Understanding and Designing Human Data Relations

Alex Bowyer

Table of Contents

# 1 Literature Review

## 1.1 Data-Centrism

### 1.1.1 What is Data?

Data is an oft-used word that carries multiple meanings. In everyday speech, it might refer to mobile phone bandwidth, a filled application form or a collection of files. Even experts have a variety of definitions of data, as well as the related concepts of information and knowledge(Zim, [2015](#ref-zim2015)). In this study, we refer to data by its accepted definition as information or knowledge stored in a form suitable for computer processing. Wellisch expressed this as ‘the representation of concepts or other entities, fixed in or on a medium in a form suitable for communication, interpretation, or processing by human beings or by automated systems’(Wellisch, [1996](#ref-wellisch1996)), which is a useful definition as it includes the fact that both humans and algorithms can use data, and that data is something that needs interpretation.

From a strict grammatical stance, ‘data’ is a plural of the singular ‘datum’ thus it is more correct to write ‘the data are correct’ - but this usage is rapidly declining from use(‘Data’, [n.d.](#ref-grammaristData)) and throughout this thesis I use the more widely adopted usage of treating data as a singular mass noun, as in ‘the data is correct’.

The concepts of ‘data’ and ‘information’ are closely related, so much so that they are often used interchangeably. Ackoff presented a model for distinguishing data, information, knowledge, understanding/intelligence and wisdom, in which he describes data as the physical symbols, effectively the 1’s and 0’s stored in a computer or the ink marks on a page, which becomes useful when humans or algorithms are able to deduce facts from those symbols to answer simple questions - at this point it becomes ‘information’. Being able to interpret deeper how and why questions allow information to become knowledge and understanding, towards the ultimate goal of wisdom(Ackoff, [1989](#ref-ackoff1989)). This is often represented as the DIKW pyramid (DIKW being shorthand for the data-information-knowledge-wisdom transformation that occurs as you move up through the layers), the origin of which is unknown(Wallace, [2007](#ref-wallace2007)). Figure 1 builds upon a representation by George Pór(Pór, [1997](#ref-por1997)) of the pyramid as a ‘wisdom curve’, showing how increasing meaning and value can be obtained from data as deeper questions can be asked of it. This theme of obtaining meaning and value from data is an important aspect of my research that I will refer back to.



REDRAW Figure 1: Making Data into Meaningful Information

This model that turning data into information can be thought of as using that data to answer questions is consistent with the idea that “information can be thought of as the resolution of uncertainty”(‘Information’, [n.d.](#ref-wikipediaInformation)). The exact origin of this definition is unknown but it is often attributed to mathematician Claude Shannon(Shannon, [1948](#ref-shannon1948)). Indeed from an etymological stance, one who is informed is one who has received knowledge or concepts as a result of what has been communicated to them. Thus we can consider that data is the material from which that information can be received. It follows also that data contains uncertainty that must be resolved in order for it to become meaningful information.

### 1.1.2 The Rise of Data-centrism

The earliest computer systems used data to store mathemical and scientific facts. Data processing allowed for previously manual operations to be performed with greater speed and accuracy, most famously the work of Alan Turing and the case of the Enigma code breakers during World War II(Hutton, [2012](#ref-hutton2012)). This work was the advent of general-purpose computing - machines that could be applied to any problem provided you could reduce that problem to data. Businesses over the following decades began to apply computers to myriad new problem areas in all different fields of work and life, and doing so began the encoding of information about people as data, be it for statistical purposes like censuses or research, or simply to enable the more efficient serving of customers by storing databases of customer records.

The personal computer revolution(‘The personal computer revolution’, [n.d.](#ref-britannicaPCrevolution)) of the late 1970s and 1980s put computers in every office and eventually every home too, and it soon became commonplace that each individual would have data stored about them in companies’ databases. In the subsequent years three factors have combined to accelerate this trend of storing data about people: i) labour costs have remained high and companies have sought ways to automate their businesses and to implement online services and call centres in place of in-person staff interaction, ii) computer processing and storage has become ever cheaper thanks to the advent of cloud computing, meaning that many business processes could be reduced to data processing tasks or entire businesses be moved online, and iii) the rise of smartphones and web-enabled devices have meant that the public are now ready and willing to conduct much of their daily business online through the web and apps. These factors have encouraged both commercial and civic providers to centralise their services and to ‘go digital’ to the greatest degree possible. In doing so they collect ever more data about people (now ‘service users’ or just ‘users’). Data is now seen as a resource which can be mined for value, and harnessed for profit and business efficiency - ‘the new oil’(Toonders, [2014](#ref-toonders2014)). Zuboff, in her 2019 book on ‘surveillance capitalism’, characterises this new digital world as the collection of human behaviour data so that it can be used as free raw material and converted into profit through hyper-personalised advertising and targeting by software platforms(Zuboff, [2019](#ref-zuboff2019)). This philosophy is also known as ‘data-ism’(Brooks, [2013](#ref-brooks2013)).

