would be applied typically to a single service provider, app or platform (which could also be a public sector service) is fourfold:

1. **Gather Evidence:** Using all available means, gather personal data and information from the target organisation.
2. **Determine Current Practice:** Analyse the gathered data and information, using targeted approaches to dig deeper where needed, to establish facts about how the target organisation stores and handles personal data.

3. **Identify Desired Change:** By consideration of the specific practices observed from the analysis, specific problematic practices, gaps or areas where improvement or change is needed are identified.
4. **Appeal for, or Compel, Others to Act:** Using all available means, directly or indirectly persuade or force the target organisation to execute the desired change.

I describe this approach as **Discovery-Driven Activism**. The discovery phase, which aims simply to establish facts about the past or current data practices of the target organisation, can be broad (*'let's see what we find'*, as in the digipower investigation), or highly targeted, such as when *The Citizens* (a non-profit pressure group in the UK who *'use impact journalism to hold government and big tech to account'* ('The citizens - about us', 2020)) used Subject Access Requests to investigate a breach of personal data by the Labour Party in the UK (Colbert, 2022).

1.2.1 Obtaining Your Personal Data

Subject Access Requests and **Data Portability Requests** (Information Commissioner's Office, 2018) are two of the most powerful tools for this kind of investigation. Similarly to Freedom of Information Requests which have used to obtain otherwise hidden data and information from governments and public sector organisations (BBC News, 2014), so these new data access rights are beginning to be used to force commercial organisations to release personal data or information about data processing. There are challenges in non-compliance, as discussed in 5.4.2 and 5.5.1, but the ability for the individual to ask very broad or very precisely targeted questions and to be able to threaten a complaint to a Data Protection Authority if that question is not answered, is a significant new power that HDR reformers can exploit.

INSIGHT 9: Individual GDPR requests can compel companies to change data practices.

In this inset box, I will show, using a mini case-study from my own direct personal experience, how one person can apply the discovery-driven activist approach to compel a multi-billion-dollar international data-centric organisation to improve their HDR.

INSIGHT 9: Individual GDPR requests can compel companies to change data practices.

As an avid user for several years of the music streaming service Spotify who has built up a large library of playlists, I have made a number of GDPR requests to get copies of my personal data. When I was first given a copy of my personal data, I was returned a basic ZIP file including 12 JSON files containing playlists, search queries, account information, my last 12 months of track play history, and inferences about my musical tastes. Spotify also make an extended data download available, including technical log data, and extended play history (which covers the lifetime of my account). I requested this extended download and received a much larger dataset with 175 JSON files, including granular details of when I had used different interface features and the precise details of every song I had ever played. Thinking that I would like to use this data to build a view of my listening history that was not tied to the Spotify platform (in line with the idea of increasing agency by separating one's data from the service that holds it [8]), I examined the streaming history and playlist data with this purpose in mind. What I found was that individual songs were identified only by textual strings of the title, artist and album name. This information is insufficient for a programmer's use - there is no unique identifier or Uniform Resource Indicator (URI) to uniquely identify the specific version and release of a track played. Also without such an identifier, it would not be possible to generate a thumbnail image of the track, or build functionality such as a clickable link to 'play this track in Spotify'.

INSIGHT 9: Individual GDPR requests can compel companies to change data practices.

This highlights a common issue that occurs with data access requests, as highlighted in 5.4.3 - there is ambiguity over whether providers should identify data in a machine-readable way (useful for programming), or in a human-readable way (to optimise understanding). In my case, I needed both. I e-mailed Spotify back and was provided with an alternative file set which contained only Spotify Track URIs, such as `spotify:track:5CKqyYTZqp6Nb4b3kJjUL5`. These met the programmer need to uniquely identify the track, but not the human need - I had no idea which artist or track each of these URIs corresponded to, as there was no human-readable text accompanying each entry. So, I e-mailed Spotify back, making the case that my GDPR rights had not been fully satisfied, because I needed *for each play history entry*, both machine-readable ID and human-readable track title and artist name. I sent Spotify over 30 e-mails on this matter between October 2020 and May 2021. There is little continuity of conversation between support agents, and hard to be escalated to the correct staff with the technical or legal expertise to assist with such nuanced questions. However, by persistently and politely repeating my questions and not accepting No for an answer, I was able to achieve a notable outcome, Spotify **changed the format of their data returns**, not just for me but **for all future customers**. Now, each item in the playback history data you get back from Spotify, every item includes textual track and artist details AND a Spotify track URI. The data can now be understood by both human and machine. The likely interpretation here is that I successfully able to persuade their Data Protection Officers (who handle GDPR requests) the importance of returning data that is both machine-readable and human-understandable. Perhaps they also recognised the amount of work they had invested in supporting my query, and wanted to **avoid having to do such work ever again** should I or any other customer make the same request in future. This was a tiny impact, but a lasting one, and it shows that the discovery-driven activism / civic hacking approach can have an effect in improving HDR with a target organisation.

INSIGHT 9: Individual GDPR requests can compel companies to change data practices.

A larger scale example of individuals forcing giant corporations to change is seen in the case of Facebook. In the early 2010s, Austrian lawyer Max Schrems began to pressure Facebook to disclose more personal data to their users. He created a tool to enable people to make their own data access requests, which over 40,000 people used. Faced with an overwhelming volume of work and massive liability of future data access requests, Facebook was forced to launch the self-service ‘Download Your Information’ (DYI) download tool, increasing transparency for all Facebook users worldwide (Solon, 2012). Facebook was forced to increase its transparency further when Paul-Olivier Dehaye (now CEO of Hestia.ai) made a GDPR request (later backed by legal action) to force Facebook to disclose more information about which advertisers Facebook had enabled to target him using the Facebook Custom Audiences feature. Apparently in order to avoid being embarrassed in court, Facebook updated DYI so that your downloaded information includes a list of advertisers who have added you to a Custom Audience (Dehaye, 2017). Dehaye and Schrems both continue to act as HDR reformers and civic hackers following the discovery-driven activism approach, through their organisations Hestia.ai and privacy rights organisation noyb.eu (‘none of your business’) (Schrems, 2017) respectively.

Facebook’s DYI tool, mentioned in the insight above, represents a useful class of tool in the arsenal of the activist HDR reformer. Along with Google Takeout, it is one of number of *data download portals* that allow users to download their own data. Since GDPR’s introduction in 2018, an increasing number of large online platforms including Facebook, Google, Apple, Netflix, Twitter, Spotify, Uber, Instagram and Strava, faced with the need to reduce the cost impact of GDPR request handling for their large userbases, have developed and augmented online self-service portals available where users can download a copy of their personal data. This has some advantages over Subject Access Requests in that data can usually be obtained within minutes or hours rather than taking up to 30 days, but has some disadvantages in that the data returned is a voluntary offering by the company, that may not cover the data that the individual is seeking and does not provide any ability to ask follow-up questions. This technique was sometimes used as a fallback means to obtain data in Case Study Two, and was used more strategically in the digipower project [ARI7.2], where its merits and limitations are discussed (Bowyer, Pidoux, *et al.*, 2022)

1.2.2 Data Flow Auditing

Both access requests and download portals rely on the organisation in question to be transparent, accurate and thorough in their provision of information, but an alternative technique of *data flow auditing* allows individuals to investigate and collect data on the *actual* behaviour of a target organisation. This was used effectively in the digipower investigation [ARI7.2]. Using an Android app

called TrackerControl (Kollnig, 2021), a service provider's app can be monitored while the user is using it normally, to see which servers or domains that app is contacting (and one can imply, exchanging data with). Apple has recently introduced an equivalent function on iOS known as App Activity Reports (Apple, 2022), providing iPhone users with the same ability as part of the phone's operating system. This has limitations in that the content of the data exchanges is not known, but can serve as a valuable tool to verify claims made in privacy policies or GDPR responses, and also as a means to *generate questions* for further investigation, for example by identifying third parties such as data brokers which the target organisation may be sharing personal data with. This technique is described further in (Bowyer, Pidoux, *et al.*, 2022), along with a comparison of the different techniques of data flow auditing, data download portals and data access requests.

In general, what the discovery-driven activism approach highlights is that there is a role for pro-active citizens to play in challenging the power of data-holding organisations by treating those organisations as a subject of investigation, both in research (Walby and Larsen, 2012) and in the pursuit of improving civic society (Schrock, 2016).

1.2.3 How Activists Pursue Change

Once information has been obtained, the HDR reformer activist can use a variety of means to try to bring about the desired change:

- If a target organisation fails to comply with a data access request, or a demand to erase or correct data, they can be *reported to the appropriate Data Protection Authority*. In some cases even the threat of this (which can carry a large fine) can be enough to compel the organisation to change.
- If a breach of law is found, the target organisation could be *taken to court*, as seen in the Schrems case above, which resulted in new legislation that Facebook had to comply with (Kuchler, 2018). As well as individual cases, this also often happens in the form of class action lawsuits, as with Facebook and Cambridge Analytica (Bowcott and Hern, 2018). Increasingly, unethical or illegal data practices are being challenged. In some cases, such extreme measures are not needed.
- Simply *making data available to the public* can be empowering to society at large. This approach has been demonstrated by the UK website TheyWorkForYou, which increases democratic accountability of MPs by making MP's votes and public statements more readily accessible (mySociety, 2004).
- As well as structured *impact journalism* such as that conducted by The Citizens as mentioned above, another technique available to individual activists is public shaming of misbehaving organisations**, especially on Twitter. While the ethics of this are complex and it does not always succeed, the technique has been used effectively to force organisations to change, in order that they might avoid further bad publicity (Silver, 2014; Braw, 2022).

INSIGHT 10: Collectives can compare and unify their data and use their pooled knowledge to demand change.

Increasingly, the Internet experience that individuals experience is not the same as anyone else's. Thanks to recommendations, targeted ads and social media feeds personalised to your interests, no two people will see the same digital reality. This means it is very difficult for regulators or individuals to hold digital service providers to account. In recent years, many activists have embraced **the power of collectives**, and realised that together, they can discover far more than they can alone.

An example of this is the WhoTargetsMe project, launched in 2017 (Jeffers and Webb, 2017). The objective of this project was to monitor political advertising in the UK. Recognising (as larger studies have shown (Bakshy, Messing and Adamic, 2015)) that everyone was seeing different advertisements, the goal was to have each individual report what adverts they see on Facebook, so that these can be pooled and compared with others. Over 50,000 people participated, building up an otherwise unavailable clear picture of the ways in which different political demographics were being targeted. This is a powerful mechanism available to collectives in this space: the ability to have individuals obtain their own datapoints and then compare them.

Another example is seen in the Worker Info Exchange ('Worker info exchange', 2022), a collective that helps gig economy workers such as Uber drivers and Deliveroo riders to make data requests. Using the pooled data, they conduct investigations to understand algorithmic inequalities and identify unfair treatment of worker by employers. They then help those workers to fight for better working conditions, much like a traditional trade union, but powered by collectively-sourced data. This resulted in Uber being taken to court, and some gains being made for drivers (Lomas, 2021; Foucault-Dumas, 2021).

As the aforementioned case with Max Schrems showed [Insight 9], collectives can be particularly powerful when exerting their data access rights *en masse*, and this can improve HDR and force greater transparency. René Mahieu and Jef Ausloos have published an exhaustive list of collective actions taken using GDPR rights, addressing issues such as discrimination by US colleges, corporate surveillance of climate activists, identifying gaps in data disclosures, and manipulation of users on dating apps (R. Mahieu and Ausloos, 2020). The authors identify that the GDPR provides an *architecture of empowerment* and have called for better enforcement and for European authorities to provide better support for the ability for collectives to make data access requests together (R. L. P. Mahieu and Ausloos, 2020). Hestia.ai's digipower investigation concluded that data-discovery driven collectives are a vital step on the road to a more digitally empowered society (Pidoux *et al.*, 2022, p. 70). It is clear that organised collectives exploiting data access rights represents a powerful vector for impactful discovery-driven activism.

1.2.4 Data Access & Understanding Services

Having< identified that there is a clear trajectory where individuals and collectives can obtain data to empower them, it is clear that this complex work can be supported. We see the emergence of what I would call *data access & understanding services*, with entrepreneurs and activist enthusiasts:

- creating *tools to help people get their data back* from organisations, such as Tap My Data ('Tap my data', 2022), AccessMyInfo ('Access my info', 2022) and My Data Done Right ('Privacy salon & my data done right', 2022);
- creating *tools to help people understand their personal data*, such as Ethi (Jelly, 2021) and digipower.academy ('About digipower.academy', 2022); and
- helping collectives and journalists with *training, workshops and skills development* to take advantage of their data rights and *understand data ecosystems*, like Hestia.ai (Dehaye, 2019).

If such emergent endeavours can be supported and enabled to flourish, that could make HDR reformers using the discovery-driven activism approach more successful by ensuring that a lack of legal, technical or investigative skill does not become a barrier to any HDR reformer wanting to use this approach.

This approach shows that there is a role for independent actors and organisations to carry out discovery-driven activism - access requests, complaints, legal challenges, public campaigns and more. Discovery-driven activism can empower individuals and collectives to incrementally work towards building the world of better HDR that this thesis outlines.