As a result of data-ism, the collection of data about people has become an inevitable part of modern life. We live ‘digital lives’(‘Our digital lives’, [2018](#ref-ted2018)) where we each interact directly and indirectly with hundreds of digital systems every day - as you shop, socialise, or browse online; as you listen to music or watch TV; as you interact with governments or healthcare services; as you travel, and many more. Every one of those interactions indicates the presence of data about you stored in a company database. Every aspect of our lives involves the input, processing and output of data – either provided by, collected from, or generated about, us. And the digital data we create and consume (whether consciously or not - data sharing is often unwitting(Crabtree & Tolmie, [2018](#ref-tolmie2018))) has a direct influence on our lived experience - from decisions about what we are entitled to and what opportunities we will be offered, to the advertisements and content recommendations we are shown while we browse.

Unfortunately, the large-scale systems which collect data about us now function as ‘data traps’(Abiteboul et al., [2015](#ref-abiteboul2015)) - where data about us is easily gathered but very hard to remove or even to access. This creates a lack of agency for the individuals living in this data-centric world. The World Economic Forum’s “Rethinking Personal Data” project recognised the critical role that data - specifically personal data - data created by and about people - now holds, and identified that ‘an asymmetry of power exists today […] created by an imbalance in the amount of information about individuals held by industry and governments, and the lack of knowledge and ability of the same individuals to control the use of that information’(World Economic Forum, [2011](#ref-WEF2011), [2013](#ref-WEF2013), [2014b](#ref-WEF2014context), [2014a](#ref-WEF2014lens)).

### 1.1.3 Data Protection & GDPR

Since as early as 1973, the need to protect individuals’ rights over their data has been recognised(US Department of Health Education and Welfare, [1973](#ref-USDOHEW1973)). The 37-nation organisation OECD in 1980 stated that “the right of individuals to access and challenge personal data is […] the most important privacy protection safeguard” and issued recommendations that individuals should be given basic privacy rights, including the right to be informed whether data is stored about them, and the right to an intelligible copy of that data(Organisation for Economic Co-operation and Development, [1980](#ref-OECD1980)).

Over the subsequent decades, lawmakers began to enact laws to deliver these rights to individuals, notably the UK’s Data Protection Act 1984 (which set up an independent body, the Data Protection Registrar (now the Information Commissioner’s Office) with which organisations were required to register their usage of personal data), Ireland’s Data Protection Act 1988 (which introduced the concept of a ‘duty of care’ for data collectors - that they are expected to avoid causing damage or distress to data subjects), the EU’s Data Protection Directive in 1995 and the UK’s Data Protection Act in 1998. However, such laws were generally found to be ineffective - in 2002 Simon Davies, director of Privacy International said that the UK’s DPA was “almost useless in limiting the growth of surveillance”(Millar, [2002](#ref-millar2002)).

It was only in 2018, when the EU’s General Data Protection Regulation (GDPR) came into force, carrying with it significant designed-to-hurt fines for non-compliance(Kelly, [2020](#ref-kelly2020); Leprince-Ringuet, [2021](#ref-zdnet2021)), that individuals have been able to practically exercise their data rights to any meaningful degree(Atebits.com, [2020](#ref-atebits2020)). The GDPR – which gives individuals key rights including rights to timely data access, explanation, erasure and correction(Information Commissioner’s Office, [2018](#ref-ico2018)) – can be seen as the first serious attempt to rebalance the aforementioned power imbalance over data between citizens and organisations and is generally regarded as a landmark piece of legislation and a strong template for individual data protection. Around the world, companies have overhauled their privacy policies and updated their business practices to comply with the GDPR and other similar legislation, such as Japan’s 2017 Act on the Protection of Personal Information, India’s 2019 Personal Data Protection Bill and the 2020 California Consumer Protection Act. In the USA, there has been no national privacy law yet, but the GDPR’s influence is being felt in court rulings(Hoofnagle et al., [2019](#ref-hoofnagle2019)).

Also in 2018, the Cambridge Analytica scandal(‘Facebook–cambridge analytica data scandal’, [2014](#ref-wikipedia2018cambAna)) broke; the personal data of 87 million people, acquired from Facebook, was exploited with the apparent intent of influencing voting outcomes including the UK’s 2016 Brexit referendum and the USA’s 2017 election of Donald Trump. This combined with widespread public information campaigns about GDPR have led to a heightened awareness of personal data rights(European Union Agency for Fundamental Rights, [2020](#ref-EUAFR2020)) and at the time of writing in 2021, personal data protection laws and individual digital rights remain a rapidly evolving area.

From the GDPR and its antecedents, a number of key terms have been established which I will adopt in this thesis, specifically (Information Commissioner’s Office, [2014](#ref-ico2014); The European Parliament and the Council of the European Union, [2016](#ref-GDPR2016)):

* *Personal data* means any information relating to an identifiable natural person - one who can be identified directly or indirectly by reference to an identifier such as a name, identification number or location or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person.
* The *data subject* is the identified individual, living or deceased, who the personal data relates to.
* A *data controller* is the legal entity (company, public authority, agency, individual or other body) which collects or stores personal data about an individual and determines the means and purposes for which it is processed. Liability for data protection compliance rests with the data controller.
* A *subject access request* is the right to a copy of your personal data.
* *Data portability* is the right to receive a copy of all stored data about you, not just that which you provided, in an accessible and machine-readable format such as a CSV file, so that you can transport it to another service or make use of it.