1.3 Approach 2 to Improving HDR: Building the Human-centric Future

The approach to HDR reform presented in this section, illustrated in Figure 9.3 above, focuses on the gaps in individual data interaction capability that exist today. The objective here is to design and build proofs of concept for **novel human-centric information systems** that can deliver people **new capabilities over their data**. In this approach, the focus is more introspective than Approach 1: it is about how the individual can improve their relationship with data in the context of their own digital life. The bulk of this section describes **specific design ideas** developed by myself and colleagues at BBC R&D during my 2020-2021 research internship on the Cornmarket project. As established in Insight 2, one of the most promising models for giving people a new and improved relationship with their data is to create a place where one's scattered (Abiteboul, André and Kaplan, 2015) personal data can be stored and aggregated in one place (Jones, 2011), and based on Insight 3, that place should offer individuals the means to use their life and ecosystem information as a material they can explore, examine or repurpose as they see fit. This leads to

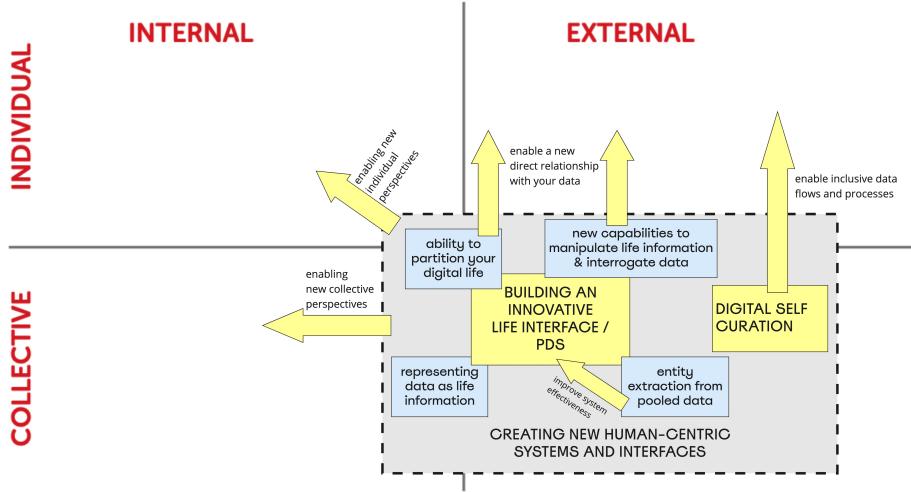


Figure 3: Figure 9.3: HDR Approach 2: Building the Human-centric Future

the vision of a Personal Data Store [2.3.4] or into which one can unify the data from the different parts of your digital life (as depicted in the AllOfMe vision video ('AllOfMe.com Teaser Clip', 2008)), and meet public demand for 'control over your data' (Teevan, 2001; Hartman *et al.*, 2020). The Cornmarket R&D project therefore sought to develop a human-centred (i.e. non-commercial) PDS proof of concept, as shown in the conceptual model I developed for the BBC Cornmarket project depicted in Figure 9.4.

1.3.1 Life Interfaces

The first challenge in designing such a system is to consider what data will be stored, and that data can be represented as life information. At a high level, I identified a number of different types of data that a user might wish to store in a PDS:

Then, considering the earlier observation by myself (Bowyer, 2011) and proponents of temporal PIM systems [2.2.2] that time can be used as a unifying concept for personal information, I considered how we might represent occurrences within personal data as *happenings*, anchored against a particular point in time. In effect, this would focus on different properties of the data, much like Karger's *lenses* [Karger *et al.* (2005); 2.2.2].

In modelling data as life information using the concepts shown in Figure 9 and Figure 9.6, it is important to come up with as simple a model as possible, so that the *life interface* is not overwhelmingly complex and unmanageable. Over several iterations, I was able to reduce the modelling of life information in a PDS down to the following four types - *activities*, *interactions*, *transactions* and *datapoints* [Figure 9.7]. These could correspond to four views or lenses within

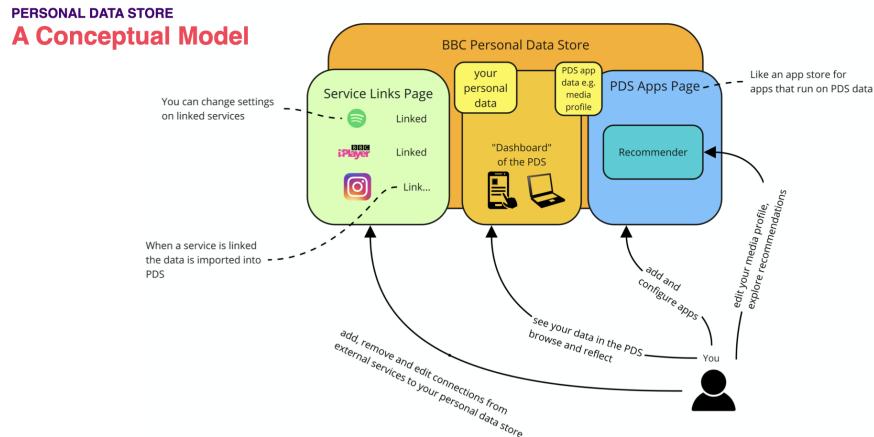


Figure 4: Figure 9.4: Conceptual Model for a Personal Data Store System

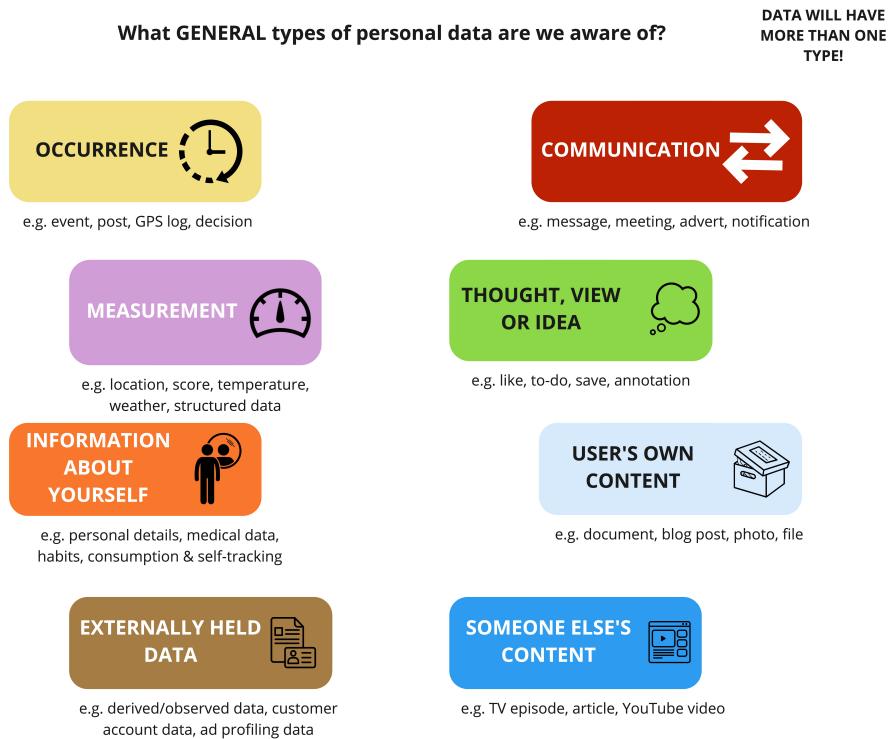


Figure 5: Figure 9.5: High Level Data Types

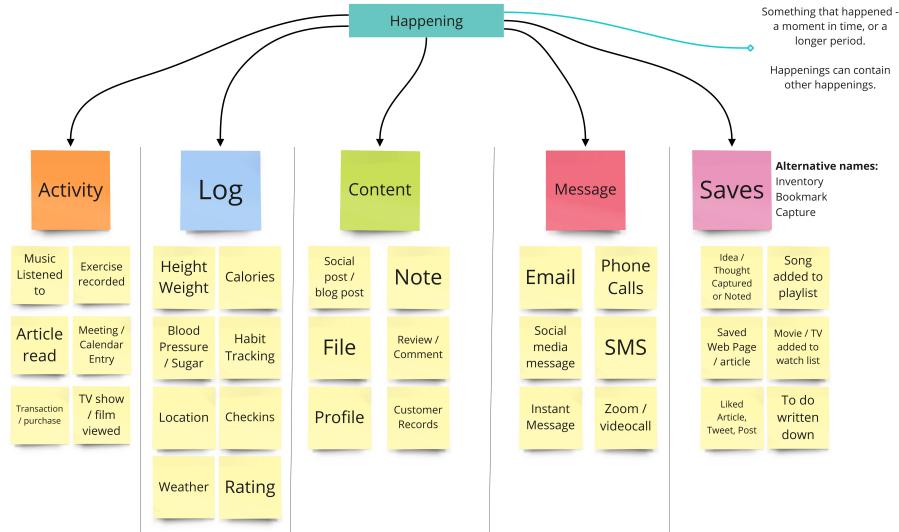


Figure 6: Figure 9.6: Life Information Modelled as Happenings

the user interface.

Having decided upon models for the information within a PDS, it becomes much easier to design visual presentations of that information. Figure 9.8 shows a design mock-up designed by Alex Ballantyne of BBC R&D of some of the views of life information that one could offer within a PDS:

1.3.2 Life Partitioning

What has been shown in Figure 9.8 is the beginnings of imagining what a *life interface* might look like, a single holistic interface covering all aspects of one's digital life.

Previous HCI design work [2.3.1] has focused on, in the first wave, the functionality of the machine; in the second wave, on the common tasks of a work environment; and in the third wave, on classes of users and the commonalities of experiences in everyday life. But the challenge designers of life interfaces face is even more difficult. As Lindley noted, no single arrangement of information will suffice because in the same context two people may have different needs [Lindley *et al.* (2018); 2.2.2]. Because of the *subjectivity principles* [Bergman, Beyth-Marom and Nachmias (2003); 2.2.2], we are now faced with the need to build an interface that is suited to the individual, even though **every individual has different needs**. The design requirements for the interface are likely unique to the individual. Therefore, any life interface design must be able **support different mental models** - and as the life sketching exercise [Figure 5.2] in Case Study Two showed, people have many different ways of compartmentalising

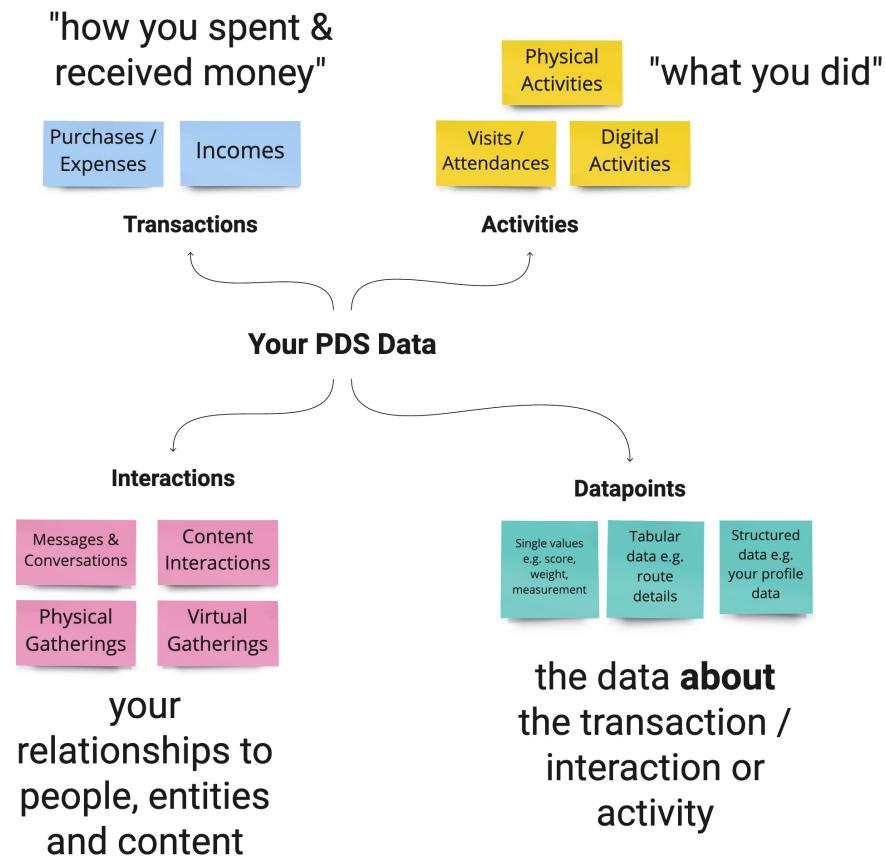


Figure 7: Figure 9.7: A Simple PDS Life Information Presentation Model

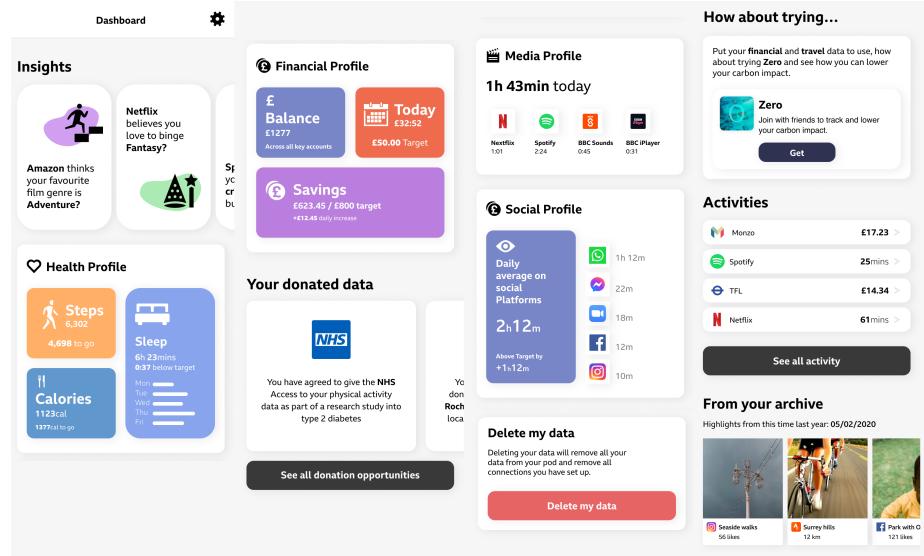


Figure 8: Mock-up of Life Information Presented in a PDS Interface

their life.

During a life interface design hack week at BBC R&D, colleague Jasmine Cox and I explored this problem. We considered that a key element of the interface design would need to be the versatility to *partition one's life in different ways*, and then to be able to assign different collections of life information to the different partitions. As a visual illustration of one possible mental model for life partitioning, consider Figure 9.9, where, based on an idea I had had years earlier, I modified a CluedoTM board to represent the 'rooms of your digital life':

We also imagined other mental models, such as partitioning your life according to parts of the body (heart for relationships, body for fitness, brain for current projects, etc.) or a landscape with forest, lake and buildings representing different things.

Whatever the visual metaphor engaged, the important thing is the functionality - being able to use these partitions to *filter* [2.1.4] your life information and focus on a particular *perspective* [Lansdale and Edmonds (1992); Krishnan and Jones (2005); 2.2.2]. Deciding to focus on a particular aspect of one's life is analogous to fixing a *conceptual anchor* [Teevan (2001); 2.2.2].

We then produced a mock app workflow for assigning different elements of your life (such as people, places or topics) to different partitions of your life, and viewing a visualisation of those different partitions in some subjectively meaningful form before filtering on a particular life partition and then being

³Cluedo board design is a copyright of Hasbro, Inc., fair use applies.

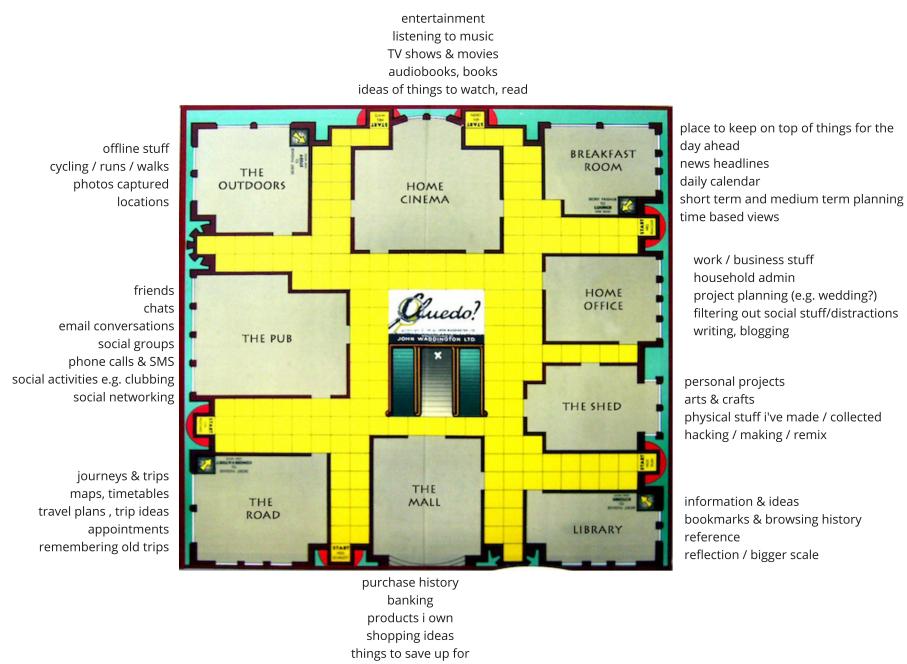


Figure 9: Figure 9.9: Life Partitioning Analogy using a Cluedo™ board³.

presented with a *timeline* [2.2.2] of life information pieces associated with that part of life. This is illustrated in Figure 9.10, with artwork by Jasmine Cox:

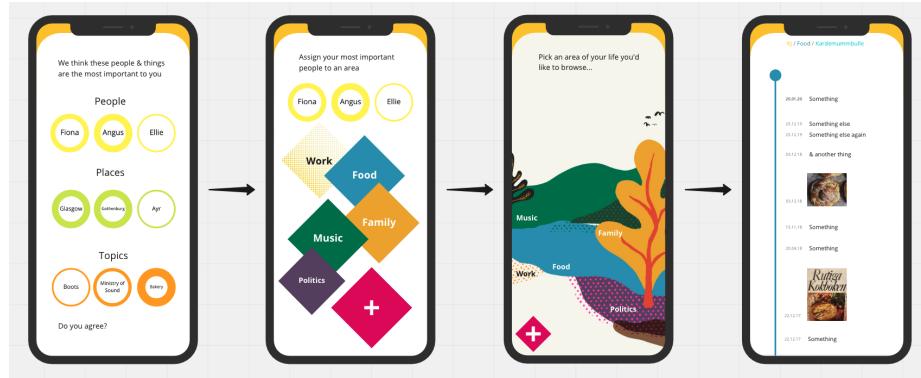


Figure 10: Figure 9.10: Mock-up: Browsing by Areas of Life

1.3.3 Entity Extraction From Pooled Life Data

INSIGHT 11: Automating the Identification of Entities can enhance Machine Understanding and Unburden Life Interface Users.

Having identified the need to assign every piece of a user's life information to a particular partition (or multiple partitions) of their life, it quickly becomes apparent that this would be **too much work for the user** to do alone. Systems that use manual categorisation and tagging to classify information work best with a large userbase to contribute effort to the classification operation (Golder and Huberman, 2006). As part of the explorations of PDS approaches at BBC R&D I therefore also examined how this challenge might be addressed (considering also that having a lot of effort could be a deterrent to adoption [Objective 5]). I identified an approach that could help with this problem: If the *entities* (for example, a person, a place, an event or a topic) associated with a piece of data can be identified, then a lot of the **assignment of data to life partitions can be handled automatically**. For example, association with your office location would indicate that any data associated to that location is likely to relate to the 'work' part of your life, and this could be done automatically, reducing the effort for the PDS user. The process of **identifying entities within data**, known as *entity extraction* or *named entity recognition (NER)* is a well-established technique, which relies on the trained recognition of proper nouns and keywords combined with the statistical analysis of sentence grammar (Marshall, 2019). This technique is used extensively in text-mining products within the Content Analytics industry such as those produced by my former employer, OpenText ('What is text mining and content analytics?', 2022). However, in the context of a PDS, I propose that new techniques can be applied, making use of the data touchpoints into different parts of an individual's life to identify entities relevant to them personally (including, for example, names of friends or private projects that a standard NER solution would not detect). Data is full of references to entities that have personal relevance in your life. Finding these allows meaningful metadata to be attached to each datapoint. Figure 9.11 shows how a large number of entities could be detected from different parts of an individual's data once it has been imported into a PDS environment:

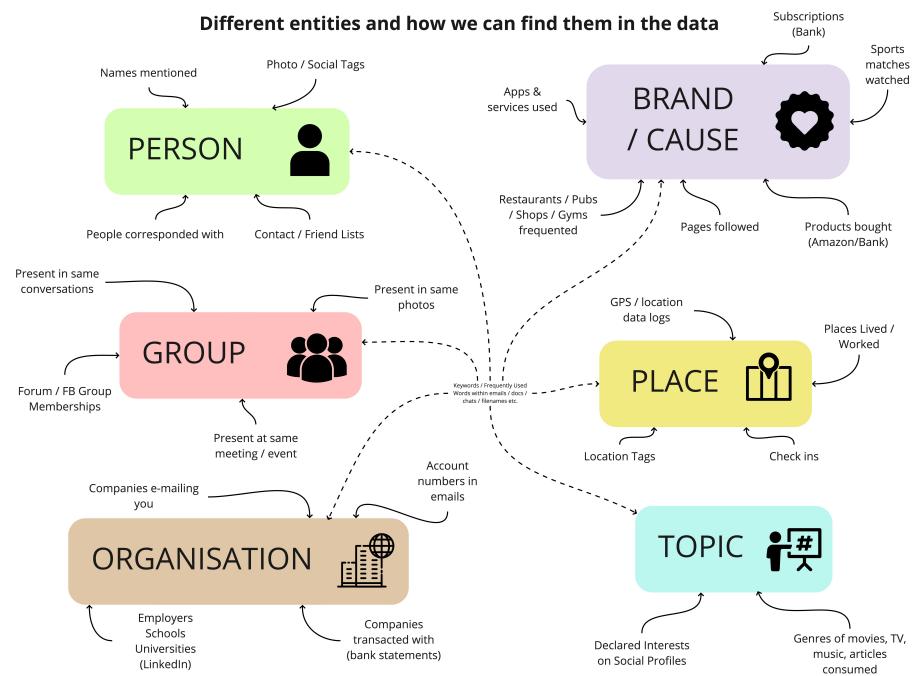


Figure 11: Figure 9.11: Identifying Entity Associations in Data

(continues...)

This sort of approach could be quite powerful in reducing the effort for life interface users. By scanning the data, the most prevalent entities could be identified, and the user need only assign the entities to different parts of their life, as illustrated in the first two frames of Figure 9.10. This would then allow hundreds of associated data points which had been programmatically associated to that entity, to be assigned to the correct ‘bucket’ or life partition. While such an approach would not be perfect, and there would need to be some corrections made by the user, this is far preferable to them having to provide all the classifications and is likely to motivate greater engagement. I have observed in user experience design and consideration of productivity systems that users are more motivated to correct errors, than to fill in a blank page. Philosophically, we are moving here towards a *learning system*, a system that can be told when it is right and when it is wrong, and get better at classifying things correctly, analogous to the way an executive might train an assistant to anticipate his/her needs better, a sort of *digital life assistant* (Bowyer, 2018a). Bayesian classification techniques could also be used to help with the learning here (Authors, 2022b). . This approach is also useful for *ecosystem detection*, as outlined in Insight 4 - as identification of relationships with external entities is a key first step to mapping a user’s ecosystem.

Insight 11 offers a practical, theory-informed approach as to how we might start to build systems that are more able to understand the meaning of human information [Insight 8]. This is a technique that is already being used by corporations. For example, Facebook has developed an internal system called *word2vec*, shown in Figure 9.12, through which it mines information about the world from the data that people generate through social media posts, messages and interactions. This information is exploited for commercial benefit. It is about time similar techniques were used for individual benefit and personal data empowerment.

Automatic ‘best guess’ attempts to arrange personal data in a PDS can go further than just association of related entities. In many cases, it is possible to programmatically determine the nature of a piece of data. Figure 9.13 shows a detailed approach for identifying and classifying data automatically in a PDS, drawing on the previously identified data types in Figure 9.5, the data attributes in Figure 9.15, and the entities identified in Figure 9.11. In Figure 9.14, a more abstracted approach is shown looking at how data could be classified according to the four different life information concepts in Figure 9.7.

1.3.4 Life Information Manipulation Capabilities

Moving beyond the initial challenge of classifying and arranging human information within a *life interface*, a core consideration in designing such an interface is to consider what capabilities the user might be given over the pieces of life

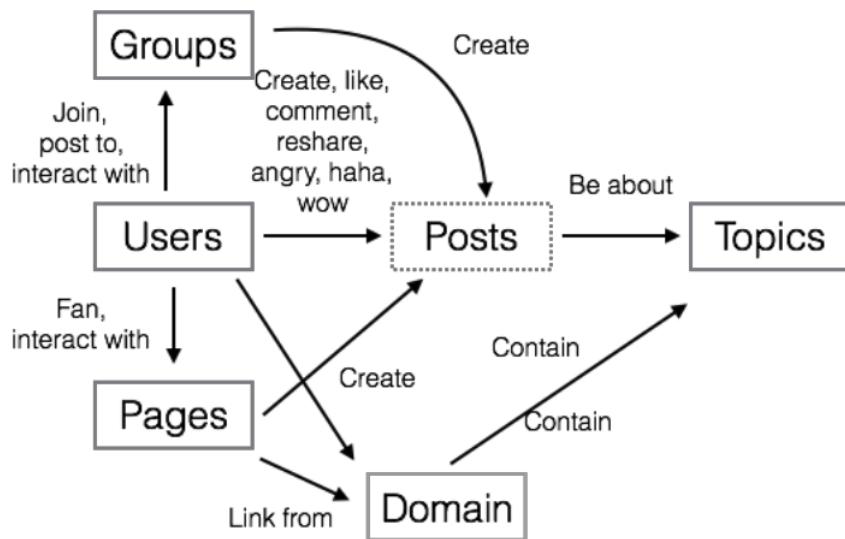


Figure 12: Figure 9.12: Facebook's World2vec Model, Semantically Modelling Human Information from Social Media Posts on Facebook (Peysakhovic and Wu, 2020)

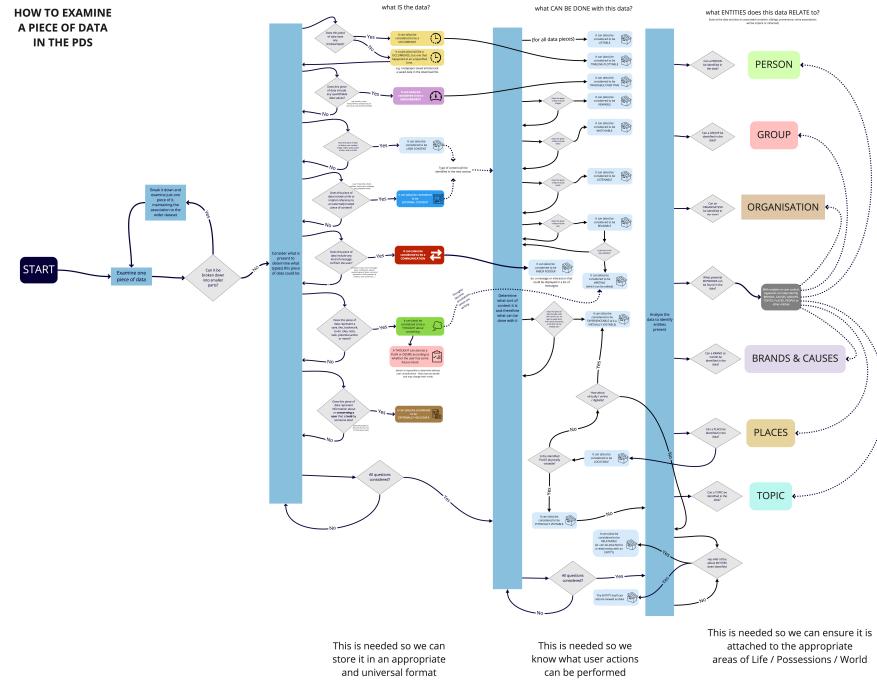


Figure 13: Figure 9.13: Identifying the Attributes of Data

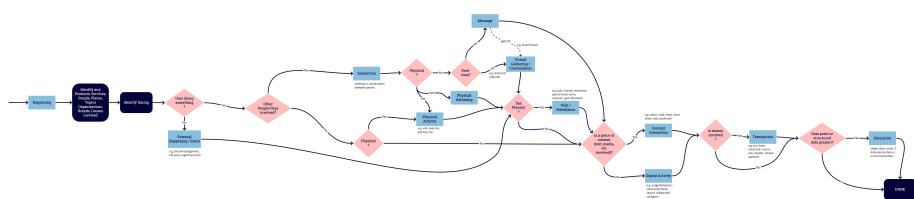


Figure 14: Figure 9.14: Determining the Nature of a Piece of Data

information represented in the interface. It is through the provision of a wide range of useful operations upon information items that the information will start to feel like *a material* as described in Insight 3 and deliver the *new capabilities* needed for success Insight 7.

There are two ways to consider this problem - one starting with the data, and another starting with the user.

Starting with the data, and considering the different types of data that a user might be able to gather from across their digital life, I identified that each datapoint or dataset is likely to have properties which determine *what can be done with it*. A number of possibilities are shown in Figure 9.15:

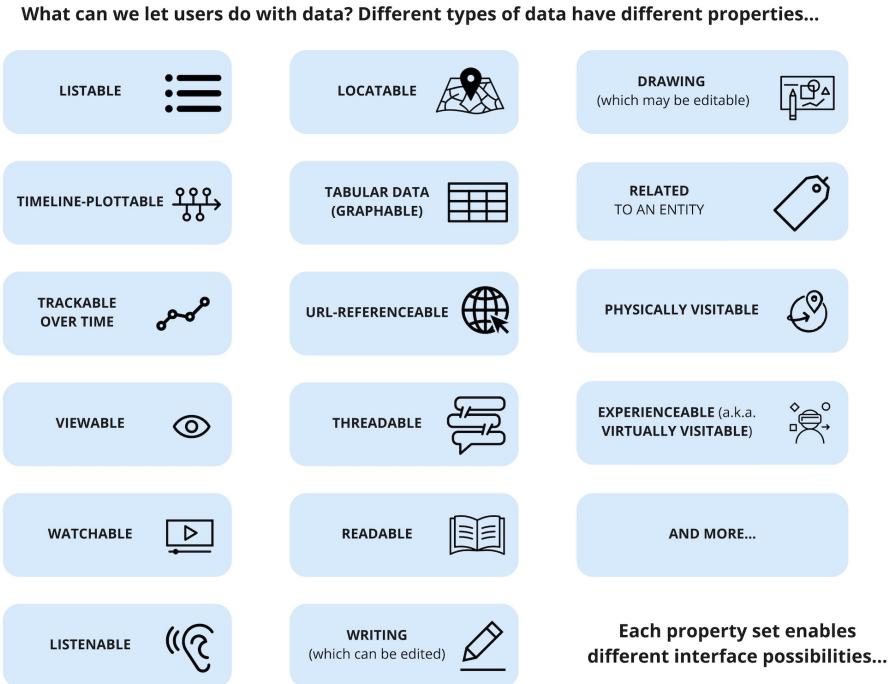


Figure 15: Figure 9.15: Attributes of Data

These attributes could then be used to determine which datapoints would be eligible for inclusion in different views (or *lenses* (Karger *et al.*, 2005)) of the data, such as tables, graphs, maps, lists, media thumbnail views, etc.

If we consider capabilities from the user's perspective, it is useful to recognise that visualisations alone (such as those in Figure 9.8 are not sufficient to meet the need for either interrogability or malleability [8.2]. To address the need for malleability, we need to determine what actions an individual might want to perform on a piece of life information. To explore this design question, we can

draw on multiple sources of inspiration:

- actions identified in prior PIM research & design work [2.2.2] such as arranging, browsing, keeping, and associating;
- actions needed for SI [2.2.3] such as combining, history viewing, interpreting patterns;
- actions needed for effective access [2.1.4] such as filtering, zooming, getting detail on demand; and
- actions inspired by the eight lenses of how people think of data [Table 7.1] such as summarising, deriving or remixing.

Having developed a large set of initial design ideas for possible actions, I worked with my BBC colleague Chris Gameson to distil this information design thinking into a 12 *data card* designs that could encapsulate some of the most common *actions a user might want to perform* upon personal data in a PDS. These are shown in Figure 9.16. These had a similar purpose to the cards in my pilot study [1.3.1] and Case Study One [4.2.1], acting as ‘*things to think with*’ (Bowyer *et al.*, 2018).

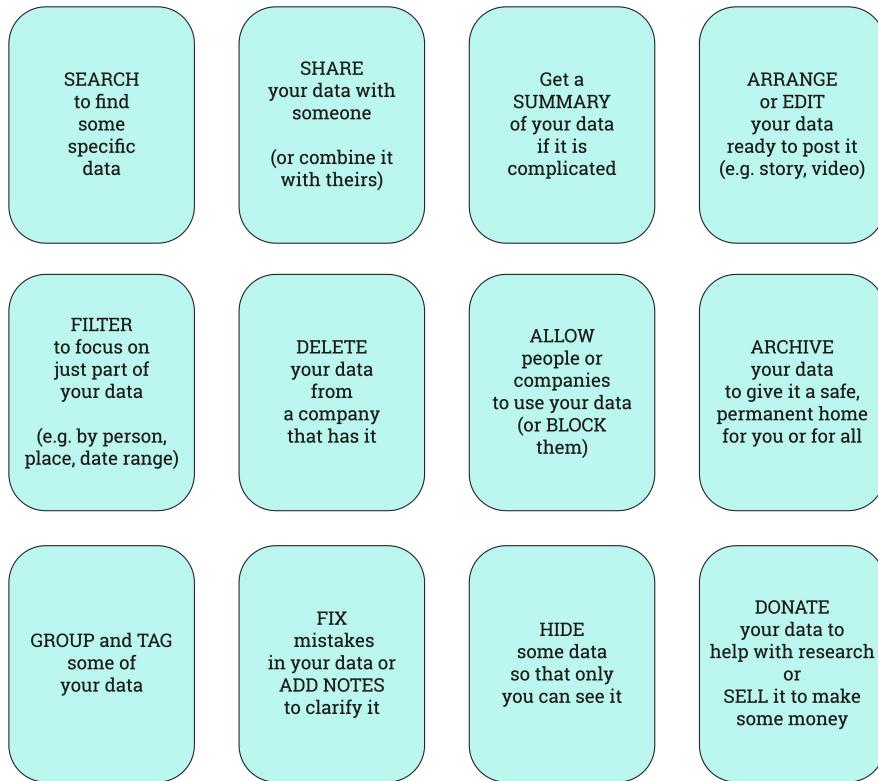


Figure 16: Figure 9.16: Actions One Might Perform on Life Information

As identified in chapter 5 [5.3.3; 5.4.2] and explored further in 6.1.3, being able to ask questions of data is important to individuals. Addressing this need for interrogability in a similar approach, we produced 10 data card designs to encapsulate some of the different common questions that users might want to ask of their data, as shown in Figure 9.17.

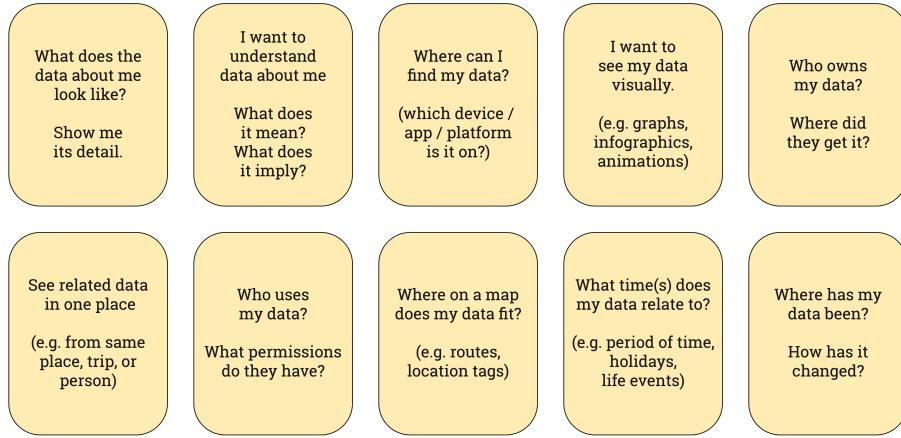


Figure 17: Figure 9.17: Questions One Might Ask of Life Information

These actions and questions act as a design building blocks, a precursor that can inform the design of features for a life interface that could provide the user with new capabilities to interrogate and manipulate the data in a much more interactive way.

Once information has been correctly identified semantically (using approaches such as those in Figure 9.13 and Figure 9.14), and associated to entities, it then becomes possible to enrich users' search and browse capabilities by enabling the use of *facets* {Teleciciute (2020)} to allow users to precisely target a set of data points or a particular data item. An example of how this could be done is shown in Figure 9.18, another design artefact from my work at BBC R&D. This is not a user interface mock-up but rather a three-stage strategy, with examples, for the different types of facets that one could offer a user to help them explore their data in powerful ways:

1.3.5 Digital Self Curation & Inclusive Data Flows

A final element of Approach 2, which, unlike the previous subsections, has a broader focus than just the creation of an effective life interface because it has sociotechnical/business process design implications too, aims to address the problem identified in 8.4 - the intractable data self. Drawing upon the philosophy behind VRM [2.3.4], we can imagine (as implied by both Case Studies [4.3.3; 5.5.2]) that were individuals able to **create or contribute to the digital**

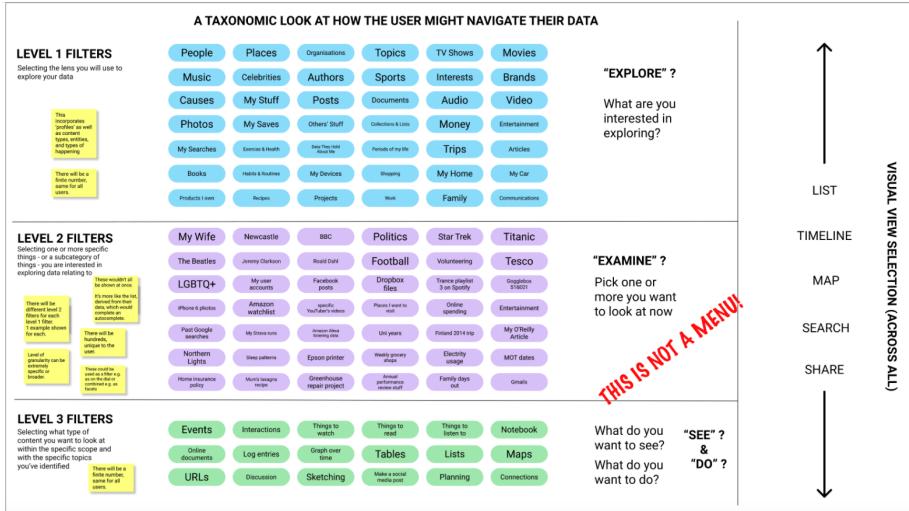


Figure 18: Figure 9.18: Example Taxonomies for Life Information Navigation

representation of themselves in data, that this would be more accurate. Furthermore, this would provide an opportunity to address the identified HDR want of *involvement* [6.2.3]. In the BBC Cornmarket project, this was explored through the concept of profile-based recommendations. Functional interfaces were built which allowed users to construct and edit a representation of their own musical/media tastes known as a '*media profile*', using personal data such as viewing/listening history imported from multiple sources as inputs. This profile would then be used to recommend new content, which would be likely to be better suited to the individual's tastes since it would be based on a more accurate representation of self. This aspect of the Cornmarket project (Orphanides, 2021) is not only a practical manifestation of the ideas of pull-centric/VRM approaches, but is also illustrative of a new way of thinking about personal data, where it can be **directly generated (or edited/fixed) by the individual**, rather than just asserted on inferred by the service provider without verification.

Such thinking could be expanded further to accommodate the idea that data is not static. Once we make the mindset shift from data as a static bundle of facts to **data as an ever-changing flow or stream** [2.2.5; 4.3.3; 6.2.2], it becomes much easier to think about how processes and interfaces might be updated to give users a direct role in the evolution and shaping of that data. This also makes it easier to maintain knowledge of data's provenance [Insight 5]. Such **inclusive data flows** and processes could yield benefits for users in terms of increased ecosystem negotiability Objective 4, but also for businesses, as will be explored in Approach 4.

The future-centric designs and insights⁴ I have presented as part of Approach 2 show that while developing interfaces and processes for a better HDR future is a challenging design problem (due to the variety in individuals' mental models and the breadth of data across an individual's digital life), it is certainly possible to tackle these challenges. It is likely that the reasons that few have ventured into this space are largely due to the lack of an obvious business model or incentive rather than any question of the merits of these approaches. It is possible to **empower individuals with new capabilities** over the life and ecosystem information encoded within their data, and even to design **new data interaction approaches** that give users an ongoing role in the curation of that data, and HDR reformers can help make a world of better HDR a reality by **building, or investing, in the building of these more life-centric technologies and processes**.

1.4 Approach 3 to Improving HDR: Defending User Autonomy and Hacking the Information Landscape

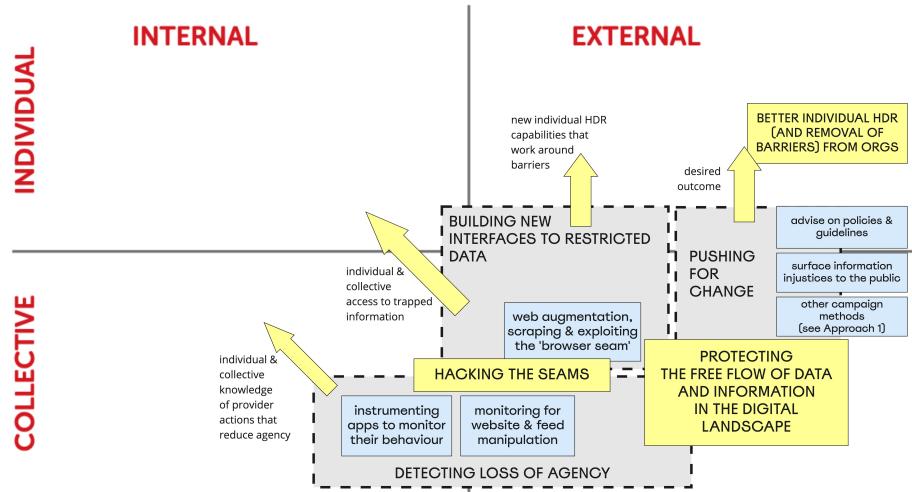


Figure 19: Figure 9.19: HDR Approach 3: Defending User Autonomy and Hacking the Information Landscape

1.4.1 A Deceptive Landscape

While Approach 2 focuses on designing a better future, and Approach 1 focuses on identifying and campaigning for positive changes in provider practices, there is a need for a more direct approach to tackle an urgent and evident problem head on: the active diminishing of users' agency [8.4]. In pursuit of mass market appeal

⁴The research and design work and insights from my time with BBC R&D's Cornmarket project are documented in more detail in the external publications detailed in 1.3.4).

and cost reduction, there is a trend towards the *dumbing down* of technology, which invariably leads to fewer features and less agency. Increasingly, companies encourage users to think of technology as a black box, which you cannot and should not look inside, as depicted in Figure 9.20, and discourage users from varying their usage, digging deeper or challenging the normal way of using a product:



Figure 20: Figure 9.20: The Modern ‘Black Box’ View of Technology

This disempowering trend is seen across hardware, software and service industries, where providers attempt to **restrict the ways in which technology can be used** even after it has been purchased - cars that can only be serviced by official dealers, printers that can only use officially branded ink cartridges, smartphones that cannot be repaired without special tools and warranty invalidation, media files that cannot be copied or shared, online content that cannot be accessed in certain geographies, operating systems that block or discourage certain software from running, and banking applications that forbid screenshots. Apple, for example, encourages users to consider technology as ‘magical’, rather than as tools to be harnessed and understood. This approach is highly problematic for user agency:

“Magical design prioritises pleasing and surprising a passive user who can only use the solution as authorised.”—Cristiano Storni (Storni, 2014)

In order to challenge this trend towards disempowered, compliant users and **protect individual autonomy** and our ability to make sound judgements and hold organisations to account, direct activism and grassroots civic action is called for. This forms the core objective of this approach, which is represented in Figure 9.19 above: to **identify ways in which agency is being diminished**, and to find ways to **circumvent the new barriers** in order to maintain existing levels of capability.

It is bad enough when a particular technology blocks you from doing what you want to do, but the trend is even worse when viewed at a broader level: As the digipower investigation highlighted [Insight 6], the information landscape that

we inhabit as digital citizens is being manipulated to **change what we see, what we understand of the world, and what we can do, in order to manipulate our behaviour**, which can have harmful effects on democracy, agency and individual autonomy. Therefore a further goal here is to protect our ability to **experience unbiased and unmanipulated information landscapes**. A recent US Congress investigation shows how the big five use their dominant positions as device manufacturers, search engines or major Internet platforms to promote their own interests (Sisco, 2022). In a paper looking at the growing manipulation of search results, Shah and Bender consider how this landscape could be protected and accountability preserved. They describe their approach as establishing '*guardrails for the status quo*', and I find this phrase to be an excellent summary of what Approach 3 is trying to do (Shah and Bender, 2022). At the time of concluding this thesis, August 2022, this issue of the unreliability of the social media landscape has again come to public attention, as former Twitter head of security Peiter 'Mudge' Zatko revealed his grave concerns for the security of our information infrastructure and landscape:

"Your whole perception of the world is made from what you are seeing, reading and consuming online, and if you don't have an understanding of what's real and what's not, that is pretty scary."—O'Sullivan, Duffy and Fung (2022)

<https://t.co/QU823RBnN1>

1.4.2 The Battle at The Seams

We can also take this idea further through the idea of **adversarial design** (DiSalvo, 2012) which advocated **using design processes to challenge the status quo**. To understand how to gain traction here and carry out '**design after design**' (Storni, 2014), it is important to understand the concept of **seams**, how they can be exploited, and the risks of losing them. This concept is explained in Insight 12.

INSIGHT 12: The ‘Seams’ of Digital Services need to be Identified, Exploited and Protected.

As we have already identified [8.4] product design (be it hardware or software) is political. Designers pass some power to the user through their design, but also, users should be able to take some power on their own terms. This is the case made by Cristiano Storni in his 2014 paper on '*the politics of seams*', Cristiano Storni identified the idea of '**empowerment-in-use**' which advocates the idea that people need to appropriate their technologies to different users than the designers may not have foreseen (Storni, 2014), something that current black box, limit-what-the-user-can-do thinking blocks. Central to this capability is the concept of **seams** - those **exposed areas which the user is free to change**. This concept was proposed by Mark Weiser and developed by Chalmers and others (Weiser, 1994; Weiser and Brown, 1997; Chalmers, MacColl and Bell, 2003). The changes such as closures of APIs or removal of ports as described in 8.4 can be seen as the removal of seams. As Storni highlights, the availability of design seams is a **critical determiner of user power**, and companies gain power and reduce agency when they remove those seams. It follows therefore that by identifying, exploiting and protecting the seams of digital services and devices, user autonomy and the viability of data-unification efforts can be protected.

An unseen battle is for the free flow of information is underway at the seams of today's digital products. Hackers, civic activists and makers seek to repurpose and exploit the edges of products for their own means, while digital service providers and platforms try to block such activity. For example:

- A successful tool called Findings allowed people to clip and share their favourite quotes from Kindle books. Amazon blocked and banned this tool, and the company shut down (Owen, 2012; Maldre, 2012).
- Louis Barclay created a tool called Unfollow Everything, which allowed Facebook users to automatically unfollow all friends and pages, in order to give them greater control of their News Feed reading experience and avoid being manipulated into reading more than they want to. He was banned for life from Facebook and threatened with legal action should he ever build any tools that manipulated the Facebook experience (Barclay, 2021).
- Various activist groups have for several years been fighting to give individuals the legal **right to repair** their own products (Miller, 2021), which has often been blocked through planned obsolescence, inaccessible seams or restricting access to parts. The problem has been described as **device tenancy**, the idea that our relationship with our technology products is more like a tenant, where a landlord retains overall control and permits us to perform certain activities (Tufecki, 2019). New laws have been introduced in the EU (Tett, 2022), forcing companies to support individuals to repair their devices. Apple has subsequently released self-service repair kits, though these themselves force parts to be paired with particular phones, limiting the utility of self-repair (Moore-Colyer, 2022).

INSIGHT 12: The ‘Seams’ of Digital Services need to be Identified, Exploited and Protected.

- As detailed in my co-authored paper with Louis Goffe and colleagues (Goffe *et al.*, 2021, 2022), ***web extensions*** and ***web augmentation*** offer a powerful technique for modifying web experiences and repurposing user interfaces. This is because once a website is loaded into your browser, it is no longer under the control of the remote site, and by creating a *web extension* to run code within your local web browser that loaded website can be edited, *scraped* (P., 2021), or otherwise repurposed. This has been successfully used to stop clickbait, dispute fake news, combat addiction, filter explicit words and more. However, in order to re-assert control over these customisations, Google has announced changes to the way Chrome extensions will work, which will ‘*stifle innovation*’ and limit what developers can do within the web extension (Miagkov, Gillula and Cyphers, 2019).
- An example from 2016 shows how seams can be exploited to obtain information and increase transparency. By brute force querying of a Facebook API, researchers were able to identify a complete list of 282,000 interests on Facebook and identify the relative popularity of each interest. (Havlak and Abelson, 2016).
- A number of HDR reformers, myself included, had identified a new seam for subverting some of Facebook’s control over how its content is consumed [8.4]: *accessibility tags* or *ARIA tags* (Authors, 2022a). These are specially marked-up tags which are used by screenreaders to display or read content in a more accessible way for partially-sighted or blind people. Because these show page content in a standard format (whereas the HTML of most web pages varies widely and often changes), they present a reliable way to more easily scrape content from the loaded web page within a web extension. In experiments at Open Lab we successfully scraped posts from friends’ feeds (which Facebook do not make available anywhere except the News Feed) so that they could be consumed separately in a more human-focused user experience. This technique has been used successfully to monitor Facebook ads by NYU’s Ad Observatory (Watzman, 2021). In 2021, Facebook was found to have deliberately obfuscated content within ARIA tags to prevent such investigations, resulting in **visually impaired users being unable to differentiate ads from posts and hearing junk characters read aloud**. This can be taken as an adversarial stance against researchers, activists and HDR reformers, and shows that companies like Facebook will go to extraordinary lengths to assert their dominance and reduce user agency (Faife, 2021).

INSIGHT 12: The ‘Seams’ of Digital Services need to be Identified, Exploited and Protected.

- One reason why many companies and services have produced apps is because these are much more locked down and controllable than the web browser environment; there are fewer seams. However, adopting the same philosophy as using web extensions to modify web-based experiences, and drawing on data flow auditing technologies like TrackerControl [9.3] researchers at Oxford University have now developed techniques by which mobile apps can be reverse engineered and modified to change user experiences to better meet users' needs, offering the promise of a '**right to fair programs**' (Kollnig, Datta and Kleek, 2021). These examples make it quite clear that Storni was right: product seams are the place where control can be asserted or regained. They are the setting for the battle for the freedom and integrity of today's information landscape, and it is important for HDR reform that this space is specifically targeted. The role of the HDR reformer here is twofold:

- 1) To **surface information injustices** and especially the closures of seams.
- 2) To **push the seams to gain transparency and re-assert control**, using both and campaigns for policy change.

In this context, the work of whistleblowers such as Frances Haugen (Horwitz *et al.*, 2021) and Edward Snowden (Macaskill *et al.*, 2013) is particularly validated and important. Whistleblowers can expose internal procedure changes that affect the information landscape's integrity that are not otherwise visible. In order to hold online platforms to account, the public must be **aware** and **able to attribute** any restriction in freedom or information access to the correct source, in other words to know that the information or functionality is being modified or restricted. These ideas are explored further in (Bowyer, 2017). Seams should be much more in the public consciousness than they are.

1.4.3 Regulating to Free the Information Landscape

The reason that seams are so important, is that because they a point of friction, a point of resistance to the idea that *data should be separable from services* [Objective 5]. One of the goals of HDR, and particularly this approach, is to wrest control of data away from service providers, so that it can be truly free-flowing, as I envision in (Bowyer, 2018b). In this approach, I have identified some of the key avenues for progressing this goal - through web extensions, accessibility tags, API exploitation and reverse engineering. But in the face of such extreme power, it is clear that such grassroots advocacy will not be enough. For example, web extensions are a powerful approach but are limited in that only those 'power users' who install the extension will gain the benefits. They do not directly help the layperson with their diminishing agency. Therefore, as outlined in [5.5.1], HDR reformers must also work to educate and persuade policymakers of the need for change. As an example of this, in my Case Study Two paper (Bowyer, Holt, *et al.*, 2022) I called for better guidelines from the EU towards data holders. The European Data Protection Board launched a

public consultation on a new set of guidelines they had constructed around GDPR (European Data Protection Board, 2022) and I contributed a detailed set of recommendations to that consultation, emphasising some of the important insights I have identified in this thesis, including a focus on human information, delivering ongoing understanding, establishing standards, and viewing data as an ongoing flow over time (Bowyer, 2022). This sort of engagement by HDR reformers is critical to help shape future policies and guidance, contributing expertise to help the civil servants making those rules. As HDR reformers we must continue to push for better regulation - as it really is the only force that can significantly change data-holding organisations' practices. Better regulation is needed not only to help individual user autonomy, but to combat online extremism (Arthur, 2017). In this context it is important to note that the European Union is, at the time of writing in July 2022, developing a wide range of new laws that could help to improve the agency of individuals and the integrity of the information landscape:

- The Data Act (Board, 2022)
- The Data Governance Act (Commission, 2022a)
- The Digital Markets Act (Commission, 2022b)
- The Digital Services Act (Commission, 2022c)
- The Artificial Intelligence Act (Brakel, Uuk and Boine, 2022)

It is too early to evaluate exactly what these laws will change or what their impact will be, but HDR reformers should observe carefully: they will almost certainly play as pivotal a role in protecting the information landscape as the GDPR has in opening up data access to individuals.

In this approach, I have highlighted how important it is for HDR reformers to **seize and harness the powers we are given, and fight to hold onto them**. Groups of HDR reformers can combine development skills, innovation and disruptive design approaches to find and publicise new ways to circumvent providers' efforts to control and limit users' agency, and policymakers can prevent further erosion of individual agency through legislation and enforcement over the information landscape. As the popularised adage based on the words of John Philpot Curran and Thomas Jefferson goes, '*The price of freedom is eternal vigilance*'.

1.5 Approach 4 to Improving HDR: Winning Hearts and Minds: Teaching, Championing and Selling the Vision

Approach 1, the investigative activist approach [9.2], and approach 3, the digital freedom fighter approach [9.4], can help with users' ecosystem understanding [Objective 3] and ecosystem negotiability [Objective 4]. Meanwhile, Approach 2, the disruptive interaction designer approach, can help with user's direct understanding [Objective 1] and capabilities [Objective 2] over their data. But as outlined in [Objective 5], these approaches will not have sufficient impact until and unless the goals of HDR reform become widely accepted among business

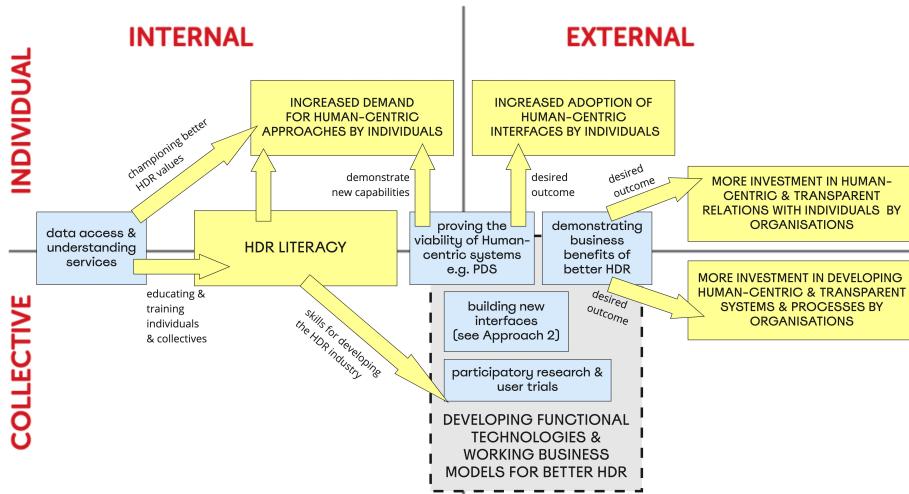


Figure 21: Figure 9.21: HDR Approach 4: Winning Hearts and Minds: Teaching, Championing and Selling the Vision

leaders, citizens, journalists and politicians. And for that to be truly accepted, **better HDR must be proven to work**.

The call for better Human Data Relations is a call for a radical reconfiguration of today's data world. As [Objective 5] outlines, where new systems are needed, system builders must invest in and **see the value of HDR ideals** (not just to individuals but to their organisations). Where new policies are needed, politicians must be persuaded that HDR's ideals are **worthwhile and have public support**. And most importantly for any change to occur, there must be a **demand for change, and an engagement and appreciation** of new HDR approaches once they become available. These are the goals of Approach 4, as illustrated in Figure 9.21 above.

Therefore, there is a complementary parallel trajectory of HDR reformer effort that is needed if the disruptive potential of HDR is to be realised. Across society, we must find ways to **demonstrate, persuade and prove the value of better HDR**. Collectively we could call this '*'motivational work'* in support of HDR reform. I have identified three aspects to this motivational work, which will be explored below:

1. Education & Data Literacy 9.5.1
2. Demonstrating Business Value 9.5.2
3. Proving the Viability 9.5.3

1.5.1 HDR Education & HDR Literacy

My interactions with participants in the pilot study [1.3.1]], Case Study One and Two have all shown me that people do not feel comfortable and confident when it comes to matters of accessing and using their own data. This impression is empirically sound: A research study which surveyed over 1,500 members of the public about attitudes to personal data conducted at BBC R&D in early 2021 highlighted a lack of understanding and confidence around personal data as one of four key findings. Other key takeaways from this research included feelings of helplessness and needing to know the basics (Sharp, 2021). A major part of any effort to overcome the lack of demand for HDR described in 8.5 therefore must begin with **educating people about data** - and more specifically, to educate people about data, life information [7.6.1], and personal data ecosystems [2.3.4; 7.6.2] from a human-centric, forward-looking HDR reform perspective.

Data literacy is already a strong education and skills focus area across public and private sector, for both children and adults. However, I identify some inadequacy in this concept when viewed through an HDR lens. Given the broad and varying perspectives of data [2.1.1], there are also varying ideas about what data literacy is. To some, data literacy is about the technical skills of number crunching, spreadsheets and data analysis (Precisely Editor, 2022). To others, it is a more high-level ability to read, understand and argue with data, and to exercise critical thinking or identify bias (Knight, 2016). A third perspective is that of technical prowess, as outlined by Gurstein as one of the needs for effective data access - the literal ability to interpret and visualise the information within your data (Gurstein, 2011). While all of these are clearly important aspects, something broader is needed to encompass **HDR literacy**. This would encompass such as aspects as:

- being able to **appreciate the value** of your personal data as containing stored facts and records about your life [5.5.3];
- being able to understand the implications of **how your personal data is used and shared** by organisations you interact with [8.3];
- having an appreciation of the importance of establishing **portable data and separation** between data and services [9.2];
- knowing how to **exercise your rights** to obtain your own data, being able to recognise incomplete data returns and knowing how to demand corrections or better responses [5.1]; and
- understanding the need for a **free and fair information landscape** and being able to identify when individual agency is being diminished [8.4; 9.3].

These skills should become part of school curricula, but also need to be taught to adults both in the private and public sector - both in their roles as citizens as well as in their roles within the organisations they work for.

As examples of the sort of educational work that can be beneficial, we need to look at organisations whose remit includes the delivery of education and training. For example:

- The BBC's primary duties are to entertain, *inform* and *educate*. In line with these latter two objectives, the efforts begun in the Cornmarket project continue into public engagement - the Cornmarket team have produced explanatory videos and blogs to convey important parts of the HDR reform messaging to the public (Sharp, 2021; BBC Cornmarket Team, 2021). At the time of writing in the summer of 2022, a new project is about to be released which will include a generally available public offering which will combine PDS technology with the 'BBC Together' watch party concept. This will allow groups of friends to have fun watching or reading BBC content together, while also gaining an understanding of what data is generated while they do and how that information might be utilised to better understand themselves and for the basis of a personal data ecosystem under their own control.
- In Switzerland, Hestia.ai is pioneering a new approach in offering interactive data-exploration workshops to businesses, researchers, educational establishments and activist collectives around Europe. This takes the approach used with EU politicians in the digipower investigation (Härkönen and Vänskä, 2021) and broadens it into general training provision known as 'digipower academy' ('About digipower.academy', 2022) which will enable the creation of workshops where participants can upload their data, explore data flows and aggregate information insights, and learn about the important issues of the personal data economy.

To support this approach and reach wider audiences, efforts such as those of BBC R&D and Hestia.ai need to be invested in, replicated, scaled up and offered to all age groups and all levels of society. This shows that there is a second important reason to support the growth of *data access & understanding services* in addition to the investigative angle in Approach 1, namely to raise HDR literacy across society.

1.5.2 Demonstrating the Business Value of Better HDR

There are two aspects to the motivational problem of generating demand. One is motivating the end users - for which I have shown an approach running through Insight 7 and Approach 2 [9.3]. The other, perhaps more challenging, is to demonstrate that the radical new approaches of HDR reform, while they will entail significant changes and new work, will be worthwhile for businesses and organisations. Some of the challenges of shifting to more inclusive and more human-centric ways of operating are explored in Case Study One [4.4.3]. In the following insight, we consider the ways in which HDR reform such as shared data stewardship [4.2.4], inclusive data flows and individually-sourced data [9.3] might be beneficial to businesses.

INSIGHT 13: It is Possible (and Necessary) to Demonstrate Business Benefits of Transparency and Human-centricity

As outlined in 8.5 and in this section, it is essential that work is done to persuade data-holding organisations of the benefits of moving towards the new paradigms outlined in this thesis. The following avenues for possible future research and advocacy toward data holding organisations have been identified:

- **Trust & Reputation:** In line with the third ‘public relations’-like aspect of HDR [7.2] as well as the recommendations in 4.3.4, 4.4.1, 5.5.2 and 6.2.1, displaying a more inclusive, open and supportive attitude to data handling could strengthen the service relationship and increase customer loyalty and trust. Organisations that are seen to have good human data relations are preferred.
 - **Consent:** In the wake of the GDPR, ensuring consent is becoming an increasing concern to organisations, and the risks of legal consequences for mistakes are high. It makes sense that a more dynamic [Bowyer *et al.* (2018); 4.4.1; 5.5.2; 6.2.2] consent approach that involves individuals [6.2.3] and keeps them in the loop, will enable individuals to speak up much earlier and express consent wishes that might otherwise go undetected.
 - **Accuracy:** The best-placed person to spot errors in data’s accuracy or fairness is the individual about whom the data is concerned. Therefore, increasing their involvement is likely to improve the quality of the data, especially if additional data is contributed or curated by the service user [4.3.3, 6.2.3]
 - **Liability:** In an increasingly litigious society, storage of personal data, especially health or financial data, is a significant liability for businesses, especially if something goes wrong. Investment in human-centred personal ecosystems would outsource the storage of sensitive data to data trusts or PDS providers, reducing liability for the service business. By ensuring that data is accessed only in ways that are centralised outside of the business and remaining in the user’s control—such as PDS company digi.me’s Private Sharing model (digi.me, 2019; Bowyer, 2020)—organisations can ensure that have negligible risk of mishandling customer data.
 - **Better Customer Targeting** The most radical, but perhaps the most persuasive, business model relating to better HDR, is the Vendor Relationship Management approach [2.3.4], where individuals express their own service or product desires explicitly, which vendors then respond to. This turns traditional models inside out, and would empower users more, but due to the inherently improved accuracy of a self-declared interest, might also give businesses a greater confidence that their investment in converting those customers to a sale would be worthwhile. It is important to remember that the current drive towards collecting more data that drives the platformisation trend is in order to improve ad targeting, so that businesses can get a better return on their investment. A VRM approach, or any other approach where the individual contributes improved data to their data self, is in line with that current business objective.
-

This section has identified the areas which need to be evaluated and explored through research or entrepreneurial investment, in order to produce data that could persuade businesses to adopt reformed HDR approaches to data handling and service user interaction.

1.5.3 Proving the Viability of Human-centric Approaches

The third aspect of the motivational work in Approach 4 is that work must be done not just to create new systems and technologies that meet the HDR objectives [7.7], but to **prove that human-centric HDR approaches work**. This involves both developing functional technical **proofs of concept** to test HDR design concepts such as those in Approach 2 [9.3] in practice, as well as starting businesses which can explore new business models to discover which forms of value in the PDE / HDR space can be sufficient to drive the space and its players - which after all will be the engines of change in pursuing HDR - forward.

As an example, development work I undertook with Stuart Wheater as part of my role in the SILVER project (Connected Health Cities, 2017) was successful in building a working system to extract citizens' health data from EMIS, the medical system used by NHS GPs, and make it available as understandable and explorable timeline-based information that could be understood by support workers, drawing on the ideas of temporal PIM systems 2.2.2. The system was extremely difficult to construct due the closed and insular nature of the different parts of the health and social care sector [8.5.2], resistance to change, and the fragmented nature of public sector systems (Pollock, 2011). Nonetheless, it was successful trialled with support workers (using test data) and received positive feedback. It proved that silos can be broken down and life information presented in new ways. A screenshot of the interface I developed is shown in Figure 9.22, and example videos have been published online (Bowyer and Wheater, 2017).

Across the *MyData* and PDE / HDR space, many small businesses have been established which have demonstrated successes in different elements of pursuing the visions of MyData and HDR. For example:

- Digi.me, along with their partners UBDI ('Whose data is it anyway?', 2019; 'Digi.me', no date), have demonstrated the technical and business viability of **human-centric solutions which allow users to securely share data**, through a personal data store-like library, with paying third parties, while maintaining individual oversight and information-viewing functionality. CitizenMe (CitizenMe, 2021) has also demonstrated success in this field, focusing on individual data generation and personal rewards.
- Tim Berners-Lee's company, Inrupt, has successfully produced an effective open source **PDS platform technology called Solid**, which is in use in several organisations across the world (Olivo, 2020). The Solid community has also resulted in the first practical standard for unifying human information from the personal data diaspora [8.1.2], as called for in [8.5],

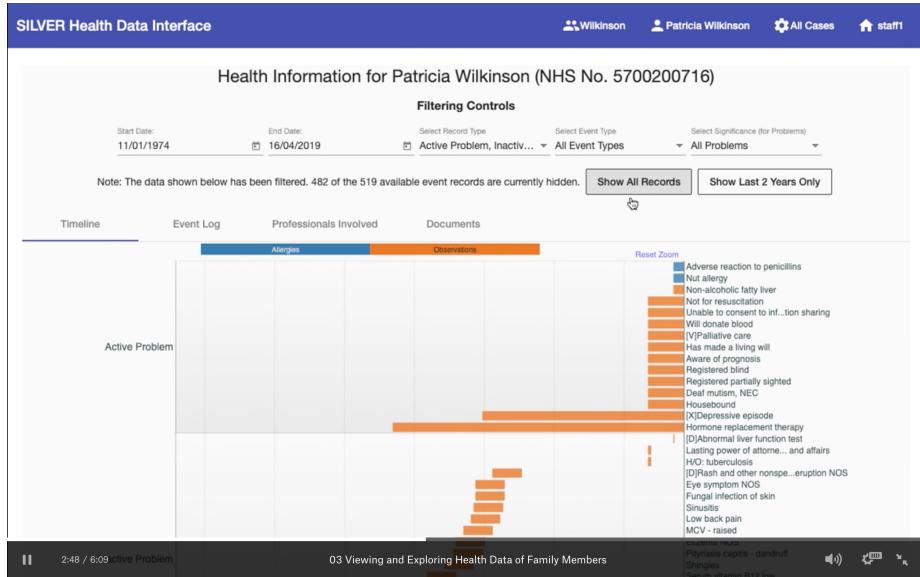


Figure 22: Figure 9.22: SILVER Health Data Viewing Interface

data shapes ('ShapeRepo: Make your apps interoperable', 2022), which is gaining adoption.

- BBC R&D's Cornmarket project has come up with a viable set of designs and mock-ups and a working R&D prototype. Built upon Solid technology, it connects to real data sources including Netflix, BBC and Spotify. If supplemented by some of the features and strategies described in 9.3, it could deliver a **real-world PDS-based life interface**. While not yet tested in the marketplace, the designs have received positive feedback in a 1,500-participant study, including specific appreciation for many of the ideas raised in this chapter and this thesis, including life insights and goal setting, personal data unification, individual oversight, data exploration, the twin goals of LIU and PDEC [7.6], and self-contributed data (Sharp, 2021).
- With several customers including non-profit research organisations, universities, researchers, consumer organisations, local authorities and transport providers, Hestia.ai has proven that there is a **paying market for data access and understanding services** (Dehaye, 2022), which are needed for the reasons outlined in Approach 1 [9.2] and Approach 3 [9.4].

Looking at Approach 4 as a whole, this section has shown that pursuing HDR is not purely a data interaction design problem, nor solely a political problem, nor solely a technical problem. It is all of these, but also, it is a business problem. In our capitalist society it is essential to find a path to better HDR that companies can get behind. And this approach too has shown that motivation and education

go hand in hand with all of the above. It is not just a case of building new things, it's about beginning and catalysing a cycle of constant feedback, of data-enabled design and action research or iterative software and business model development - finding what works, championing it, communicating it and selling it.

1.6 Summation of Chapter 9, and of Part One

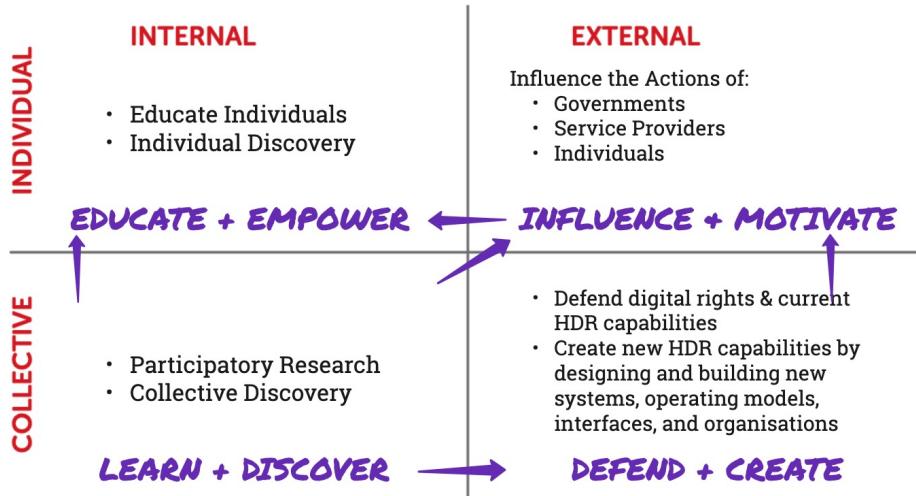


Figure 23: Figure 9.23: Summary of Generalised Change Strategies for Pursuing Better HDR, Using the ToC Model

This chapter examined the expanded research question [7.1] of how better Human Data Relations might be achieved in practice. Through detailed practical examples drawn from the peripheral research settings [7.2] and elsewhere, and building upon the 13 insights Insights introduced throughout Chapters 8 and 9, I have, as illustrated and described four distinct trajectories for improving HDR. These provide practical strategies for HDR reform that could begin to address the expanded research question [7.1] of how better HDR can be achieved.

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

The common elements of these four approaches are summarised in abstract in Figure 10.1 above, and can be considered in terms of the four ToC quadrants:

- **Learn & Discover** (CI): In this quadrant, individuals, researchers, activists and other stakeholders work in groups to understand data attitudes

and user needs, and to gain collective knowledge of data collection and usage practices which are sometimes hidden.

- **Defend & Create** (CE): In this quadrant, activists work to ensure current HDR capabilities are not eroded, while researchers, designers, technologists and social innovators design and create new technologies, operating models, organisations and interface designs, the structures enabling a world with better HDR.
- **Influence & Motivate** (IE): In this quadrant, individuals' relationships with data and with data holders, as well as data holders and policymakers' relationships, can be improved. For us, as external actors seeking change, the task is to influence the many parties by showcasing and facilitating newly created structures and capabilities, and to harness our new collective knowledge to advocate the benefits of changing data-related behaviours.
- **Educate & Empower** (II): In this quadrant, individuals' ways of thinking about data and data holders grow and evolve towards a state we could call '*feeling empowered*' [6.3]. This would constitute definitive progress in delivering the better HDR that I and the participants of Part One's research called for. This change is driven through education, improving HDR literacy, and through the experience of new capabilities and changed relationships with data and with data holders. Such changed relations would empower individuals to hold a more aware and equitable position in every aspect of their digital life.

Taken together, these multi-pronged efforts show how better HDR might be achievable in practice, even despite the identified obstacles. The glossary included at the back of this thesis provides a convenient way for future researchers, innovators, activities and other HDR reformers to quickly locate insights and designs of value within this the thesis, so that others can build on this research to pursue improved Human Data Relations.

Those who wish to pursue these approaches need not be technical or experts. The HDR reform agenda, and in particular these four approaches, can be supported or pursued in many different ways, which may appeal to different readers:

- prototyping and creating proofs of concepts;
- fundraising or investment;
- design activities;
- documentation and knowledge sharing;
- market research, participatory research or usability research;
- 'early adopter' testing and quality assurance of PDE / PDS offerings;
- promotion, advocacy and journalism of HDR issues;
- critical audits of provider practices;
- policy design;
- political pressure on governments and regulators;
- participation in open data, PDE, civic hacking or *MyData* communities; or even
- individual self-experimentation with HDR tools, rights and capabilities

This concludes Part Two of the thesis. In the next and final chapter [Chapter 10], the legacy and contributions of the whole thesis are discussed.

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.
- ‘About digipower.academy’ (2022). available at: <https://digipower.academy/about>.
- ‘Access my info’ (2022). available at: <https://accessmyinfo.ca/>.
- ‘AllOfMe.com Teaser Clip’ (2008). YouTube. available at: <https://www.youtube.com/watch?v=JWYqt4WL6xE> (accessed: 21 March 2021).
- Apple (2022) ‘About app activity report’. available at: <https://support.apple.com/en-us/HT212958>.
- Arthur, C. (2017) ‘Theresa May wants to tackle online extremism. Here’s how to do it’, *The Guardian*. available at: <https://www.theguardian.com/commentisfree/2017/may/26/theresa-may-online-extremism-google-facebook-twitter-manchester>.
- Authors, V. (2022a) ‘ARIA accessibility tags’, *Mozilla Developer Network*. available at: <https://developer.mozilla.org/en-US/docs/Web/Accessibility/ARIA>.
- Authors, V. (2022b) ‘Naive bayes classifiers’, *GeeksForGeeks*. available at: <https://www.geeksforgeeks.org/naive-bayes-classifiers/>.
- Bakshy, E., Messing, S. and Adamic, L. A. (2015) ‘Exposure to ideologically diverse news and opinion on facebook’, *Science*. American Association for the Advancement of Science, 348, pp. 1130–1132. doi: 10.1126/SCIENCE.AAA1160/SUPPL_FILE/PAPV2.PDF.
- Barclay, L. (2021) ‘Facebook banned me for life because i help people use it less’, *Slate*.
- BBC Cornmarket Team (2021) ‘Building trusted data services and capabilities’. available at: <https://paper.dropbox.com/doc/Building-trusted-data-services-and-capabilities--BmF6FlBd0cKGw0nC654th7V3Ag-Us49Ek0nex7yClKughPN4>.
- BBC News (2014) ‘More than 400,000 freedom of information requests made since 2005’, *BBC News*. available at: <https://www.bbc.co.uk/news/uk-politics-30632221>.
- 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.
- Board, E. D. P. (2022) ‘The EU’s data act: Data protection must prevail to empower data subjects’. available at: https://edpb.europa.eu/news/news/2022/eus-data-act-data-protection-must-prevail-empower-data-subjects_en.
- Bowcott, O. and Hern, A. (2018) ‘Facebook and cambridge analytica face class ac-

- tion lawsuit', *The Guardian*. available at: <https://www.theguardian.com/news/2018/apr/10/cambridge-analytica-and-facebook-face-class-action-lawsuit>.
- Bowyer, A. (2011) 'Why files need to die'. available at: <http://radar.oreilly.com/2011/07/why-files-need-to-die.html>.
- Bowyer, A. (2017) 'Designing for human autonomy: The next challenge that civic HCI must address'. available at: <https://eprints.ncl.ac.uk/273832>.
- Bowyer, A. (2018a) 'A grand vision for post-capitalist HCI: Digital life assistants'', *CHI Workshops 2018*. available at: <https://eprints.ncl.ac.uk/273826>.
- Bowyer, A. (2018b) '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*.
- Bowyer, A. (2022) 'Response to guidelines 01/2022 on data subject rights - right of access version 1.0'. available at: https://edpb.europa.eu/sites/default/files/webform/public_consultation_reply/Response%20to%20Guidelinesv1.0-AlexBowyer-March2022.pdf.
- Bowyer, A. and Wheater, S. (2017) 'SILVER family health data interface: Demo screencasts / showcase'. available at: <https://eprints.ncl.ac.uk/273839>.
- Brakel, M., Uuk, R. and Boine, C. (2022) 'The artificial intelligence act'. Future of Life Institute. available at: <https://artificialintelligenceact.eu/>.
- Braw, E. (2022) 'How corporate boycotts could backfire', *Foreign Policy*. available at: <https://foreignpolicy.com/2022/03/28/russia-sanctions-ukraine-corporate-boycotts-could-backfire/>.
- Brest, P. (2010) 'The Power of Theories of Change', *Stanford Social Innovation Review*, 8(2), pp. 47–51.
- Chalmers, M., MacColl, I. and Bell, M. (2003) 'Seamful design: Showing the seams in wearable computing', *IEE Eurowearable*.
- CitizenMe (2021) 'Become a Citizen and unlock the value of your data'. available at: <https://www.citizenme.com/for-citizens/> (accessed: 23 August 2021).
- Colbert, M. (2022) 'Labour members left in the dark over data breach', *Byline Times*. available at: <https://bylinetimes.com/2022/06/06/labour-members-left-in-dark-over-data-breach/>.
- Commission, E. (2022a) 'Data governance act explained | shaping europe's digital future'. available at: <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained>.
- Commission, E. (2022b) 'The digital markets act: Ensuring fair and open digital markets'. available at: <https://ec.europa.eu/info/strategy/priorities>

- 2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en.
- Commission, E. (2022c) ‘The digital services act: Ensuring a safe and accountable online environment’. available at: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en.
- 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).
- Dehaye, P.-O. (2017) ‘Facebook forced to disclose more information about its ad targeting’, *Medium*. available at: <https://medium.com/personaldata-io/facebook-forced-to-disclose-more-information-about-its-ad-targeting-7e6c0127722>.
- Dehaye, P.-O. (2019) ‘Hestia.ai: About us’. available at: <https://hestia.ai/en/about/>.
- Dehaye, P.-O. (2022) ‘Hestia.ai: achievements’. available at: <https://hestia.ai/en/#realization>.
- digi.me (2019) ‘Digi.me private sharing: See how you can do more with your personal data’. YouTube video; YouTube. available at: https://www.youtube.com/watch?v=pGcnK_KraXs.
- ‘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.
- Es, M. van, Guijt, I. and Vogel, I. (2015) ‘Hivos ToC Guidelines: Theory of Change Thinking in Practice’. The Hague, The Netherlands: Hivos.
- European Data Protection Board (2022) ‘Guidelines 01/2022 on data subject rights - right of access version 1.0 - public consultation’. available at: https://edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-012022-data-subject-rights-right_en.
- Faife, C. (2021) ‘Facebook rolls out news feed change that blocks watchdogs from gathering data’, *The Markup*. available at: <https://themarkup.org/citizen-browser/2021/09/21/facebook-rolls-out-news-feed-change-that-blocks-watchdogs-from-gathering-data>.
- Foucault-Dumas, C. (2021) ““This case forces Uber to do a lot more than ever before”, says data protection expert”, *Medium*. available at: <https://medium.com/personaldata-io/uber-vs-drivers-trial-interview-data-protection-expert-rene-mahieu-55359f8cdd9d>.
- Goffe, L. et al. (2021) ‘Appetite for disruption: Designing human-centred augmentations to an online food ordering platform’, in *34th british HCI conference*, pp. 155–167.
- Goffe, L. et al. (2022) ‘Web augmentation for well-being: The human-centred design of a takeaway food ordering digital platform’, *Interacting with Computers*.
- Golder, S. A. and Huberman, B. A. (2006) ‘Usage patterns of collaborative tagging systems’, *Journal of Information Science*, 32, pp. 198–208. doi: 10.1177/0165551506062337.

- 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.
- Härkönen, T. and Vänskä, R. (2021). Sitra. available at: <https://www.sitra.fi/en/projects/digipower-investigation/#what-is-it-about>.
- Hartman, T. et al. (2020) ‘Public perceptions of good data management: Findings from a UK-based survey.’, <https://doi.org/10.1177/2053951720935616>. SAGE PublicationsSage UK: London, England, 7. doi: 10.1177/2053951720935616.
- Havlak, H. and Abelson, B. (2016) ‘The definitive list of what everyone likes on facebook’, *The Verge*. available at: <https://www.theverge.com/2016/2/1/10872792/facebook-interests-ranked-preferred-audience-size>.
- Horwitz, J. et al. (2021) ‘The facebook files’, *Wall Street Journal*. available at: <https://www.wsj.com/articles/the-facebook-files-11631713039>.
- Information Commissioner’s Office (2018) ‘Your data matters - Your rights’. available at: <https://ico.org.uk/your-data-matters/>.
- Jeffers, S. and Webb, L. K. (2017) ‘About who targets me’.
- Jelly, M. (2021) ‘The Mission’. ethi.me. available at: <https://web.archive.org/web/20220312232859/https://www.ethi.me/the-mission> (accessed: 12 March 2021).
- Jones, W. (2011) ‘The Future of Personal Information Management Part I: Our Information, Always and Forever’.
- 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>.
- Knight, M. (2016) ‘What is data literacy?’, *Dataversity*. available at: <https://www.dataversity.net/what-is-data-literacy/>.
- Kollnig, K. (2021) ‘TrackerControl’. available at: <https://www.trackercontrol.org/>.
- Kollnig, K., Datta, S. and Kleek, M. V. (2021) ‘I want my app that way: Reclaiming sovereignty over personal devices; i want my app that way: Reclaiming sovereignty over personal devices’, *CHI Extended Abstracts*. ACM. doi: 10.1145/3411763.3451632.
- Krishnan, A. and Jones, S. (2005) ‘TimeSpace: Activity-based temporal visualisation of personal information spaces’, *Personal and Ubiquitous Computing*, 9(1), pp. 46–65. doi: 10.1007/s00779-004-0291-x.
- Kuchler, H. (2018) ‘Max schrems: The man who took on facebook — and won’, *The Financial Times*. available at: <https://www.ft.com/content/86d1ce50-3799-11e8-8eee-e06bde01c544>.
- Lansdale, M. and Edmonds, E. (1992) ‘Using memory for events in the design of personal filing systems’, *International Journal of Man-Machine Studies*, 36(1), pp. 97–126. doi: 10.1016/0020-7373(92)90054-O.
- 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. (2021) ‘Dutch court rejects uber drivers’ “robo-firing” charge

- but tells ola to explain algo-deductions’, *TechCrunch*. available at: <https://techcrunch.com/2021/03/12/dutch-court-rejects-uber-drivers-robo-firing-charge-but-tells-ola-to-explain-algo-deductions/>.
- Macaskill, E. et al. (2013) ‘NSA files decoded: Edward snowden’s surveillance revelations explained’, *The Guardian*. available at: <https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded>.
- Mahieu, R. L. P. and Ausloos, J. (2020) ‘Harnessing the collective potential of GDPR access rights : towards an ecology of transparency’, *Internet Policy Review*. available at: <https://policyreview.info/articles/news/harnessing-collective-potential-gdpr-access-rights-towards-ecology-transparency/1487>.
- Mahieu, R. and Ausloos, J. (2020) ‘Recognising and Enabling the Collective Dimension of the GDPR and the Right of Access’.
- Maldre, M. (2012) ‘Amazon makes kindle less social -’. available at: <https://www.mattmaldre.com/2012/10/01/amazon-makes-kindle-less-social/>.
- Marshall, C. (2019) ‘What is named entity recognition (NER) and how can i use it?’, *Medium*. available at: <https://medium.com/mysuperai/what-is-named-entity-recognition-ner-and-how-can-i-use-it-2b68cf6f545d>.
- Miagkov, A., Gillula, J. and Cyphers, B. (2019) ‘Google’s plans for chrome extensions won’t really help security’, *Electronic Frontier Foundation*. available at: <https://www.eff.org/deeplinks/2019/07/googles-plans-chrome-extensions-wont-really-help-security>.
- Miller, A. (2021) ‘The right to repair movement, explained – BMC software | blogs’, *The Business IT Blog*. available at: <https://www.bmc.com/blogs/right-to-repair/>.
- Moore-Colyer, R. (2022) ‘Apple launches iPhone self service repair kits — but there’s a big catch | tom’s guide’, *Tom’s Guide*. available at: <https://www.tomsguide.com/news/apple-launches-iphone-self-service-repair-kits-but-theres-a-big-catch>.
- mySociety (2004) ‘About TheyWorkForYou’. available at: <https://www.theyworkforyou.com/about/>.
- O’Sullivan, D., Duffy, C. and Fung, B. (2022) ‘Ex-twitter exec blows the whistle, alleging reckless and negligent cybersecurity policies’, *CNN Business*. available at: <https://edition.cnn.com/2022/08/23/tech/twitter-whistleblower-peiter-zatko-security/index.html>.
- Olivo, O. (2020) ‘The road ahead for inrupt and solid’. available at: <https://inrupt.com/Solid-roadmap-preview>.
- Orphanides, K. G. (2021) ‘The BBC’s radical new data plan takes aim at netflix’, *Wired UK*. available at: <https://www.wired.co.uk/article/bbc-data-personalisation>.
- Owen, L. H. (2012) ‘Betaworks’ findings shifts to web clipping, as amazon bans kindle clips’, *GigaOm*. available at: <https://web.archive.org/web/20120922222936/http://gigaom.com/2012/09/18/betaworks-findings-pivots-as-amazon-bans-kindle-clips/>.
- P., R. (2021) ‘Web scraping extensions: The easy gateway to web data’, *Medium*. available at: <https://raluca-p.medium.com/web-scraping-extensions-the-easy>

- gateway-to-web-data-40e8592e13bf.
- Peysakhovic, A. and Wu, L. (2020) ‘CS 4803 / 7643: Deep learning guest lecture: Embeddings and world2vec, a guest lecture by ledell wu’. available at: https://web.archive.org/web/20211018015836/https://www.cc.gatech.edu/classes/AY2020/cs7643_spring/slides/L13_EMBEDDING_WORLD2VEC_FINAL_version.pdf.
- Pidoux, J. et al. (2022) *Digipower technical reports: Understanding influence and power in the data economy*. doi: 10.5281/zenodo.6554155.
- 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).
- Precisely Editor (2022) ‘Data literacy: What is it and why it matters’, *Precisely*. available at: <https://www.precisely.com/blog/data-integrity/data-literacy-what-it-is-and-why-it-matters>.
- ‘Privacy salon & my data done right’ (2022). available at: <http://www.privacysalon.org/my-data-done-right-right>.
- Schrems, M. (2017) ‘Noyb.eu - our detailed concept’, *noyb.eu*. available at: <https://noyb.eu/en/our-detailed-concept>.
- Schrock, A. R. (2016) ‘Civic hacking as data activism and advocacy: A history from publicity to open government data’, *New Media and Society*. SAGE Publications Ltd, 18, pp. 581–599. doi: 10.1177/1461444816629469.
- Shah, C. and Bender, E. M. (2022) ‘Situating search’, in *CHIR 2022 - Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*. Association for Computing Machinery, Inc, pp. 221–232. doi: 10.1145/3498366.3505816.
- ‘ShapeRepo: Make your apps interoperable’ (2022). available at: <https://shaperepo.com/>.
- 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>.
- Silver, J. (2014) ‘Shamed on twitter, corporations do an about-face’, *Ars Technica*. available at: Shamed on Twitter, corporations do an about-face.
- Sisco, J. (2022) ‘New internal documents show tech giants pushing out competitors’, *Politico*. available at: <https://www.politico.com/news/2022/07/19/documents-antitrust-case-google-amazon-00046522>.
- Solon, O. (2012) ‘How much data did facebook have on one man? 1,200 pages of data in 57 categories’, *Wired*. available at: <https://www.wired.co.uk/article/privacy-versus-facebook>.
- Storni, C. (2014) ‘The problem of De-sign as conjuring: Empowerment-in-use and the politics of seams’, *Proceedings of the 13th Participatory Design Conference on Research Papers - PDC '14*. New York, New York, USA: ACM Press, pp. 161–170. doi: 10.1145/2661435.2661436.
- ‘Tap my data’ (2022). available at: <https://tapmydata.com/>.
- Taplin, D. H. and Clark, H. (2012) ‘Theory of change basics: A primer on theory of change’, *ActKnowledge*, p. 9. available at: http://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/ToCBasics.pdf.
- Teevan, J. B. (2001) ‘Displaying dynamic information’, in *Conference on*

- human factors in computing systems - proceedings*, pp. 417–418. doi: 10.1145/634067.634311.
- Teleciciute, J. (2020) ‘What is faceted search and how does it improve ecommerce UX?’ Searchanise. available at: <https://searchanise.io/blog/what-is-faceted-search/>.
- Tett, G. (2022) “Right to repair” campaign forces rethink by big tech’, *Financial Times*. available at: <https://www.ft.com/content/aabe2aee-cd2b-42e4-9a0b-51f838da89db>.
- ‘The citizens - about us’ (2020). available at: <https://the-citizens.com/about-us/>.
- Tufecki, Z. (2019) ‘We are tenants on our own devices’, *Wired*. available at: <https://www.wired.com/story/right-to-repair-tenants-on-our-own-devices/>.
- Walby, K. and Larsen, M. (2012) ‘Access to information and freedom of information requests: Neglected means of data production in the social sciences’, *Qualitative Inquiry*, 18, pp. 31–42. doi: 10.1177/1077800411427844.
- Watzman, N. (2021) ‘The political ads facebook won’t show you’, *Medium*. available at: <https://medium.com/cybersecurity-for-democracy/the-political-ads-facebook-wont-show-you-e0d6181bca25>.
- Weiser, M. (1994) ‘Creating the invisible interface: (Invited talk)’, in *Proceedings of the 7th annual ACM symposium on user interface software and technology*, p. 1.
- Weiser, M. and Brown, J. S. (1997) ‘The coming age of calm technology’, in *Beyond calculation*. Springer, pp. 75–85.
- ‘What is text mining and content analytics?’ (2022). OpenText. available at: <https://www.opentext.com/products-and-solutions/products/ai-and-analytics/opentext-magellan/magellan-text-mining>.
- ‘Whose data is it anyway?’ (2019). 04: UBDI. available at: <https://www.ubdi.com/blog/whose-data-is-it-anyway> (accessed: 31 March 2021).
- ‘Worker info exchange’ (2022). available at: <https://www.workerinfoexchange.org/>.