### 1.1.4 The Need for Practical and Effective Data Access

[Target 400 words]

Para 1 - to tackle the imbalance, need transparency, control, trust and value. Larsson, to assess value people need awareness

Para 2 - effective access and what it entails

Para 3 - what is missing in terms of ability to use our data. distributed data - need for associativity, can’t do that due to silos etc. time, timelines - the needs of filtering, data detail. etc.

it’s about finding meaning. there is today no interface that offers such functionality over our digital data.

### 1.1.5 Research Gap: The Human Experience of Data

[Target 400 words]

so the gap is - what value and meaning fo people find of their data. what mental models do they have around data. what is data to them, what is their relationship to it (and what would they like it to be). what is data to people living in a data centric world?

## 1.2 Personal Data Interaction

[Target 2,500 words]

### 1.2.1 Computers as General Purpose Information Tools

[Target 300 words]

### 1.2.2 Personal Information Management

[Target 600 words]

### 1.2.3 Personal Informatics & The Quantified Self

[Target 650 words]

### 1.2.4 The Emergence of Complex Digital Lives

[Target 600 words]

### 1.2.5 Research Gap: The Data Beyond The Individual

[Target 350 words]

## 1.3 Human-Centric Computing

[Target 2,500 words]

### 1.3.1 Human Computer Interaction Foundations

[Target 400 words]

### 1.3.2 Data Transcendence & Human Data Interaction

[Target 600 words]

### 1.3.3 People In Context: Human-Centred Design

[Target 600 words]

### 1.3.4 The Personal Data Economy

[Target 600 words]

### 1.3.5 Defining the Research Agenda for Human-Centricity in Practice

[Target 350 words]

## 1.4 Research Gap

[Target 500 words]

# Bibliography

Abiteboul, S., André, B. & Kaplan, D. (2015) *Managing your digital life with a Personal information management system*. 58 32–35.

Ackoff, R.L. (1989) From data to wisdom. *Journal of Applied Systems Analysis*. 16 (1), 3–9.

Anon (n.d.) *Data*.

Anon (2014) *Facebook–Cambridge Analytica Data Scandal*. 16.

Anon (n.d.) *Information*.

Anon (2018) *Our Digital Lives*.

Anon (n.d.) 'The personal computer revolution', in *Britannica*. [Online].

Atebits.com (2020) *The GDPR: Does it Benefit Consumers in Any Practical Way?*

Brooks, D. (2013) *The Philosophy of Data*.

Crabtree, A. & Tolmie, P. (2018) The practical politics of sharing personal data. *Personal and Ubiquitous Computing*. 22 (2), 293–315.

European Union Agency for Fundamental Rights (2020) *Your Rights Matter: Data Protection and Privacy 2020*. 20.

Hoofnagle, C.J., Sloot, B. van der & Borgesius, F.Z. (2019) The European Union general data protection regulation: What it is and what it means. *Information and Communications Technology Law*. 28 (1), 65–98.

Hutton, D.M. (2012) *Turing’s Cathedral: The Origins of the Digital Universe*.

Information Commissioner’s Office (2014) *Data controllers and data processors: what the difference is and what the governance implications are*. 20.

Information Commissioner’s Office (2018) *Your data matters - Your rights*.

Kelly, R. (2020) *The Biggest ICO Fines Ever Issued*.

Leprince-Ringuet, D. (2021)

Millar, S. (2002) *UK singled out for criticism over protection of privacy*.

Organisation for Economic Co-operation and Development (1980) *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*.

Pór, G. (1997) 'Designing Knowledge Ecosystems for Communities of Practice', in *Advancing organizational capability via knowledge management*. [Online]. 1997

Shannon, C.E. (1948) A mathematical theory of communication. *The Bell system technical journal*. 27 (3), 379–423.

The European Parliament and the Council of the European Union (2016) *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data*. 16–32.

Toonders, J. (2014) *Data Is the New Oil of the Digital Economy*. 1.

US Department of Health Education and Welfare (1973) *Records Computers and the Rights of Citizens*.

Wallace, D.P. (2007) *Knowledge management: Historical and cross-disciplinary themes*. Libraries unlimited.

Wellisch, H.H. (1996) *Abstracting, indexing, classification, thesaurus construction: A glossary*. American Society of Indexers.

World Economic Forum (2011) *Personal data : The emergence of a new asset class*. 1–40.

World Economic Forum (2014a) *Rethinking Personal Data : A New Lens for Strengthening Trust*. 35.

World Economic Forum (2014b) *Rethinking personal data: Trust and context in user-centred data ecosystems*. 35.

World Economic Forum (2013) *Unlocking the Value of Personal Data: From Collection to Usage Prepared in collaboration with The Boston Consulting Group Industry Agenda*. (February),.

Zim, D.C. (2015) *What is the meaning of "data", "information", and "knowledge"?* 3 (1), 2.

Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile.