UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

Electronics and Computer Science

Volume 1 of 1

Personal Data: Definition and Access

by

Brian Parkinson

Thesis for the degree of Doctor of Philosophy

April 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

Electronics and Computer Science

Doctor of Philosophy

PERSONAL DATA: DEFINITION AND ACCESS

Brian Laurence Parkinson

The terminology around personal data is used inconsistently, the concepts are unclear, and there is a poor understanding of their relationships. As a result, debate is hindered and individuals are increasingly concerned about the wider and more pervasive set of digital services that create inconceivable amounts of data which are collected, curated, matched, and compared by corporate and governmental actors.

This research focuses on all data descriptive of an individual, named the digitally extended self, how it may be categorised, modelled, and accessed, then the issues associated with that access. A lexicological analysis of the terms used to describe personal data is conducted, and used to identify common concepts, proposing a model of the digitally extended self, showing how these concepts of personal data fit together. The model is then validated against key publications.

The author's personal data was collected, using an auto digital ethnographic method, from a purposive sample of organisations representing a range of sectors in the UK then snowballing to the rest of the EU and beyond. An analysis of this data, and the process used to collect it, is conducted, demonstrating that individuals cannot discover their full digitally extended self. Variations between categories of data, organisational sectors and location of the organization are examined.

Reasons for these variations are explored through nine semi-structured interviews with experts including legislators, IT management, Data Protection Officers, and a think tank director. Content analysis of the interview transcription points to a lack of willingness and capability as the reasons for the poor performance and lack of transparency, evident in government bodies.

There are four claims to original knowledge; first the categorisation and model of personal data; second, the analysis showing variations in organisational performance; third, the analysis illustrating the impossibility of knowing one's own digitally extended self, and fourth, an assessment of, and reasons for, the poor performance of government organisations in responding to subject access requests.

Table of Contents

Table of Tables

Table of Figures

Academic Thesis: Declaration of Authorship

I, Brian Laurence Parkinson declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Personal Information, A Model, Its Validation, and a Way Forward

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
   Parts of this work have been published as:

   Parkinson, B. et al. (2017) The Digitally Extended Self: A lexicological analysis of personal data. *Journal of Information Science*, 016555151770623.

Signed: .....

Date: .....

Acknowledgements

I would like to thank my supervisor David Millard, whose commitment, sagacity and guidance has been invaluable, and without whom this work would not have been possible. I am also grateful to my other supervisors Kieron O'Hara and Richard Giordano who have both provided insightful advice and help when needed.

Thanks also go to the individuals who have provided input to this research; the many employees who have responded to my requests for details of my own personal data held on their organisation's systems, and the interviewees who gave their time and expert opinion to shape the conclusions to this research.

My parents would have been both proud and amazed that their son could have produced this work, it's not what people from our background are expected to do. But then they did not have the pleasure and privilege to meet my wife Judy Sebba. It is she who encouraged me to embark on this journey and who has stayed the course guiding me through my post-graduate adventure, by offering wise words of advice, encouragement, and support.

Thank you.

Chapter 1: **Introduction**

1.1 Background

This thesis focusses on personal data, the volume of which is growing beyond imagination. Much of the new information collected each year is associated with individuals (Reinsel et al., 2017). It can be viewed in two ways; either from the stance of its usage (e.g. Lupton and Michael, 2017), and impact on people and society (e.g. Mayer-Schönberger and Cukier, 2013). Alternatively, one can focus on its composition, origins and nature, which is the stance taken in this research. The thesis examines what personal data is, where it resides, who controls it, how it can be accessed, and as a result: why it is impossible for an individual to know their digitally extended self. Without this knowledge, individuals will not have the opportunity to examine:

- where and how their privacy is affected;
- the agencies that manipulate their data perhaps beneficially, or alternatively in ways that limit their life options;
- those attempting to affect their actions.

Personal data is tightly connected with ideas of personal privacy. Floridi proposes that individuals may be seen as constructed of data, in which case a breach of privacy is an act of aggression towards the person (Floridi, 2006a). Indeed, it has been seen as 'the most valued of rights' (Brandeis, 1928, p. 277 U. S. 478) and is now a universal right under Article 12 of the Universal Declaration of Human Rights (United Nations, 1949).
However, privacy is not a new concern. Gutwirth (2002) discussed the history of privacy reaching back to its unacceptability amongst the Roman elite, through to the privacy associated with family life in the early 19th Century, and its constraints upon family members (other than the male head of the family), with the emergence of personal privacy for all family members in the late 19th Century through to current times.
Arguably, since the introduction of the printing press by Gutenberg in the 15th Century, and certainly since the production of the portable camera in the late 19th century, the effect of technology on personal privacy has been of concern. It was the portable camera that was referenced by Warren and Brandeis when arguing against new technologies affecting people's privacy, and the sad state of the U.S. Press in that they were turning to gossip rather than news, and using new photographic and printing technology such that:

> ‘numerous mechanical devices threaten to make good the prediction that “what is whispered in the closet shall be proclaimed from the house-tops”’ (Warren and Brandeis, 1890, p. 2).

It was this that led to the early definition of informational privacy, ‘the right to be let alone’. (Brandeis 1928, p. 277 U. S. 478).

The subsequent development of computers in the late 1940s and early 1950s led to large scale commercial use of ‘mainframe’ computers in the 1960s and 1970s. This was seen as a new threat to privacy, but not only that, Westin described the computer as a new technology of privacy invasion, framing informational privacy in terms of power or control:

> ‘the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others’ (Westin, 1967, p. 7).

The next major step change was a combination of four elements; first personal computing, in the late 1970s, spread computing capability into the hands of individuals and conversely individual information into the personal computer. Second, the transnational movement of information through networked devices in the 1990s, allowed the transfer of more personal data to other connected computers, and to data stores. Third, the creation of data warehouses in the late 1990s facilitated large scale matching and analysis of personal data. It was in 1999 that Scott McNeally the co-founder of Sun Systems stated that ‘you have zero privacy anyway, get over it’ (Sprenger, 1999, p. 1). Perhaps in response to these changes, privacy was couched not in terms of control but in terms of appropriate use of data, judged by the data subject, as ‘contextual integrity and the reasonable expectation of privacy’ (Nissenbaum, 2010, p. 233). Finally, data storage costs have decreased and the storage density increased so that it is now often cheaper to keep data than to delete it (Hypponen, 2014). As a result, it is estimated that in 2016, 16.1 zettabytes of data had been generated (Reinsel et al., 2017). The response from the privacy community is a call for the right to be forgotten (Ausloos, 2012) or at least for corporate amnesia (Mayer-Schönberger, 2009).

The increase in data has been enabled by the technology but driven by new facilities in the areas such as search engines, social media, messaging, location services, video and music streaming, and health measurement self-quantification. Alongside the increasing range of functions in attempts to describe, analyse, forecast and recommend courses of action a number of often ambiguous terms have been developed to label or characterise elements of personal data.

It is in this context that this research examines firstly terminology used to label personal data from a compositional lens, and then explores some of the scale and accessibility issues associated with exploring one's own digitally extended self.

In summary, it is the interaction between technology and personal data which threatens privacy that has motivated this research. It is the intention that its results will enable individuals, organisations and the legislature to better understand the nature of personal data and encourage increased transparency about the data organisations hold, and what they do with it.

1.2 Research Questions

This thesis will explore personal data, which is considered an extension of the individual within the digital domain and labelled *the digitally extended self.* The terms used to describe personal data will be analysed to investigate whether a consistent nomenclature can be developed which will help with the understanding of personal data. This will be applied to real data from a range of organisations enabling validation of the model and providing some insight into organisational transparency in relation to personal data. The availability of data constituting parts of the digitally extended self will then be addressed, together with possible barriers to its access.

The questions to be addressed through this process are:

RQ1: What are the components of the digitally extended self and how do they relate to one another?

RQ2: How feasible is it for an individual to obtain the information, held by organisations, which is descriptive of them?

RQ3: What is the quality of the personal data returned by organisations when it is requested by individuals?

RQ4: What are the reasons for the variations found in the performance of different classes of organisations?

1.3 Research Framework

This research is formed of three pieces of work which produce five outputs as illustrated in Figure 1.1. The first is a lexicological analysis of the terms in the literature relevant to personal data, in order to define a model of the digitally extended self. The second is a subject access request investigation of the author's own personal information in order to populate an example model, and the third consists of interviews with domain experts in

order to better understand the barriers and challenges faced by organisations in providing personal data. Each phase together with the outputs are discussed below.

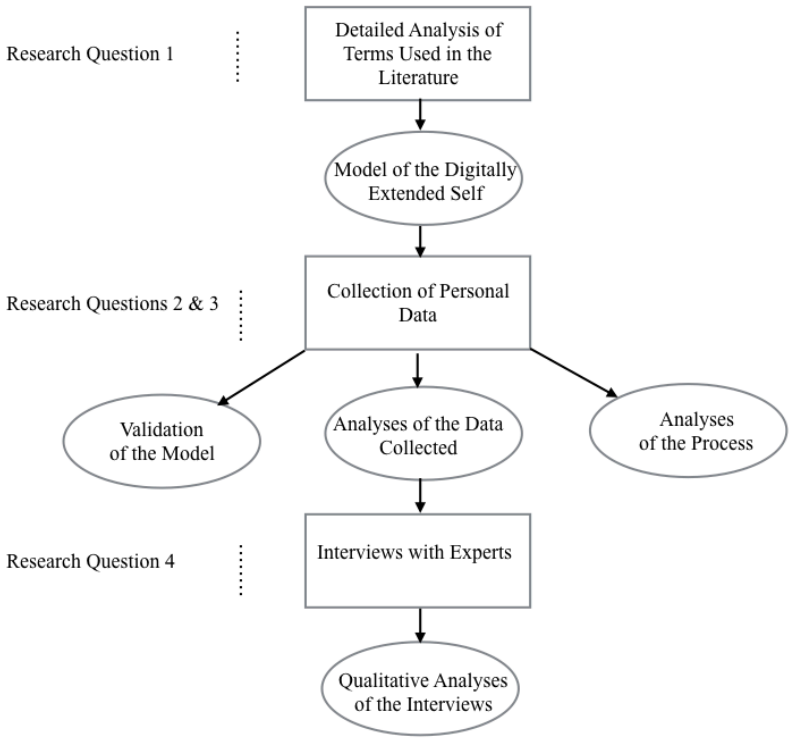


Figure 1.1 Research Framework, showing key research work, and their relationships to the research questions

Detailed Analysis of Terms Used in the Literature

Chapter 3 reports on the first stage of this research, where Google Scholar is searched for terms related to personal data. A sample of publications then is taken from the results based on the number of citations relative to age of the publication, and the resulting papers and books are examined for the meaning of each term. Additional terms are extracted from the sample literature and new searches completed. The terms and meanings are analysed, highlighting inconsistency in use, and a proposed standard set of terms are recommended, their inter-relationships being represented by a model of the digitally extended self. The model is then validated by referencing a subset of the sample based on citation level and range of terminology.

Subject Access Request Investigation

In the second phase, described in Chapter 4, the author's personal data was requested from a purposive sample of organisations using subject access requests. The number of organisations is then extended by using a snowball sample, based on data extracted from the replies from the original sample. The data collected is then compared to the categories defined in the model, and this forms the second validation of the proposed terminology.

The responses from the organisations are examined and an analysis of the responses presented illustrating how well they provide data from the different areas of the model. In addition, the process of data collection itself is measured and analysed.

Interview with Experts - Chapter 5

In Chapter 5, the results from the analyses of the personal data collected in the previous phase form the basis of interviews with nine experts. They provide some insight into the reasons for the outcomes of the subject access request investigation. It is recognised that the robustness of the findings is constrained by the number and type of experts available for interview, but nevertheless interviews with domain experts provides specific, specialist opinion and explanations not available through other methods.

1.4 Contribution to Knowledge

This research addresses four gaps in previous knowledge. First, the lack of a consistent classification system for personal data, which it does by analysing the use of terms that describe personal data in a wide range of literature, and as a result of that analysis, proposing a nomenclature and model for personal data (Parkinson et al., 2017). Second, the absence of research into the performance of organisations in responding to requests for copies of personal data which is achieved through the process of sending requests for personal data to organisations and analysing the completeness of their responses. Third, the lack any evaluation regarding the practicality of individuals collecting the full range of their own personal data, achieved through the measurement of processes followed in the second phase of this research. Finally, an assessment of, and reasons for, the poor performance of certain organisations in responding to requests for personal data, identified through content analysis of interviews with domain experts.

1.5 Publications of this work to date

Parts of this work have been published as:

- Parkinson, B. et al. (2017) The Digitally Extended Self: A Lexicological Analysis of Personal Data. *Journal of Information Science*, pp. 016555151770623.

1.6 Structure of this thesis

Chapter 2 outlines the background to this research by drawing on the literature to examine what is meant by personal data and privacy, and why this may matter to people. The use of

data and its threats to privacy are discussed, as well as the beneficial uses of data, before exploring some of the initiatives to protect privacy. Finally, the self and its extension into the digital world are examined putting this research into context.

A lexicological analysis of the terms used to label personal data is then described in Chapter 3, which illustrates an inconsistent use of terms, and that no move to establish standard nomenclature has been made. A standard classification for personal data is proposed and a model created to illustrate the relationships between data categories. This categorisation and model is then validated against the literature and used as the basis for the following research.

Further validation of the model, against real personal data, is demonstrated in Chapter 4, through the use of the author's data obtained by submitting subject access requests to UK based organisations, and requests for data to those outside the UK. The results of further analysis of this data are then discussed in relation to the process and its implications for people wanting to discover the extent and use of their data, and with respect to how organisations performed when answering requests for personal data.

In order to understand some of the variations in organisational performance a panel of experts are interviewed and the transcriptions analysed. An examination of the results suggests underlying reasons for organisational behaviour observed in Chapter 4, and this together with reflections on the model are presented in Chapter 5.

Finally, Chapter 6 concludes this thesis by drawing together the findings presented in this work in the context of the research questions, before examining some of its limitations. The contributions to knowledge are then discussed before implications for policy and practice are considered and future research identified with respect to new legislation and changes in the use of data.

Chapter 2: **Background**

There is a large volume of personal data curated in the world, and therefore potential privacy issues. In 2010 Eric Schmidt, then CEO of Google, stated that:

> 'between the birth of the world and 2003 there were 5 exabytes of data created ... in the last bit we create 5 exabytes in 2 days' (Schmidt, 2010, p. t18:14).

In 2017, a report produced for Seagate Technologies indicated that 16.1 zettabytes of data had been generated in 2016, much of which was probably personal data (even if in the form of sound or video) (Reinsel et al., 2017). That is 16,486 exabytes or 912 times that thought to be created in the same period 7 years earlier. These are large numbers, people are now familiar with a gigabyte of data (a CD holds .7 gigabytes), so 16.1 zettabytes are 17,702,137,207,193 gigabytes. By comparison, the computer that was used in Apollo 11 to land men on the moon held 0.000061 gigabytes of data. The increasing availability and use of personal data has led to ongoing privacy concerns (Acquisti et al., 2015).

This chapter will begin by examining what personal data is, then address privacy as a topic, before considering why personal data and privacy should be of concern. It then presents some of the uses of data in order to set the context for a brief examination of the threats to privacy that large scale access to personal data can facilitate. The benefits to be reaped from the exploitation of this data are considered next, followed by an examination of some initiatives aimed at preserving privacy. Then other implications for personal data and privacy are highlighted before finally the concept of the digitally extended self is interpreted. This leads into Chapter 3, an examination of the terms used to label and categorise personal data.

2.1 What is Personal Data?

Overton (2016) describes personal data as 'everything that identifies an individual, from a person's name to telephone number, IP address, date of birth and photographs'. Others have used a wider interpretation, for example the Australian Privacy Act defines personal data as 'information or an opinion about an identified individual, or an individual who is reasonably identifiable' (Australian Government, 1988, p. S6).

Three categories of data are suggested by Hildebrandt, O'Hara and Waidner (2013) volunteered data, that which people voluntarily share about themselves; observed data, that which is captured by observing a person's activities; and inferred data based on the analysis of the above. Inferred data may not identify a single person but will be attributed to a person, e.g. a propensity to buy a product. Under these definitions both voluntary and

observed data are considered personal data under EU law (European Union, 2016) whilst inferred data is not, as it is not initially linked to an identifiable person (but a group). However, it may be argued that as soon as individuals are considered a member of an inferred group, (e.g. likely to need a TV licence in the UK as they have moved to a new residence) then that attribution becomes personal data.

The wider definition of personal data used within this work will therefore be 'data that is an attribute of an individual'. First, this includes data artefacts produced by a person that is linked to, and descriptive of, that person, e.g. emails, search enquires, social media posts. Second, the definition will include data that is created by another person or item of technology, e.g. emails sent to a person, Facebook posts that identify a third party, or a comment written about another person in a business setting such as an appraisal. Third, data that is the product of some analysis that is linked to that person, due to their own data or the data of a group with which they may be associated, e.g. they are considered likely to buy an insurance policy, or that they may be a possible terrorist. A final set of data is that derived from items with which they are associated, for instance, location data associated with a mobile phone or car. When presenting potential privacy ramification from modern vehicle software and firmware, Simon and Graham (2017, p. 456) define personal data in this context as:

> 'data that are under the control of a covered entity, not otherwise generally available to the public through lawful means, and are linked, or as a practicable matter linkable by the covered entity, to a specific individual, or linked to a device that is associated with or routinely used by an individual'.

These data are attributes of a device, but if the device is linked to an individual at the time that data was created then the data becomes an attribute of the individual as well, e.g. when a person is known to be in a car with their phone as it passes an ANPR camera.

The Data Protection Act 2018 Part 1 section 3.2 defines personal data as 'any information relating to an identified or identifiable living individual' subject to certain exceptions. This covers data which when collected together identifies an individual; de-identified, encrypted or pseudonymised data but which can still be used to re-identify a person; and anonymised data for which the anonymisation may be reversed (European Commission 2018).

2.2 Privacy

The meaning of 'privacy' has been the cause of much debate and thought. US Supreme Court Justice Louis Brandeis (1928, p. 277 U. S. 478) called it 'the most comprehensive of rights and the right most valued by civilised men'. In the same judgement he inferred that it

was a part of ‘the right to be let alone'. In this, he echoed Judge Thomas McIntyre Cooley (1879, p. 29) that ‘The right to one's person may be said to be a right of complete immunity: to be let alone', although Cooley was referring to personal injury rather than invasion of privacy. In his judgement Brandeis had been reacting to changes in technology, a theme, which Westin (1967) echoed with his suggestion that as technology changes, so does the balance between privacy and disclosure. Westin (1967, p. 7) defined privacy as:

> ‘the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others’.

The topic of control is closely related to concepts of privacy. For instance, the individual’s ability to control access to what they consider to be private, as Parent (1983), Gavison (1980) and Allen (1988) discuss. However, that is insufficient to define privacy from a legal standpoint, as the choice and therefore definition of privacy in any one situation would depend on the views of each individual. But, as Allen (1988, p. 26) states ‘the ability to control access for the sake of achieving desirable states of privacy can be exceedingly important'. Indeed Gavison (1980, p. 421) defines privacy as, ‘a limitation of others' access to an individual’.

On the other hand, Posner (1978, p. 393), considered privacy from an economic standpoint, suggesting that the concept is economically inefficient, whilst noting that ‘one aspect of privacy is the withholding or concealment of information'. Gobetti, looking at the historical progress of the public and private, acknowledges that traditionally the ‘distinction between private and public runs along the lines separating economic and the political domains' (1992, p. 6). Etzioni (1999), however, does not engage in such historical dichotomies, but argues for a balance between privacy and the needs of the community, and O’Hara (2010) argues that privacy should be seen not just as an individual right but a public good.

Other writers take a different approach. Lessig (2006, p. 231) considers that the right to privacy would be stronger if it was conceived more as a property right, and that ‘individuals should be able to control information about themselves'. Laudon (1996) went further and suggested the creation of personal information banks, which would allow individuals to deposit their data and be paid interest. Parent (1983, p.306) also viewed ownership as important stating that ‘Privacy is the condition of not having undocumented personal knowledge about one possessed by others'.

An alternate perspective defines privacy as ‘the right to live in a world in which our expectations about the flow of personal information are for the most part met’ (Nissenbaum, 2010, p. 231) rather than as a right to control the access to one’s data or to have the access to that data restricted as described above. Nissenbaum names this contextual integrity. It is based upon an individual’s expectations in relation to the norms

of information flow within a society and is closely allied to the concept of reasonable expectation of privacy. Gutworth (2002) makes an additional point, that privacy is also a relative, contextual concept, which is dependent upon not just norms of information flow but of the institutional, social, cultural, religious, historical and epistemological contexts within which privacy expectations are nurtured.

Many definitions of privacy have been offered and debated which is reflected by Post's view that:

> 'Privacy is a value so complex, so entangled in competing and contradictory dimensions, so engorged with various and distinct meanings, that I sometimes despair whether it can be usefully addressed at all' (Post,1963, p. 2087).

More recently, Solove, after more than 10 years of study, has termed privacy 'a concept in disarray' and that '(n)obody can articulate what it means.' (Solove, 2008, p. 1). For legal purposes, and Solove is a lawyer, this may be true, but the ideas of Lessig in *Code 2.0* are attractive and demand further attention. In particular, he suggests regulation of cyberspace in order to defend a space where culture can be shared, to empower individuals to control what is known about them, and to defend the individual from capitalist exploitation. which

Finally, the feminist critique of privacy should be acknowledged, proposing the line of privacy be redrawn from the economic and political, to the family and the rest of society (Gobetti, 1992). Such a critique may argue for transparency in that privacy itself is detrimental to women, as it can be used as a shield to cover, control and abuse (MacKinnon, 1991), suggesting in the extreme, an end to privacy. Elshtain (1993) takes a more moderate stance defending private life and the family, whilst promoting self-development and democratic participation.

Some of the many views on what constitutes privacy have been considered. The right to be let alone is a negative form of freedom whilst on the other hand the right to decide when, how and what data is shared is a positive freedom inferring control. Solove sums up the situation:

> 'Currently, privacy is a sweeping concept, encompassing (among other things) freedom of thought, control over one's body, solitude in one's home, control over personal information, freedom from surveillance, protection of one's reputation, and protection from searches and interrogations. Philosophers, legal theorists, and jurists have frequently lamented the great difficulty in reaching a satisfying conception of privacy.' (Solove, 2008, p. 1).

However, for the purpose of this work, the meaning of privacy will be considered from a positive standpoint and be limited to control over personal data (as defined in the previous

section) rather than physical or other forms of privacy. It concerns informational privacy, and as with other forms of privacy is contextual. As technology changes it affects the balance between privacy and disclosure, and its use by ourselves or others affects the level of control that individuals have over their lives.

2.3 Why does privacy matter?

Rachels (1975) writes that there is no single reason that privacy is valued by people. Rössler (2005) argues that the value of privacy lies in its protection of individual autonomy, which is a necessary condition for a rewarding life in a liberal democracy. This, he proposes is based upon decisional privacy, which enables an individual to claim, with proper justification, that a matter in none of the business of other people; informational privacy, control over what people can know about oneself; and local privacy, the right to live in protected spaces. The value of autonomy is in turn explained by Raz thus:

> 'The value of personal autonomy is a fact of life. Since we live in a society whose social forms are to a considerable extent based on individual choice and since our options are limited by what is available in your society, we can prosper in it only if we can be successfully autonomous.' (Raz, 1986, p. 394)

Bloustein (1964) on the other hand, suggests that the value of privacy lies in its protection for human dignity and independence in addition to autonomy. Others, however, argue that privacy's worth is as a safeguard for intimacy. For instance, Fried (1970) whilst defining privacy as control over information about oneself, additionally argues that it has intrinsic value and is fundamental in enabling people to develop intimate relations. Shoeman (1984) observes that privacy enables individuals to control intimate information about themselves, benefiting relationships with other people and allowing the development of one's personality, whilst Inness (1992) concludes that it is necessary in order for individuals to fulfil the need for loving and caring. This argument may be extended to cover the development of other forms of social relationships by both controlling information and also access (Rachels, 1975).

It may be argued that privacy gives an individual control over part of their lives and information, therefore breaches in privacy threaten that control, and thus threaten the individual:

> '… considering each individual as constituted by his or her information, and hence by understanding a breach of one's informational privacy as a form of aggression towards one's personal identity' (Floridi, 2006a, p. 111).

In arguments for greater disclosure of personal data it is often said that if you have nothing to hide you have nothing to fear, or if you keep within the law, you have nothing to fear. O'Hara and Stevens took an alternative viewpoint:

> 'If you keep within the law, and the government keeps within the law, and its employees keep within the law, and the computer holding the database doesn't screw up, and the system is carefully designed according to well-understood software engineering principles and maintained properly, and the government doesn't scrimp on the outlay, and all the data are entered carefully, and the police are adequately trained to use the system, and the system isn't hacked into and your identity isn't stolen and the local hardware functions well, you have nothing to fear' (O'Hara and Stevens, 2006, p. 251).

I would add, and if the record matching is accurate and the algorithms are sound, then you may have nothing to fear.

Each instance of data storage comes with its own risk. For instance, even a CCTV camera, which may seem benign and is expected to reduce crime, (although a report by Gill and Spriggs (2005) casts doubt on this), can be problematic. In 1994, a boy suffering from depression walked down the high street in Brent at 11:30pm, holding a knife with which he tried to kill himself. This was a private act at a quiet time of the day, and at the time there were no issues with knife crime. He was recorded on a CCTV camera and his image later appeared upon BBC television in a 'Crime Beat' programme, and in newspaper articles describing how CCTV cameras were being used to fight crime. The boy's identity was clear to those who knew him and he had been labelled a criminal, when what he had needed was help. The European Court of Human Rights found that under Article 8 of the European Convention on Human Rights (Council of Europe, 1950) Brentford Council had infringed the boys right to respect for private and family life[1].

Then there is the case of Maher Arar, a Canadian software engineer of Syrian origin. He was detained during a stopover in New York, based on information from the Royal Canadian Mounted Police which was inaccurate. He was deported to Syria, under a covert scheme for extraordinary rendition, and spent a year in jail where he was tortured. He eventually received C$10.5 million in compensation and an apology from the Canadian Government. The USA government accepted no responsibility (Abu-Laban and Nath, 2007).

[1] Peck v United Kingdom, no. 44647/98, §24, ECHR 1999-II

Individuals are risk assessed by algorithm not just to obtain credit but to obtain freedom, as police forces, and courts use scores from software such as Northpointe's COMPAS, or Durham Constabulary's HART applications. COMPAS in the USA was used in courts to help with sentencing, whilst HART is used in Durham, UK to help decide if a person is sent to court or to a rehabilitation program. Both systems make assessments on likelihood of the individual's re-offending, however, the algorithms are not transparent and appear to show bias e.g. against black people, or people from poor areas (Angwin et al., 2016).

Not all implications are so serious, it may just be that access to a website may be restricted as happened to Muhammad Khan when he was refused access to the online game Paragon because he was on the Specially Designated Nationals list, a US government blacklist (Hern, 2016).

As far back as the 17th Century Locke, Filmer, Grotius, and others insisted that life in the private realm is a prerequisite for life in the public sphere. This is reflected now in a submission from the Electronic Privacy Information Center to the US Office of Science and Technology Policy (EPIC, 2014) which states that the use of predictive analytics undermines freedom of expression, and affects people's ability to fly, obtain a job, get clearance for particular roles, obtain credit, and has a chilling effect on online interaction and participation. This is because of opaque algorithms that may consider a person's race, nationality or political views.

Finally, the inter-relationship between personal data and privacy still matters today because as Rosenberg asserted, privacy is never won but is always in conflict with civilisation. It must be defended again by each successive generation as 'we are continually changing our life environment; society may be altered so frequently that safeguards that in the past adequately protected our liberties become obsolete' (Rosenberg, 1969, p. 14).

2.4 Use of Data

The previous sections discussed personal data, privacy and why it matters. The following part of this chapter moves on to consider the importance of data, its value, where it originates and how it is used.

2.4.1 Importance of data

In both the U.S.A. and the UK business context, the driving objective has been to maximise shareholder value (Gamble and Kelly, 2001). Friedman (2007) goes further, stating that the object of a business is to maximise its profits whilst staying within the

‘rules of the game' and obeying the laws of the land, indeed a corporation has no social responsibility. Corporate executives are agents of the corporation and as such, should act without social responsibility unless they are acting on their own behalf. Other views suggest that corporate social responsibility can be attained without damaging ‘financial performance' (McWilliams and Siegel, 2001). Godfrey et al. (2009) however conclude from their research that corporate responsibility shown towards trading partners has little benefit, whereas that shown to secondary stakeholders (or society at large) seems to protect shareholder value if a negative event damages the company. Thus, companies strive to increase shareholder value and only hold back if actions will damage that objective. They do this through a series of initiatives and projects aimed at increasing profit, through taking in raw material and producing goods with added value. With respect to data, information is gathered in and used to enhance sales, increase profit margins, develop products and be sold at increased value.

The ability to analyse data is key to success (Marr, 2010) and for many companies, data are used to segment their customers so that marketing can be more specific and thus cost effective (Tapp, 2008). Recognising the power and value of data, Tesco, for example, acquired the remaining stake in Dunnhumby Ltd in April 2010 (Kathryn R, 2015). Dunnhumby are the market research firm which holds all Tesco Clubcard data and, reputedly, data on all UK residents (Tomlinson and Evans, 2005).

Technology progress has also provided opportunities for new companies based on, or leveraging, data. Google was worth around $2.3 billion (Stockport, 2010) in 2010 and in 2018 as Alphabet, its market capitalisation is $726 billion (Archer, 2018). Amazon, on the other hand, originally sold conventional products, its power lying in having virtual shops, and until recently no real estate costs, but also in the ability to target recommendations based upon customers’ previous actions, their demographic information, and the actions of other customers whose activities have similar patterns (Pathak et al., 2010). The company now streams music and videos and has exceeded its expected growth in its Alexa product. It has a market capitalisation of $710 billion (Gill, 2018). Twitter (market capitalisation £11.3 billion (Lucas, 2018)) feeds are analysed for marketing, and social purposes (Cheong and Lee, 2009). Facebook can target adverts very specifically (Vogelstein, 2009) and this data driven social network platform is valued at $519 billion (Woodhouse, 2018). The competition for data relating to the views, actions, and needs of individuals is intense. This may best be shown by examining Google, with their quest for data whether it be web surfing details, books, email, application data, and street images. Compare this with Facebook, with its repository of messages, individual statements, and photographs (many

with geolocational data). There is intense rivalry for the ownership of individuals' data, with Google precluded from searching and indexing the Facebook data stores. These systems concern detailed knowledge of individuals and also general trends (Marsden, 2010). It is the detailed knowledge of the individual, which enables more highly targeted, effective marketing, and in turn results in higher sales at lower advertising costs.

Compare these recent data-based organisations with ‘conventional’ companies for example General Electric once the largest of organisations, currently has a market capitalisation of $155 billion (Crooks, 2017) and Marks and Spencer $6.9 billion (at March 2018 exchange rates), to gauge the value that can be generated from the capture, curation and analysis of data.

Baker (2008) describes the analysis of data, producing slivers of knowledge regarding individuals. It is the accumulation of these slivers of knowledge combined with the base information, which has changed the nature of data, and resulted in the remarkable growth of the organisations described above.

Governments have different responsibilities than private organisations, their aims are to protect the nation, provide services that public organisations cannot, and to invest in citizen capabilities (Slaughter, 2017). Data provides value to government by helping with these responsibilities. For instance, in order to protect national security, GCHQ collects and analysis vast amounts of data, and continued to do so after the Snowden revelations and indications that it had been doing so illegally (Steiger, 2017). Recommendations to resolve the social care crisis in the UK were made on the basis of statistical analysis of large amounts of data (Commission on Funding of Care and Support, 2011). HMRC uses analytics to assess risks thus allowing better decisions to be made by UK Customs officials (Okazaki, 2017). HMRC also analyses data from banks, credit cards, land registry, DVLA, social media, online market places (e.g. AirBnB, and ebay), web browsing, emails, and financial information from over 60 countries to feed the Connect system in order to improve the collection of taxes and increase revenue for the UK government (Maciejewski, 2017, Suter, 2017). In Los Angeles and New York, analysts have been used to help reduce crime and increase public safety and in Boston smartphone data was used to find potholes in streets so they could be repaired (Desouza and Jacob, 2017). The range of governmental use of data is wide and helps to create the society within which we live.

2.4.2 Value of data

Data therefore has value to both the private and public sectors. The price that business places on data may be observed in at least three ways. As Tech Crunch reported, it is

possible to note the compensation received by people when their data is revealed. Comcast paid $100 to each person whose data they had disclosed even though the individuals had paid a fee to ensure the information was kept private (Glikman and Glady, 2015). Second, data may be valued by acquisition. In 2014, Facebook bought WhatsApp for $21.8 billion thus paying $55 for each of the 600,000 users (Deutsch, 2014). When Facebook had earlier bought Instagram in 2012, the valuation was $28.57 (Statista, 2018b) and when Microsoft bought Minecraft in 2014 the valuation per user was approximately $46 (Miller, 2014). However, the price per user is often based on the value of the underlying technology to the purchasing company, and so there may be variations as can be seen from figure 2.1.

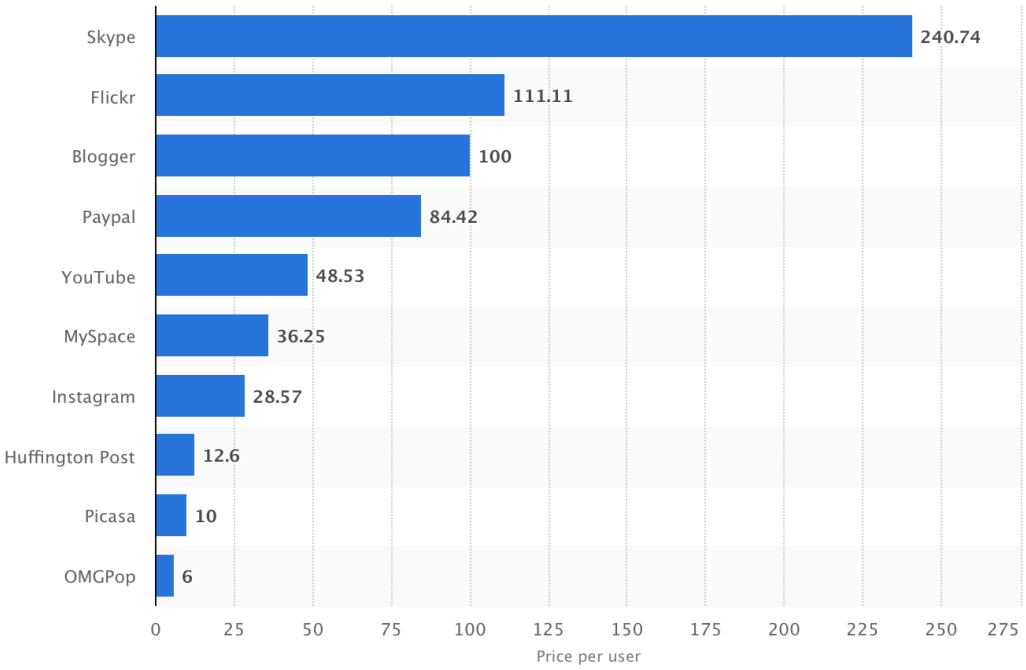


Figure 2.1 Acquisition cost per user in $: Source (Statista, 2018b)

Third, data brokers trade in personal information. For general information about an individual (age sex, and locality) $0.0007 is charged. It is more expensive to get more specific data, for instance individuals suffering from a specific disease may cost in the region of $0.30 per name (Glikman and Glady, 2015). Of course, the magic of data is that when it is transferred or sold, the seller still has the data (unlike selling groceries) and so brokers will sell the same data many times.

On the other hand, each individual may value their data differently, studies having shown that difference types of personal data have variations in perceived risk (Robinson, 2017a). Research on how people value categories of data was undertaken in Korea and showed that, across the sample, the highest value was placed on basic personal information, medical information was the next most valued and purchase list and payment data third (Lim et al., 2018). However, variations were noticed with differences in the value

attributed to medical information, with those who had experienced some form of privacy leak, ranking it more highly than others. Robinson (2017b) argues that more sensitive or risky items should hold higher monetary value than other data. However, Li et al. (2010) observed that, when offered monetary rewards in exchange for personal information, people are less likely to disclose it than when data thought to be relevant for a transaction is requested. People may not be able to act as economically rational agents when it comes to personal privacy and are unable to assess, or are unaware of, the value of their data (Acquisti, 2004).

2.4.3 Sources of data

Data are stored as a result of three possible actions, the data are entered by some means, are replicated or are the result of analysis of other data. A series of reports from IDC have reported on and forecast the amount of new data created each year since 2006, when it was estimated that the amount of digital data created, captured and replicated in that year was 161 exabytes (or 161 billion gigabytes). As these data were equivalent to about 3 million times the information in all the books ever written (Reinsel, 2007) it raises the question of where that data came from. The 2007 report estimated that ‘(b)etween 2006 and 2010, the information added annually to the digital universe would increase more than six-fold from 161 exabytes to 988 exabytes' (Reinsel, 2007, p. 1). Of this it was estimated that 25% would be original data (the rest being replicated), and between 25% and 30% would be data created in the ‘workplace' (the rest being mostly digital photographs, videos and TV signals and pictures). The updated document for 2010 reported that 800 exabytes had been created in 2009 and that 1,200 exabytes were expected to be created in 2010 (21% higher than expectation). It was estimated, that in 2010, 70% of data was generated by individuals rather than companies (Gantz and Reinsel, 2010, p. 10). The 2012 Digital Universe document (Gantz and Reinsel, 2012) reported annual data creation of 2,759 exabytes of which 23% would be suitable for analysis (although only 3% was suitably tagged). The projected data volume for 2020 was 13,000 exabytes. Again, this was an underestimate. The 2017 report states that 16,100 exabytes of data were created in 2016 (exceeding the 2020 estimate by over 23% four years early). The estimate for 2025 is 163,000 exabytes, driven by the increased internet connected devices. It is estimated that the connected person will interact 4,800 times a day with some connected device, that is once every 18 seconds (Reinsel et al., 2017). What is the point of listing these data points? It is to illustrate the colossal amount of data created each year of which more than half is generated by individuals. It is a lot of personal data.

Figure 2.2 illustrates the increasing trend for data created each year (1 zettabyte is 1,000 exabytes). It is estimated that by 2025, 5,200 exabytes of data will be analysed each year of which 1,400 exabytes would be touched by cognitive systems.

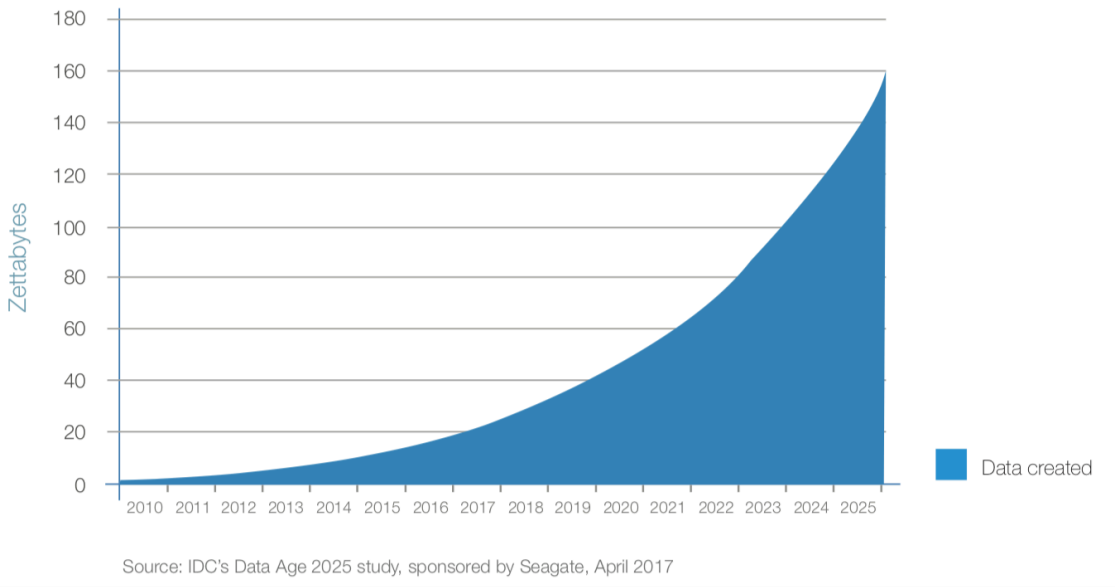


Figure 2.2 Estimated growth in data from 2010 to 2025 (Reinsel et al., 2017)

Increasingly, personal data is being collected from individual's interactions with social media, personal tracking devices and the Internet of Things. An estimation of internet activity in an average minute during 2017 from GoGlobe (2018) is indicative of the scale of the data available for collection, curation, matching and analysis, and gives some insight into just some of the sources of personal data;

- c. 700,000 hours of videos watched and more than 400 Hours of videos uploaded on YouTube,
- > 3.8 million searches on Google,
- > 243,000 photos uploaded and 70,000 hours of video content watched on Facebook,
- > 350,000 tweets sent on Twitter,
- > 65,000 photos uploaded on Instagram,
- > 210,000 snaps uploaded on Snapchat,
- > 156 million e-mails sent,
- > 29 million messages processed, 1 million photos and 175,000 video messages shared on WhatsApp,
- > 25,000 posts on Tumblr

- c. 16,550 video views on Vimeo
- > 500,000 apps downloaded
- > 1 million swipes and 18,000 matches on Tinder
- > 2 million minutes of calls done by Skype users
- > 800,000 files uploaded on Dropbox

Much of the personal data enumerated above is the result of consumer commodification, people are, more and more, the products, not the consumers, of the data broker industry and commercial surveillance. The basic asymmetry of power between individuals and providers of applications is the cause of this, and will continue, if left unchecked (Crain, 2018).

Many people now understand the phrase 'if you are not paying you are the product' but in reality, the user is always the product. Publisher websites could be used to illustrate this, most are small businesses such as The Family Cookbook Project (http://www.familycookbookproject.com). These websites normally use third parties to provide facilities like search, advertising, social media links, payment services, account management, video hosting, and comment, all of which collect personal data. As a result, an individual who is paying for a product and may expect their data to be curated by the publisher website would find that it was all collected by many other organisations (Gopal et al., 2018).

Self-quantification is another source of rich data. People are increasingly wearing items, such as step counters, digital watches, life logging cameras, and carrying mobile devices with self-quantification apps. These devices allow individuals to monitor health (e.g. heart rate, exercise, breathing, food, sleep); work productivity; and leisure activities such as travel and music. Cameras can also be worn that will automatically take photographs to diarise the day. There are privacy issues associated with many areas of personal data but perhaps none more so than with these devices. Despite this, wearables are still used, and it is expected that 245 million of these devices will be sold in 2019 despite individuals having privacy concerns (Maltseva and Lutz, 2018).

In summary, this section has shown the importance of data, for both private and public organisations, and presented examples of how personal data has been valued before illustrating the large amounts of data created each year and giving an indication of its origin. Finally, two sources of personal data were discussed as examples of more recent trends in personal data collection. The following part of this chapter moves on to examine the threats to privacy resulting from the collection, curation, matching and analysis of personal data.

2.5 Threats to privacy

This section examines a number of threats to privacy relative to personal data. It starts with the argument that privacy is beyond saving, before looking at why people appear not to value privacy and the asymmetry of power between the individual and large corporations. Next the increasing amounts of data, datafication and repurposing of data are covered before a discussion of liquid surveillance and finally the risk from criminal activities.

In the last 20 years, the way people live has changed, in many countries, due to the pervasive use of digital technologies. Whether it is shopping, exercise, health, social arrangements, what they think, what they like, or what they own, people record the minutiae of their everyday lives. This is augmented by third party actors recording payments, tracking digital identifiers, and videoing movement. These are some of the circumstances of our lives that give rise to privacy concerns and have led to the belief that there can be no privacy within the digital universe.

In a controversial statement in 1999 Scott McNealy stated 'you have zero privacy anyway, get over it' (Sprenger, 1999, p. 1). Others argue for reciprocal transparency. Brin (1999) accepting that technology advances would lead to ever increasing surveillance (and lack of privacy) called for equal transparency amongst the public and private sectors, effectively a call for sousveillance (Ali and Mann, 2013). Schneier (2008) however, discounted this proposal arguing that the power imbalance is too great between the individual and organisations, calling for individual privacy but openness within organisations. Power asymmetry is discussed below in this section.

Big data exceptionalism takes another approach, arguing that big data and privacy are mutually exclusive, before suggesting that collection of personal data should not be regulated as it is inevitable, rather its use should be controlled. It is contended that this approach may make regulation easier, as when data is collected it is not possible to know how it may be used in the future, but also not exploiting big data will be costly to society as a whole (President's Council of Advisors on Science and Technology, 2014). On the other hand, Nissenbaum (2017) having dissected the arguments, concludes that efforts to strengthen both collection and use should be sustained along contextual lines.

But why strengthen privacy legislation if people do not value privacy? Research in Germany suggests that people seem to appreciate privacy as a general good and see it as a benefit to society. However, when it comes to data that are 'invisibly' collected such as location, or usage data, they tend to lack awareness (Vervier et al., 2017). In the USA, Madden and Rainie (2015) report that many citizens want control over their personal

information but are not sure that those who collect their personal data can, or will, keep it secure.

There is a paradox in the way that people talk about privacy when asked their views and the way that they behave in practice. A field experiment reported in 2012, showed that people were not willing to pay for privacy, despite 95% of them indicating that they were interested in the protection of their personal data in a post experiment questionnaire (Beresford et al., 2012). This was labelled the privacy paradox by Barnes (2006) when explaining that teenagers freely give up personal information to social networks but are outraged when parents read their journals. Surveys still show that privacy is a primary concern for citizens but, on the other hand, they disclose personal information for scant reward often to draw the attention of peers (Kokolakis, 2017). In a systematic review of 32 papers exploring 35 theories that may explain the privacy paradox, Barth and de Jong (2017) came to no clear conclusion recommending that the subject deserved more research attention.

One explanation may be found in the work of Binns et al. (2017) who argue that people base their privacy related decisions on pre-existing conceptions of and relationships with organisations supplying apps, devices, and services. In particular, impressions regarding size, level of regulatory scrutiny, relationships with third parties, and pre-existing data exposure may result in people selecting solutions with a lower privacy potential.

Another explanation is contextual in that whilst people may normally practise a privacy trade-off, when it comes to downloading mobile apps they are less likely to do so. Highlighting the intrusive nature of an app and evoking privacy concerns (on the app permissions screen) may decrease acceptance but does not work well for desirable apps. Given that people download apps because they perceive them to be of value, Wottrich et al. (2018) consider the current privacy regime to be ineffective and recommend changes in regulations to limit their data collection.

Finally, it is suggested that media coverage on internet-enabled services tends to emphasise the benefits but minimise any negative aspects, such as diminution of privacy. By legitimising the exploitation of personal data, it has been normalised (Cichy and Salge, 2017). In addition, the move towards greater digitisation has blurred the contextual boundaries. For example, Apple a well-respected brand (see Binns et al. above) has moved from the context of consumer electronics into music streaming and more recently with HealthKit into analysing health related data for medical research. People who value Apple as a trusted provider of consumer electronics may then transfer that trust to wearables collecting health related data.

The increasing difficulty people have in understanding privacy policies has led to privacy fatigue. In addition, the loss of control associated with the seemingly impossible task of managing personal data, combined with successive data breaches being reported (Hunt, 2018), leaves individuals weary of having to consider their on-line privacy to a state where they do not bother (Choi et al., 2018). It may be considered then that the normalisation of digitisation, combined with social environments, misplaced confidence, the need for immediate gratification, and despair with the practicalities of protecting one's privacy, has led to what is known as the privacy paradox and the increased lack of privacy within the digital universe.

Another reason for privacy fatigue may be the power imbalance related to datafication. Andrejevic (2014) points out the differences in capability between those who collect and mine data and those whose personal data it is. One 32-year-old male epitomises the powerlessness of the data subject:

> 'you end up accepting having no privacy without knowing the consequences' (Andrejevic, 2014, p. 1685).

Crain's case study of the data brokerage illustrates that the commodification of personal information is deeply entrenched and that unfairness between organisations on the one hand, and the individual on the other, will continue, that 'it is not a glitch in the system it is the system' (Crain, 2018, p. 100). In this situation, transparency, the current policy for mitigating the harms of the internet in the USA, is not working.

Power asymmetry may also be observed in Google and Facebook both of whom frame themselves as committed to the human rights of freedom of expression, and privacy. However, they focus their efforts on external actors such as governments and ignore areas where their own actions impact their customers' rights and freedoms. They use their power to set and enforce their own rules of engagement, at the expense of their users (Jørgensen, 2017).

Another example may be observed, the more nuanced issue in relation to online censorship. Here it has been found that the most marginalised communities are the ones most likely to suffer from on-line censorship. Research into takedowns from six social media platforms (including Facebook and Twitter) show that they disproportionately affect minority groups (Anderson et al., 2016).

Social media platforms also exercise their corporate power in another veiled way. Privacy settings are available for individuals to control access to their data, and user 'demonstrations' in the past have persuaded social media companies to allow greater control over privacy (Sanchez, 2009). The gains made are in protecting a person's

information from the gaze of others. A private group shares information between itself safe from the gaze of other individuals, hence protecting their data. The organisation has thus obscured an underlying privacy issue (Pieters, 2017). By providing a personal room in which private interactions may take place, people fail to notice the camera and microphone in the room observing their interactions, as the social media organisation takes the data for its own use.

Finally, there is an asymmetry of power between the tech giants and national governments. Personal data moves around the globe through cables, wireless, and satellite communication channels. It is kept in large data centres that service organisations and provide cloud storage to all. It does not respect national boundaries. As a result, large multinational organisations that manage our data are in the position to mediate competing governmental demands and approaches, and so are able to determine the rules (Daskal, 2018).

The increasing amounts of data, discussed above, threaten privacy. As the data universe expands we lose control of our personal data, each year people contribute more to their digitally extended self through social networks, on-line purchases, cloud storage etc. (Wiese et al., 2017) thus, putting more of their personal data and privacy at risk. Vendors obtain more and more information and hackers and third parties who gain access, put us at risk (Will et al., 2017).

New technologies and the Internet of Things drives this increase in data, as mentioned above. One such emerging technology is the connected vehicle. Already vehicles collect data (Swan, 2015) but they are increasingly connecting with each other and with public networks. Using an example of car sensors Dötzer (2005, p. 200) observes:

> ‘A very dangerous and often ignored fact about privacy is that innocent looking data from various sources can be accumulated over a long period and evaluated automatically.’

Cars are personal devices, usually long-term purchases, but even when hired are associated with an individual. They increasingly store large amounts of personal data, which when combined with other data, can identify the driver and be associated with them (Akalu, 2018).

Large amounts of data are held by a relatively small number of organisations and is used in the very large part for monetary gain or governmental purposes, and it threatens the privacy of most people. In addition, there is little sign that the data will be used for the common good. Governmental bodies make data open for use by companies who then use it to boost profits, for example big pharmaceutical companies use NHS data for analysis in

order to create new cures (Kaplan, 2016). The search for cures is thus being driven by profit rather than by pressing societal need. Nissenbaum (2016) argues that unless there is a change in regulation it is illusory to think that data will be used for the common good, and that there is little hope in any ability to uncover and regulate uses harmful to society. Value creation in the digital economy, which is generating so much data, depends upon datafication, the creation of the digitally extended self by transforming aspects of people's lives into quantified data (Mayer-Schönberger and Cukier, 2013). The links between the impacts of datafication, wealth creation and surveillance are clear (Gawer, 2017) and a threat to privacy. The data obtained from on-line users is turned into behavioural predictions and monetised through markets unavailable to individuals, an example of this is found in the Google business model. Zuboff (2015) calls this institutionalised model of value creation, surveillance capitalism, (a specific form of informational capitalism (Castells, 1996)).

The collection of data for one purpose and using it for another is known as repurposing. This issue is at the heart of Nissenbaum's contextual privacy (Nissenbaum, 2010). For instance, an individual may be happy to disclose where they spent their night in the context of a national census, but less so in discussion with their spouse at a later date. Whilst census data is not disclosed to individuals until 72 years has passed it is, however, used for research purposes and Heeney (2012) argues that this breeches the contextual integrity and categorical privacy (Vedder, 2000) expected by the participants. This may be observed in online applications such as Facebook, Waze, and Flurry whose services are used by publisher websites (discussed above). Here the services are re-used and generalised privacy agreements allow data to be used in situations the user would not expect (Breaux et al., 2015). This should not be unanticipated as data repurposing is necessary for big data analytics (Custers and Uršič, 2016) and is the basis for the data broker economy, with its already discussed effects on privacy. Law enforcement agencies have an appetite for data repurposing, for instance Google has received increasing requests to disclose user data. Figure 2.3 illustrates this, showing the number of data requests and individual accounts affected from 2009 to 2016, the latest information available. Here the data originally submitted as a search term is repurposed into law enforcement data.

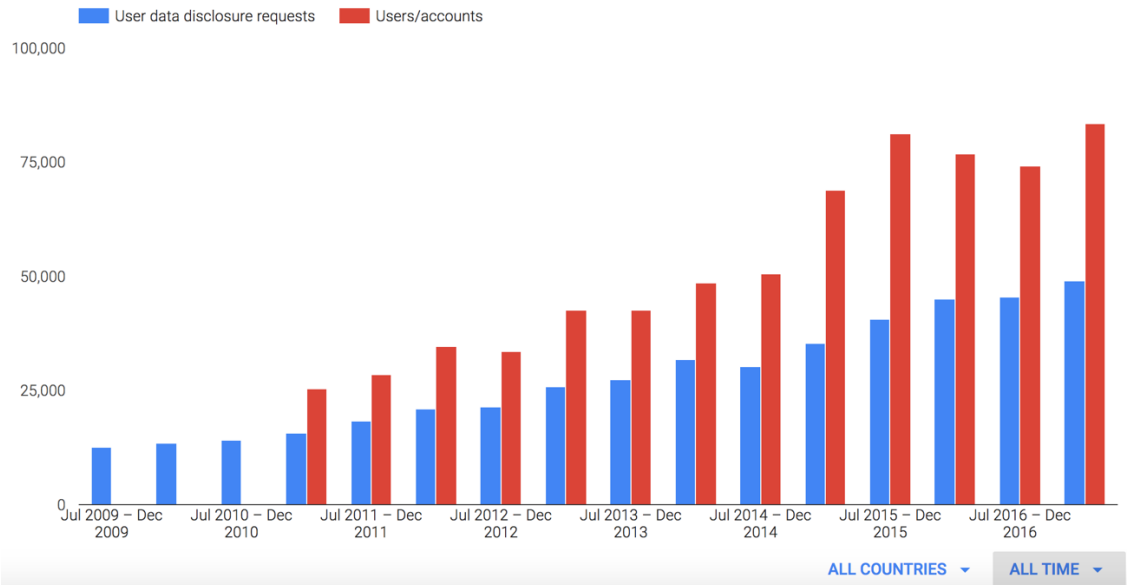


Figure 2.3 Google requests for user information from government authorities (Google, 2018a)

There are many examples of data repurposing for law enforcement reasons which can be used to illustrate this growing trend, for example, passenger name records (air traveller data), ISP telecommunications data, HMRC records, bank account details. These may be examples of repurposing for the public good, but the data is still being used for purposes which were not the original intention of the user. Two European Court of Justice rulings, *Digital Rights Ireland* and *Tele 2 Sverige* indicate that the legislation contained in Directive 2016/680 lack the essential provisions to guarantee the individual's right to data protection in cases of repurposing for law enforcement reasons (Jasserand, 2018). Pervasive computing, or the Internet of Things refers to the practice of embedding computational capability into everyday objects which are network connected and constantly available, e.g. Nest thermostats, Ring doorbells, Samsung fridges, Hue lights. This has been discussed previously as a source of large amounts of data. These devices also create a privacy risk as always on sensors monitor users 'offline' activities and transmit data about these outside the home (Apthorpe et al., 2017). The linking and analysis of data from these devices, for instance to ensure quality of service, provides insights into user characteristics which again lead to severe privacy concerns (Madaan et al., 2018). Outside the home, companies such as Tamoco use 1.1 billion proximity sensors, associated with Wi-Fi hotspots, to track the movements of 100 million smartphone users (Manthorpe, 2018). Opportunistic networks use a similar capability to build, for instance, ad hoc geo-social networks between users over Bluetooth or Wi-Fi connections. As Zakhary and Benslimane (2018) state, this creates many unique privacy related challenges. The pervasive nature of computing has been discussed between Bauman and Lyon (2013). It follows Bauman's concepts of liquid versus solid, as the postmodern world becomes

more diffuse and harder to pin down, unlike the more solid expectations of modernity. Liquid surveillance is the term used for all forms of dataveillance, the dispersed and mobile watching of ourselves and others enabled by the technologies that generate the large flows of personal data described above. Bauman considers that such is the pervasive nature of dataveillance that the inspectors of the panopticon (Bentham, 1995) can now slip away as we effectively monitor and control each other. '[S]urveillance is seeping into the bloodstream of contemporary life' (Bauman and Lyon, 2013, p. 152).

However, mutual control is insufficient for governments as they seek to harvest more data about their own and others populations in order to fulfil their prime role of providing security and safety to their citizens. This is helped by the pervasive nature of computing noted above. Trump's travel ban on people from some Muslim countries is an example, as it masks a requirement to provide data to the US authorities on all people travelling to the USA (thus putting the EU - US privacy shield under strain) (Guild et al., 2017). The collection of data that would in other circumstances be seen as unjustified is normalised by the use of fear (Svendsen, 2008). Mass surveillance of the US population has been justified in this manner. However, fear is not rational, which is illustrated by comparing deaths from terrorism with those from homicides by firearms in the USA. In 1999, terrorists, five of whom were American, killed 233 people worldwide; in 2000 the figures were 405 people killed of whom 19 were Americans; and exceptionally in 2001 there were many more deaths due to the 9/11 attacks (where about 2,670 U.S. Citizens were killed). By comparison, many more people were murdered in the USA by firearms alone, 1999, saw 10,128 homicides by firearms, in 2000 there were 10,179, and 11,106 in 2001 (Richardson, 2007). Surveillance derived from database information is not new (Lyon, 1994), but has increased with the concern regarding terrorist attacks discussed above (Grayling, 2009). Fear is used as a reason for data collection and is also therefore a threat to privacy.

Dataveillance is the underlying cause of liquid surveillance. It is normally associated with the powerful observing the weak. However, it is facilitated by self-surveillance, and the mutual observations of the weak, collected, curated, compared and analysed by the strong. It is augmented by other actors who monitor the traces people leave within the digital universe perhaps through passing CCTV cameras, or digital sensors. A variety of actors use this information. Sometimes a person makes use of their own data, usually though the data is exploited by other entities such as commercial and research organisations, government, hackers and cyber-criminals thus threatening privacy (Lupton and Michael, 2017). There is now evidence of increasing public unease about how people's data is used. Whilst there is an appreciation of the public good that can be accomplished in the fields of

security, crime prevention, public health and healthcare there is hostility towards the idea that government should sell big data to the private sector. Very few people consider they have control over how their data is collected, and used, and are concerned about their privacy (Lupton and Michael, 2017).

The most concerning issues for people using the internet in the USA is cyber-crime. As can be seen in figure 2.4, cyber-crimes are by far the most concerning issues for U.S. Internet users according to Statista reporting a YouGov poll.

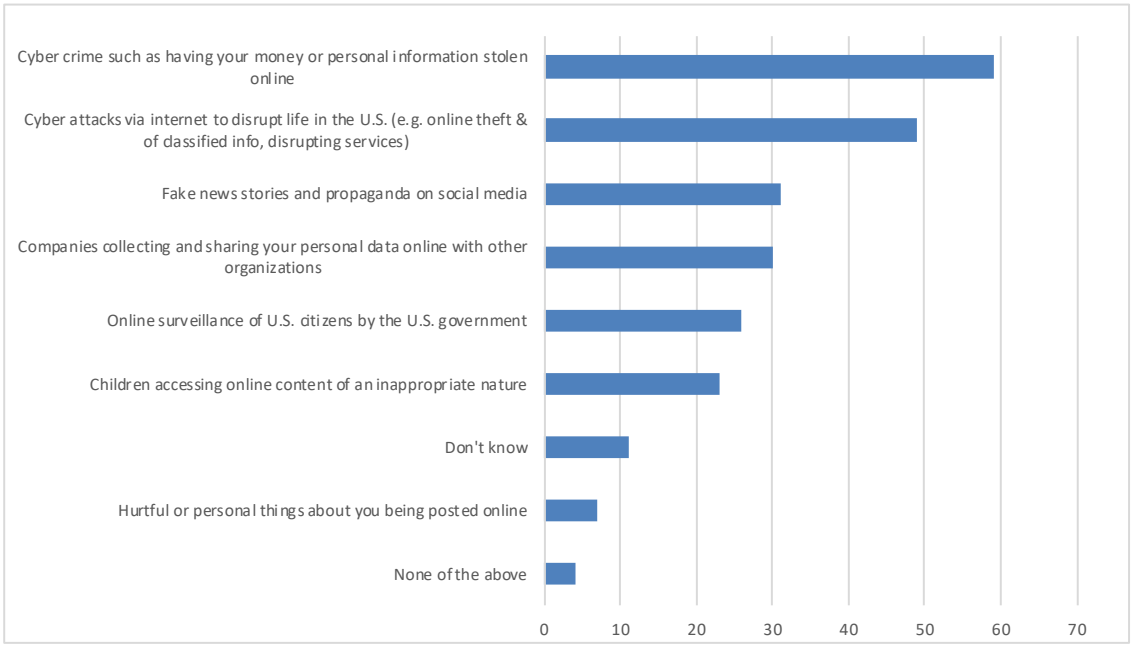


Figure 2.4 Most concerning issues about online usage according to internet users in the United States as of May 2017 (Statista, 2018a)

Cyber-crime may be split into four categories (Ngo and Jaishankar, 2017), cyber-deceptions and thefts, e.g. credit card fraud, and piracy; cyber-trespass, e.g. hacking, defacement, and viruses; cyber-violence e.g. hate speech or stalking, and cyber-pornography; the first two of which cause most concern as shown in figure 2.4, perhaps due to wider awareness.

Thus, threats to privacy relative to personal data arise from a number of sources: a belief that privacy is beyond saving; the lack of value attributed to privacy in certain contexts; the asymmetry of power between the individual and large corporations; the increasing amounts of data; datafication and repurposing of data; liquid surveillance, and finally the risk from criminal activities.

2.6 Beneficial use of data and trust

The previous section discussed some of the areas associated with personal data that may be considered threats to individual and group privacy. The following paragraphs will examine

some of the beneficial aspects to personal data, its collection, curation, matching and analysis; and some associated issues.

Databases of information do not arise accidentally but rather from well-formed projects, which will have the objective to ‘deliver beneficial change' (Turner, 1998, p. 18) from the perspective of those sponsoring the projects. Lovelock and Farhoomand (2000, p. 773) state that one of the principal reasons for building databases is ‘[t]o build knowledge by accumulating a company's individual experiences'. Given the Foucauldian relationship between knowledge and power and its association with governmentality (Foucault and Gordon, 2002) the beneficial aspects of creating such databases, from the viewpoint of central or local government organisations, can be appreciated. Very often the creation of a database is a by-product of, or a tool to fulfil, the overall objective of a project. For example, it is widely understood that electronic health record systems enable the ‘delivery of sustainable, high quality health care' (Robertson et al., 2010, p. 1). As a result, in order to accomplish these beneficial objectives, the NHS decided that the best way to progress was to create a central repository of data known as the Spine, comprising four main databases and a centralised communications service. However, as more and more citizens generate their own health data, for instance from wearables, smart phones or smart scales, an issue arises. The new data combined with existing NHS curated data has enormous scientific value (Wilbanks, 2014), but falls outside existing health data protections. In order to reap the benefit of combining and analysing these data changes to regulation and practice are needed.

Whether and how something is beneficial however, is very subjective, and depends upon the frame from which it is viewed. Identity management is one example of this. From a government perspective, there are good reasons for introducing an effective identity card system, for instance to improve border control or reduce fraud within the welfare state. However, as Crompton (2010) points out, if it is seen by the populace as a policing action, or that power lies in the hands of government to collect additional information which is then linked, used or disclosed, then it is perceived as non-beneficial by the populace and so loses public support. Anderson et al. (2009) conducted a survey of public sector databases and reported on the use of the 46 most significant. Of these only six were found, in the eyes of the investigating academics, to ‘have a proper legal basis for any privacy intrusions’ that ‘are proportionate and necessary in a democratic society' (Anderson et al., 2009, p. 2). Here then is the dilemma, on the one hand the democratic state creates databases for beneficial purposes, but, on the other hand they cause unnecessary intrusion and possible harm in certain circumstances.

The commercial use of data is also associated with the subject areas of trust, identification and surveillance, as trust is needed for relationships between individuals and groups (Hardin, 2002) and it is information about others which helps to develop, or destroy, trust. In sociology, the role of trust in social systems, which is beyond the scope of this review, has been examined for instance by Luhmann et al. (1979), Barber (1983) and Giddens (1991), whilst Misztal (1996) brought together previous work in a view of trust as an element of social cohesion. However, the area of trust, which is of interest, is that which mediates between the individual and the organisation. This is similar to trust within business-to-business dealings (Bachmann, 2001, Friman et al., 2002, Lane and Bachmann, 1998).

Trust is the basis for a transaction, and arguably the bedrock of western democracy (de Durand, 2008) although Cook (2001) argues that trust, or negotiating the lack of it, is key. In this context, the individual will want to know that organisations will look after data entrusted to them, and also that they can be trusted to deliver their part of the transaction (Sirdeshmukh et al., 2002). In the case of a web-based transaction, where high street presence is often missing, the issues are more complex and were therefore subject to early analysis (Hoffman et al., 1999). It was also discovered that the nature of the web interface can affect the level of trust, for instance whether or not it conforms to ‘normal' standards (Gefen et al., 2003). On the other hand, organisations need to trust the consumer or client. In order to do this, they need to ‘know' the individual in enough depth so that sufficient trust is formed (Mansell and Steinmueller, 2002). For online transactions this is particularly difficult, see (Mansell and Collins, 2007). An early observer of these issues was Chaum (1985) who suggested the use of ‘card computers', similar to the chip and pin cards in current use, but which were to be intermediated by certification authorities. He suggested that failure to intermediate could lead to a ‘big brother' society, a concern realised when the U.S. Patriot Act required financial institutions to gather more information from their customers (Regan, 2004). Similar legislation was enacted in the UK, which, whilst encouraging greater identification and the collection of valuable customer information, was seen by financial institutions as intrusive and unhelpful, and so completed in a meaningless way by a ‘tick box' approach (Robinson, 2004).

Governments also need to collect information in order to improve decision-making and inform debate (Pullinger, 1997). Information about citizens is necessary, for instance for taxation, conscription or crime control purposes (Lyon, 2009) or to control borders (Torpey, 2000). However, as both Lyon and Torpey illustrate, it is a short step from identification to surveillance and to control. The work of Bentham (1995) in his

Panopticon Writings and Foucault's Discipline and Punish (1991) are informative in relation to surveillance and control.

One already discussed antidote to control is transparency, and an aid to achieving that is open access. In the UK and the USA open data initiatives have made large amounts of data, previously held in government silos, available for public access at data.gov.uk and data.gov respectively. As Shadbolt et al. (2012) argue, this has enabled citizen centred service delivery, through using semantic web standards in open government data. One early product of this is the crime analysis website (http://apps.seme4.com/see-uk/#/crime/by-population/ward/41907) which shows crimes per 1,000 people (normalised by population) for all of the UK at constituency level. A more recent initiative is Ben Goldacre's Evidence Based Medicine Datalab at the University of Oxford (https://ebmdatalab.net) which in 2016, combined prescription data with GP surgery information to highlight variations in prescribing practice.

However, much of the data universe is not made available for open access. For instance, the Australian Government has ambitions to release data but is hesitating to do so because of doubts surrounding de-identification which is required to comply with privacy legislation, but also because the public service culture in Australia is yet to embrace the open data movement (Hardy and Maurushat, 2017). Of course, much of the data resides in commercial data stores which monopolise both the data and the analyses derived from it. Open data proponents argue for open access to all data so that interpretations of data can be challenged and debated for the public good (Gawer, 2017).

This section has discussed some of the beneficial aspects related to the use of personal data and the issues of trust associated with it. Use of data can be beneficial or harmful, and the outcome may not be clear. Solove (2013, p. 1890) provides an example. A person over a 10-year period lays down 50,000 pieces of data, this is collected, curated, compared and analysed by many people, perhaps invading his privacy. The next day a relatively innocuous fact gets combined with the other data and shows that the person is at risk of contracting a highly lethal and contagious disease, and so the individual's life is saved. Alternatively, another piece of information could have been analysed and proved harmful to the person. The point is that individuals do not have a way of knowing if the data is going to be beneficial or harmful.

2.7 Initiatives to protect privacy

This section examines a number of ways that the privacy of personal data may be protected. It starts by briefly discussing eight general approaches before briefly discussing European legislation and finally, market place solutions.

Eight basic approaches to protecting the privacy of personal data may be identified, transparency, purpose limitation, privacy self-management, the right to be forgotten, data amnesia, anonymity, safe havens, and privacy by design.

Transparency has been discussed above but is mentioned here as it is currently the favoured approach to personal data privacy controls in the USA. Brandeis' famous phrase, 'sunlight is said to be the best of disinfectants; electric light the most efficient policeman' (Brandeis, 1913, p. 10) highlights the role of transparency as a mainstay of liberal democratic values. As Crain (2018) points out, it is a commendable concept but there are structural obstacles within the commercial surveillance economic model. He shows in his case study of the data broker industry that the way personal data is treated, as a commodity, lies at the heart of power asymmetries. Data brokers could not give control of data to individuals without a major re-structuring of their industry. It is this power imbalance that transparency is expected to equalise, but Crain states that it runs up against insurmountable obstacles in doing so.

Purpose limitation requires that data is collected for a defined purpose and that it should not be used for other purposes at a later time, without consent (Rauhofer, 2014). There are at least two reasons for this. First there is a formal, or informal, contract regarding the reason for data being collected and also its later use. Second, data collected for one purpose it validated for the purpose, so when it is repurposed it is more likely to have erroneous data within it. There are a number of international privacy devices that include purpose limitation, OECD guidelines, Council of Europe Convention, Asian-Pacific Economic Cooperation (APEC) Privacy Framework, EU Data Protection Directive (Greenleaf, 2012), and the EU General Data Protection Regulation (GDPR) which replaces it. However, the GDPR allows for member states to show some divergence from this principle which may weaken the instrument by authorising exemptions under local laws (Rauhofer, 2014).

The objective of *privacy management, and informed consent*, is to give the power over personal information to the individual concerned. There are, however, practical and cognitive issues associated with this approach. People deal with many organisations making the task of managing data separately for each environment very difficult, also data collects slowly over time within the data universe, it may then be collated unknown to the

individual, making decisions about the costs and benefits of releasing it practically impossible (Henttonen, 2017). Solove (2013) points to four additional cognitive issues. Privacy policies are too long and complex, so do not get read, if they are read they are not understood, even if they are read and understood people do not comprehend the context enough to make an informed choice, and finally, if they are read, understood, and an informed choice has been made, their choice may be have been skewed by other decision-making complications as discussed above. As a result, the acceptance of privacy policies generally falls short of informed consent (Pascalev, 2017) and should be considered in the context of the asymmetrical power relationship where the terms of data extraction are imposed on the user (Degli Esposti, 2014). Market place initiatives to overcome these difficulties are discussed below towards the end of this section.

The *right to be forgotten*, allows a person, in some circumstances, to have digital data, deleted so that third parties can no longer trace them when it is seen to be damaging or inaccurate (Weber, 2011). This right was established in the European Court of Justice in 2014 and applies to search engine links, where the linked information is considered to be inadequate, irrelevant, or no longer relevant. However, this ruling may be seen to be in conflict with freedom of expression, which has usually been favoured in the US courts (Neville, 2017). It has also raised concerns of censorship, and the loss of data which may be needed at some time (Ayer, 2001). The right also has no jurisdiction over websites that operate outside of the EU, nevertheless, Google had received 656,101 take down requests by 6th March 2018, which affect 2,442,884 URLs. (Google, 2018b). Under the General Data Protection Regulation, the right of erasure goes further than the European Court of Justice ruling and provides a partial right to be forgotten, enabling people to request erasure and further processing under certain circumstances (European Union, 2016).

Data amnesia, is slightly different from the right to be forgotten as it would require no intervention from individuals. The existence of data creates a privacy threat because it carries the risk that the individual may be harmed at an unknown later date (Solove, 2006). Mayer-Schönberger (2009) states the case for societal forgetting, by arguing that the act of remembering takes power from the surveilled to the surveyors in a temporal panoptican thus having a chilling effect on what is said and on societal engagement. Second, it allows society to forgive people and accept that people change. The proposal would see expiry dates attached to data after which the data would decay in parts or be deleted.

Anonymity, is a strategy to remove or encrypt information that can be used to identify an individual, thus preserving their privacy. However, there is a lack of anonymisation techniques that are generally usable whilst preserving data quality (Francis et al., 2017).

Indeed, whilst data may be anonymised in one context by adding additional datasets de-anonymisation is often possible (Hartzog and Rubinstein, 2017). These are called linkage attacks, the use of auxiliary information from a different dataset is used to match to the non-sensitive data (e.g. sex, postcode) within the target dataset. The suspicion is that as Ohm (2010, p. 1704) states '[d]ata can be either useful or perfectly anonymous but never both'. However, a technique named differential privacy seeks to sidestep this dichotomy. It does this by introducing noise (or statistical inaccuracies) into the dataset in such a way that accurate statistics may be obtained from the dataset whilst preserving privacy. The level of noise may be adjusted to obtain a balance between accuracy and privacy. The technique has been used by Apple since IOS 10.0 when analysing usage patterns (Apple, 2018). Finally, intersession techniques are being developed that allow for statistical analysis of datasets without access to the original data. These use encryption of the base data before it is transmitted to the statistical engine (where the encrypted data is not retained). Sharemind is an example of this technology (https://sharemind.cyber.ee/secure-computing-platform/).

Safe havens for information is an approach which involves taking data into archival custody, where it is kept under strict access controls (thus preserving privacy) until the information may be made public. An example of this is UK Census data which is archived, used for research purposes being subject to statistical analysis, but is not released to the public until the Lord Chancellor agrees under the Public Records Act 1958, known as the 100-year rule (legislation.gov.uk, 2018b, p. S5(1)).

The last of the basic approaches to be covered is *privacy by design*, here privacy requirements are taken into account throughout the systems engineering process. In this way privacy rules are embedded within the applications, and management systems for the data, with the aim of safeguarding user privacy whilst not limiting system capability (Romanou, 2018). In Article 52 of the GDPR data controllers are encouraged to consider data protection by design and default, although exceptions can be made on the basis of cost, technical capability, nature, scope, context, and purposes for processing (European Union, 2016). Whilst Hadar acknowledges the difficulties that this presents software developers, hence their preference for policy-based solutions (Hadar et al., 2018) there is a belief that the use of privacy design patterns may be a useful tool in what is a complicated task (Caiza et al., 2017).

Legislation protects data, to a degree. But different countries use varying definitions of personal data and have different data protection laws (Spiekermann et al., 2015). It has been argued that control of data collection should be abandoned and that control of use is

more appropriate although Nissenbaum (2017) disputes this and calls for both regulation of data collection and use. In the EU the GDPR comes into effect in May 2018, and whilst attempting to balance the need for economic growth within the community with the privacy concerns of its citizens, provides regulation of both data collection and use. In the UK this will be enacted as the Data Protection Act 2018 when passed by Parliament.
The GDPR introduces some changes to previous EU legislation (European Union, 2018). The following synopsis is taken from an analysis from the law firm White & Case (Gabel and Hickman, 2016).

- The GDPR now covers any organisation that is doing business in the EU, as opposed to previous legislation which only covered those with an EU establishment.
- Data may only be collected for a specified purpose that has been notified to the individual.
- There must be a justified lawful basis for processing data, the three bases are:
    - Contractual performance, in order to fulfil a contract, and in situations that take place prior to a contract (e.g. a product enquiry).
    - Legitimate interests, in this case the interests of the data controller must be balanced against the rights of the individual and be justified.
    - To comply with legal obligations.
- Consent must be informed and be a clear affirmative action of the data subject, and it may also be withdrawn.
- The GDPR's aim is to strengthen data subjects' rights and so is expected to result in stricter enforcement.
- The data subject can challenge the legitimate reason an organisation puts forward for processing data, and the burden of proof of lawful processing is now switched from the individual to the organisation.
- Data subjects can now require an organisation to delete data where the retention is not in keeping with the GDPR.
- New systems must now be designed with compliance in mind.
- Organisations now have 72 hours to report data breaches to data protection authorities.
- Data that no longer serves the purpose for which was originally collected must now be deleted.
- Finally, fines have increased to €20m or 4% of an undertakings world-wide turnover, whichever is the greater.

Much of this accords with the approaches to protecting personal data described above, for example purpose limitation, informed consent, right to be forgotten, and privacy by design. The changes to informed consent may prove interesting, as it is debatable that accepting large companies long and complex privacy terms constitutes informed consent (Pascalev, 2017). The changes will undoubtedly be challenging for some organisations and it will be interesting to monitor how they react.

Finally, there are independent initiatives aimed at improving the privacy aspects of personal data. The following will examine a sample of interventions, proposed or in place, that address privacy issues by providing personal data spaces or stores which lie in the control of the individual, thus creating a personal data economy (Elvy, 2017). Their intended use is as virtual intermediaries which can control the sharing of data with third parties (Gawer, 2017) and alongside AI, were considered by some to be the most important area of new technology (Overton, 2016). The intention is to transform the role of users from passive data sources to active subjects and participants in value creation (Lehtiniemi, 2017). However, it is unclear how far this initiative has progressed.

There are three schemes assessed by Lehtiniemi but none to date position people as 'active subjects in value creation'. Cozy Cloud, is a personal cloud storage facility but does not perform any intermediation, Meeco allows users to control their data in one encrypted space and in March 2018 is in beta testing, and OpenPDS is still in development but offers a SafeAnswers tool similar to Sharemind's initiative covered below. Other examples follow: CitizenMe allows people to monetise their data, or give to charity; DigiMe enables people to consolidate and explore their data in one encrypted cloud; PeopleIo permits people to 'licence' their data by acting as an intermediary between individuals and brands; and finally, DataCoup is similar to Meeco and DigiMe in its consolidation features, however, data is then profiled and a value attached, currently DataCoup is purchasing the data itself, but expects to operate as an intermediary to other organisations at a future date. Of all these operations only DataCoup and PeopleIO are allowing people to monetise their data, but this is still a nascent business area. There would appear to be two practical issues with the business model. First, people will have all their data in one place which may be a security threat. Second, payments received are low, currently a few dollars a month, not enough to attract the wealthy individuals that most marketing organisations would like to target. As a result, it is not an area that the Financial Times would recommend for investment (Greenhalgh, 2015). Nevertheless, more initiatives are forthcoming see Dong (2016), Will (2017), and Belyaev et al. (2018) for health records.

There are other approaches to the privacy data marketplace, one such, a privacy exchange authority (PEA), is proposed by Pascalev, (2017). Here the PEA would allow individuals to select their preferences for informed consent and then deliver and authenticate those with the big data companies. It is, however, unclear how differentiation between organisations would be achieved for people, as described above, place trust for a variety of reasons one of which is company characteristic. Another approach is the pay for privacy model, where consumers pay an additional fee for their data to remain confidential (Elvy, 2017).

Finally, another way of providing anonymised access to sensitive data is gaining traction, e.g. ShareMind. In this case the data, for instance welfare records and educational data, remain in their protected environments but the data is encrypted on the host computers before being shipped to a specialised analysis engine. The analysis engine cannot decrypt the data but uses secure computing technology to analyse the data and produce encrypted results, thus preserving the privacy of personal data.

This section has discussed some of the initiatives and strategies currently in place, or proposed, to protect personal data. The following section focuses on the self, introducing the concept of the digitally extended self before leading into the specific focus of this thesis, and introducing the rest of this document.

2.8 The self, its extension into the digital universe, identity and some terminology

There is considerable literature on the self, much of which is of only tangential interest to this area of work. For instance, from a philosophical standpoint Olson's *What Are We? A Study in Personal Ontology* (Olson, 2007) examines the metaphysical aspects of the self and consciousness. Mead (1934) considered the development of the self to originate out of the social process through reflexivity as a result of communication with others, which is improved through the generalised understanding of others. This social interactionism of an ‘autonomous' self was later rejected by postmodernists who renounced the modernist essentialist philosophy and in so doing, also rejected the concept of the self (Callero, 2003). For instance, Foucault (1998a) considered that, whilst the self in Athenian times was ideally a construct of reflexivity, through Christian practice, with the emphasis on verbalising one's thoughts and obeying one's master, the individual will is abandoned and therefore also the self. In *The History of Sexuality: The Will to Knowledge* Foucault (1998b) also suggested that individuals are subjected to discourses, which simultaneously provide power and knowledge. As a result, people are not only controlled by the power of

the discourse, but also become self-scrutinising and self-forming with respect to that discourse. Marcuse (1968) also writes of loss of self-identity in *One Dimensional Man* as a result of consumerism. The individual becomes a controlled consumer working harder to earn money to buy additional goods, which people are encouraged to buy, but which they do not need. Postmodernists argue that the symbolic interactionism of, for instance, Meads construction of the self, is a vestige of enlightenment values found in modernist thought. The postmodernist literature in its anti-essentialist philosophical stance rejects the concept of the self.

Feminist literature on the self, emphasises women's traditional lack of selfhood which, as McDonagh (1997) states, was codified in law, and symbolised by the woman's loss of surname after marriage. There are various views of how the self is socially constructed. Chodorow (1995) sees the self resulting from the internalisation of experience especially the nurture that is received. Kristeva (1982), on the other hand, sees the self as a result of the interplay of the feminine semiotic and the masculine symbolic. Butler's poststructuralist view objects to the ‘creation' of the self through societal normalising routines, and performativity, and would see resolution through contesting such categorisations as biological sex, polarised gender, and conditioning sexuality, which tend to construct identity (Butler, 1990). Butler's battleground of categorisation, and the intersectional theory of King (1988) and Crenshaw (1993), can be seen in the arena of the database where individuals are classified, the data analysed and conclusions of identity drawn. Benhabib (1999), however, rejects the poststructuralist view and discusses the individual with a core self and multiple other understandings of the self, developed through the retelling of the story of the self, which she calls narrativity. An alternate view is that of Fromm who argues that the self is ‘essentially constituted by the role the individual is supposed to play' (Fromm, 1994, p. 117). In the *Sane Society*, Fromm (2002) writes of the need for man to market themselves to become effectively an object for sales rather than the sum of thoughts, feelings, experiences or judgements. The self is just the sum of the parts that are played for others. When we consider the self as the sum of roles, then Goffman must be considered, especially *The Presentation of the Self in Everyday Life* (Goffman, 1971) with its analogy of human interactions to those of actors in a play. The individual’s dress, actions, dialogue and perhaps setting is adjusted to guide the impression that others will have of them. Behind this front will lie the ‘real' self, but of course people understand the ‘play' and try to see through the act to the real person behind.

Castells (1996) argues that globalisation, and the information age, has caused a divide between the net (subjects and organisations) and the self, the strategies by which people

strive to identify themselves in quickly changing environments. The self, he argues, stands apart from the databases and networks, which connect them. Kember (2002, p. 5), however, argues that the net 'is regarded as an ecosystem for emergent artificial life-forms and as an entity or intelligent life-form in itself'. She talks of a net within which humans are nodes and act as neurones within a large intelligent entity. This is a common theme, Haggerty and Ericson (2003, p. 613) suggest that a new type of body is being created which is 'a form of becoming which transcends human corporeality and reduces flesh to pure information', which they label a data double, a virtual decoporealised body. These are not accurate representations of the individual, but 'pragmatics' used to differentiate individuals from larger populations.

Balka and Star (2015), on the other hand, talk of shadow bodies, created from snapshots of an individual's data, accumulations of which, become an aggregate social form and a shadow of the self.

The terminology in this area is inconsistent and terms that may be used to name entities similar to shadow bodies, or data doubles are used for other dissimilar concepts. One such term is 'virtual self', which Metzinger (2010) uses in the setting of phantom limbs, dream states and out of body experiences whilst writing on the subject of phenomenal subjectivity. In another context Shields (2003) examines the term 'virtual' in detail from many perspectives including the historical. He identifies a credit profile as a virtual identity but does not extend the argument to the virtual entities discussed above, although he does consider the virtual as possessing characteristics of the real. Others write about the 'virtual self' in the more conventional way, as an avatar, as used in *Second Life* or *World of Warcraft*. In this interpretation the 'virtual self' acts as a visual representation of a person or alter ego in some virtual community or world (Lastowka and Hunter, 2004, Clemons et al., 2007, Halbert, 2009) but is not an accumulation of data.

Others have written about the person extending into the virtual, in the context of the ever-increasing availability and power of computing devices. David Chalmers in the forward to *Supersizing the Mind Embodiment, Action, and Cognitive Extension*, (Clark, 2008), talks of new technology as extending the mind. Agger (2008, p. 1) uses the term 'virtual self' to describe 'the person connected to the world and others through electronic means' but not as a separate virtual construct. Mayer-Schönberger (2009) describes electronic data storage as an extension of memory and argues for the right to forget. Floridi (1999) considers data as it relates to the person and later (Floridi, 2008) as part of an individual, not as a separate entity. In an extension of this concept, Baker (2008) writes of us no longer being numbers but models. None of these authors, however, identifies a grouping of information, held

across multiple data stores, as a ‘virtual self’ in their writing, in the way that Haggerty and Ericson or Balka and Star have. However, Solove uses the term digital person to describe ‘a personality translated into digital form, composed of records, data fragments and bits of information' (Solove, 2004, p. 226). The collections of data, which describe the real person, are referred to as digital dossiers, the use of which is ‘shaping our lives' (Solove, 2004, p. 3). There are therefore many labels for varying versions of digital selves but this thesis does not seek to position the digitally extended self as an exploration of the self, nor does it endeavor to explore the inter-relationship of identity and the self as may be found in Sources of the Self: The Making of Modern Identity (Taylor 1992). To do so would be a distraction from the focus of this thesis on personal data. The issue of identity poses the question, to what are the conditions under which a person at one time can be said to be the same person as at another time? The bundle theory of the self suggests that people are collections of different perceptions which rapidly succeed one another and so are in a state of constant flux and movement (Hume, 2003). Data collected at one moment in time may be analysed to create a persona which differs from a similar analysis completed some years earlier. Alternatively, a sub-set of an individual’s data may be analysed creating persona one whilst the contemporaneous analysis of another subset may produce persona two. In this respect it is possible that they may be reflecting the different roles played by an individual (Goffman, 1971) or different aspects of an individual’s identity relating for instance to class or race (du Gay, Evans, and Redman, 2000). The self may therefore be projecting varying identities. However, the Concise Oxford Dictionary defines identity as: ‘the sameness of a person or thing at all times or in all circumstances; the condition or fact that the person or thing is itself and not something else; individuality, personality’ (Murray 1971, p. 1368).

The need to identify an individual is discussed above in relation to trust and the beneficial use of data (section 2.6). The issue is one of tying data to a single person, correctly. Failure to do so may result in many things, amongst which are unexpected bank credits (Molloy, 2016), identity theft (Solove, 2004), restrictions on boarding a flight (Fife, 2018), or even torture (Abu-Laban and Nath, 2007). However, any data accredited to an individual, even incorrectly, is considered that individual’s personal data and is analysed as such and would contribute to their digitally extended self, the definition of which will be presented in Chapter 3 as the total of the data descriptive of the individual that exists within the digital universe.

2.9 Summary

This chapter has explored the boundaries of personal data and defined it for the purpose of this thesis as ‘data that is an attribute of an individual’. To label an item as personal data infers that it describes an aspect of a person. However, it may also indicate possession, it is data that belongs to a person, which in turn infers control over the data. Control over access to information is at the heart of digital privacy (Parent, 1983, Gavison, 1980, Allen, 1988), and some of the rationale for privacy in this context has also been discussed.
Much of the threat to privacy has come from the repurposing of data and the large amounts of data produced. Data has been called the new oil as it is now considered the most value resource (The Economist, 2017), but that is inaccurate. Oil is limited in supply, and is a single use commodity, whereas data is not, in addition data gains in value in combination with other data and through analysis. Data is driving economic growth, the movement of data across borders alone generates yearly economic gains equivalent to the GDP of France (Schlosser, 2018).
Despite the benefits derived from the use of data, there are dangers as well, some of which have been discussed above in relation to privacy. Work to mitigate the dangers has also been identified including the new European legislation, which seeks to strengthen people’s rights whilst enabling exploitation of data. It has been mentioned that terminology is used inconsistently when personal data and the extension of the self into the digital universe is discussed, and it is that issue which is picked up in the next chapter.

Chapter 3: A Proposed Classification and Model for Personal Data

3.1 Introduction

There are continuing concerns about the use of personal data, especially with respect to privacy, informed consent, and the right of access to data, which drive a need for well-defined and consistent terms to describe that data. This thesis focuses on data that are descriptive of an individual, and this first phase focuses on the use of terminology, addressing the first research question:

RQ1: What are the components of the digitally extended self and how do they relate to one another?

This is important because of the increasing use of personal data, and the resulting markets in personal data (Spiekermann et al., 2015), which have led to concerns regarding issues of privacy (Acquisti et al., 2015), privacy-related decision making (Kehr et al., 2015), informed consent for organisations to collect, process, curate, compare, and transfer their data to other bodies (Heeney, 2012), and also an individual's right of access to data descriptive of them (L'Hoiry and Norris, 2015).

Given these concerns it is surprising that there is no common terminology around personal data, and that no previous initiative to propose one can be identified. What nomenclature should be used for digital data that is descriptive of an individual? What collective nouns can be used to classify the data and how are they related to each other? A variety of terms present themselves in the literature, for example *digital footstep, fingerprint, shadow*, *profile*, *mosaic*, *persona*, *virtual self*, or *doppelgänger*. The terms are widely used but not in a consistent way. Neither are the usages critiqued. The problem of a common set of terms in the face of technological change has been noted before; for example, Bakshi (2016) highlights inconsistencies in use when discussing the digital economy in general, and Heinderyckx (2014) points to rapid rate of change of ICT terminology. Others have tried to address the problem in other domains, for example Safran (2007) defined terms when discussing health data, but none has aimed to specifically discuss the terminology associated with personal information.[2]

[2] Chapter 3 of this Thesis was, in large, originally published in the Journal of Information Science (Parkinson et al., 2017)

The use of consistent terms and concepts is important because it reduces ambiguity in academic debate and improves information sharing – particularly between service providers and their users. When giving informed consent, an individual must determine, and understand, the information that is covered by the agreement (Solove, 2013). It is also crucial for legislators to evaluate and use terminologies consistently, whether they be incorporated in organisational privacy agreements or legislative language or guidelines. Additionally, a concrete set of concepts is important for the design of systems that deal with personal data, as it may have implications for how data are stored, managed and exposed through a range of interfaces.

In order to tackle this problem, this research analyses the terminology and concepts of personal data present in the literature, with the goal of identifying common concepts (even when they are named differently) and establishing their relationships. The most popular/descriptive terms are then selected, and these concepts brought together in a model of the digitally extended self. The model is then tested against the literature and (in a second phase, described in Chapter 4) against this researcher's own personal data. It will demonstrate that this illustrates two uses, the first as a standard set of terms and the second as a high-level data model.

The remainder of this chapter is arranged as follows: Section 3.2 describes the method used for the lexicological analysis of the literature; Section 3.3 discusses the terms encountered and their relationships; and, Section 3.4 shows how these can be brought together in a coherent model. Section 3.5 presents a validation of the model against 45 key publications from the original sample, and shows how the model can be applied to a particular scenario. Section 3.6 discusses other approaches to categorisation and situates the model described in this chapter, whilst Section 3.7 concludes the chapter and leads into the second phase of this research.

3.2 Method

An initial reading of the privacy and surveillance literature enabled the extraction of a list of terms used to describe facets of data descriptive of an individual. In order to perform a lexicological analysis of the meanings allocated to these terms it was necessary to obtain examples of their usage. Several data sources for the search were considered, e.g. Web of Science, Scopus, or university-specific search engines such as Oxford's SOLO. Google Scholar was selected due to the wide range of papers and books within its base of data, the ease of integration into the chosen reference manager (Bookends), and its increasing use within the research community (Craswell and Poore, 2012, Bryman, 2008). Its weakness

with respect to Boolean searches, and the restriction to 1,000 search results (Haddaway et al., 2015) was not significant for this research.

A set of four common starting terms was identified from the initial readings (*digital footprint*, *digital mosaic*, *digital persona*, and *virtual self*), these then became seed search terms for Google Scholar. The search engine, at the time of this work, returned a maximum of 1,000 items for each search and so high usage terms were searched for by calendar year thus maximizing the number of references returned. For each term, the results were then ordered by citation (discovering a power law distribution, meaning that each term had a relatively small number of higher cited sources). A purposeful sample was then selected from these based on high citations relative to publication date, and overall size of the sample for that term. Each publication from the purposeful sample was read. The terms and their contexts were highlighted within the documents and then manually extracted from each paper and analysed. Where new terms were discovered within these documents they were added to the list to be researched and the process undertaken again, resulting in a snowball sample of 64,584 papers covering 16 search terms and resulting in a purposive sample of 247 (the terms are shown below in Table 3.1 together with the total count of results, and the number of papers selected under each term for the purposive sample). Digital Fingerprint and Second Self have a relatively low purposive sample due to the high number of spurious results. For instance, Digital Fingerprint is a common term within forensic science, and Second Self is part of common phrases such as ‘the second self-control task’, and ‘Barber’s second self-creation theory’.

In order to determine usage, each term was taken in turn, and the sample documents, containing that term, examined. Meanings were observed and common themes extracted. The terminology descriptive of personal data was then examined and through a series of iterations a standard categorisation developed, and the relationships between those categories defined in order to create a model. The naming of these categories was based upon common usage and strength of metaphor. A further iteration to validate the findings was then undertaken and is described in section 3.6.

Search Term	Google Scholar	Purposive Sample	Search Term	Google Scholar	Purposive Sample
Digital Biography	165	4	Digital Persona	1679	18
Digital Doppelganger	77	11	Digital Self	4223	23
Digital Dossier	421	4	Digital Shadow	592	4
Digital Fingerprint	4930	2	Ersatz Double	11	1
Digital Footprint	1501	29	Online Identity	9256	18
Digital Identity	9059	26	Second Self	25578	12
Digital Mosaic	834	11	Shadow Identity	171	2
Digital Person	967	39	Virtual Self	5120	43

Table 3.1 Summary of Google Scholar search results, Aug 2014.

A potential weakness of this approach is the dependence upon the work of the author to examine the literature and extract meaning. It can be argued that the use of a second researcher to independently analyse the literature and identify themes would strengthen the findings. However, as Armstrong et al. (1997) note, this type of analysis is a form of interpretation in which researchers' views have important effects. It is possible that a second researcher may have come to a different, but no more valid, conclusion. The derivation of categories and their labels was, however, subject to iterative debate between the author and his supervisors with the objective of producing a consistent set of terms that can be used when discussing personal data. Whilst others may have decided on an alternate nomenclature this research endeavoured to create a categorization and set of names that are informative, easy to understand, and remember (Glushko et al., 2013).

3.3 Results

The analysis of the terminology and their usage identified three main issues. First, terms used to describe categories of data descriptive of an individual are also used to label other

things; second, single noun phrases were used to label similar but differing groupings; and third, more than one noun phrase was used to label a single grouping.

3.3.1 Terms used to describe categories of data also used to label other things

It must be expected that variations in usage will be identified when examining the use of sixteen noun phrases, and within a discourse, the meanings tend to cover overlapping sets of things, as presented below. However, when analysing usage across discourses, as in this case, then examples of entirely different meaning were observed. Digital footprint, for example, is used to label the outline of a building on a digital map (Jones, 1990). Digital mosaic may be a collection of images used to create a larger image as in the case when illustrating the location of Dengue fever in Nicaragua (Chang et al., 2009), or a collection of videos, which together form a composite video (Ludwig et al., 1997). In another discourse, virtual self was used by Goffman (1990) to describe a role acted by an individual in their everyday life, whilst Metzinger (2010) uses the term to cover phantom limbs, dream states and out of body experiences, and Vander Valk (2008) considers that there are no humans in the world but that we all exist only in an immersive virtual environment as virtual selves. When terms are used across disciplines to label separate things, as has been illustrated above, meaning is created through use and explanation. However, when a single term is used to categorise similar but differing things it becomes imprecise, and hence problematic. This is addressed by ensuring that, in the terminology proposed in section 3.4, each term has a single and unambiguous meaning by using assigned terms within a controlled vocabulary (Svenonius, 2009).

3.3.2 Terms with multiple meanings

When looking at meaning in the discourses surrounding data that is descriptive of an individual, variations in meaning were observed. Rather than exhaustively listing these, the following illustrative examples are presented.

The term *digital footprint* is used to categorise data left behind by an individual in the virtual world (Byron, 2008, Greysen et al., 2010). The emphasis is on an individual leaving their own data trails. However, Palfrey and Gasser (2008, p. 33), amongst others, state that '[d]igital footprints are digital artefacts which can be left by the individual or by another'. Sellen et al (2009), however, assert that digital footprints are created about which the individual has little or no knowledge or control. This raises the question of whether the

subject individual, another individual, or both, create digital footprints, and whether the subject individual knows of them or not? This is resolved in the proposed categorisation by having one term for data left by an individual which describes themselves, and a second term for data left by someone that is descriptive of another person.

A second noun phrase in common use is *virtual self*, which Lyon (2000) uses to identify collections of data, and analyses, that describe an individual, and which Turkle (1994, p. 166) sees as 'extensions of ourselves we have embodied in program'. There may be a single virtual self to represent all data and analyses descriptive of an individual, or else, multiple virtual selves representing subdivisions of the data and thus perhaps replicating Goffman's contextual self-projections within the 'real world' (McInnerney and Roberts, 2004). However, is there a difference between the actors creating the virtual self, or of an organisation imposing a persona upon an individual? Lyon (2000) would appear to consider the virtual self as imposed perhaps as a result of some form of surveillance or analytics. Turkle (1999) considers the individual as the creator or persona(s). Indeed, this is the case with Pearson (2009), who describes the virtual self as a constructed online identity, whilst Bessière, Seay, and Kiesler (2007) use the term to label the self-created within an online game, and give it the synonym of *avatar*. The virtual self can however also be distinguished in another way, either as representing a para-authentic extension of the individual, or else a construction of an alternate personality (Lee, 2006), perhaps as an experimental device. Finally, a less complex projection of the self is a photograph used to represent an individual, e.g. on a social network site, but labelled a virtual self (Siibak, 2009).

3.3.3 Multiple terms same meaning

The previous sections present two examples of terms used to describe similar but differing categories of data. There is, however, a situation where multiple terms are used to describe the same thing. In the case of categorisation of data descriptive of individuals, the use of different terms to identify the same class of elements can cause uncertainty and a resultant lack of rigour. This is demonstrated below in section 3.6.1. For instance, *digital footprint* (Batchelor et al., 2012), *digital fingerprints* (Wittes, 2011) and *digital persona* (Clark, 2010), as used in the cited papers, are all synonyms and used within the context of personal data. In this instance, the use of the more commonly found term *digital footprint* would provide consistency and allow the nuanced inference of an individual's digital artefacts being used to create an online persona to be explained more fully.

3.3.4 Summary

A lack of consistency in the application of noun phrases used to label categories of data descriptive of an individual has been illustrated. One way forward would be to leave the situation unchanged and allow usage to either continue in an unclear way and hope that time will allow meanings to coalesce around the most popular noun phrases whilst others wither and die away. This research has taken an alternate approach and has developed a classification model for the data descriptive of an individual, which is presented below. The noun phrases chosen to label categories of data were selected as a result of their commonality of use and strength of metaphor. For instance, *digital footprint* and *digital fingerprint* both describe a data artefact left by some activity of an individual, which reflects itself in the virtual world. Footprints in the sand tend to disappear and can be readily observed, although they cannot normally identify an individual. On the other hand, fingerprints tend to remain for many years, can identify an individual, and are difficult to observe. In this case, although *digital fingerprint* is the stronger metaphor, *digital footprint* was selected because it is used more widely.

The terms selected were then developed into a coherent, and consistent categorisation of an individual's data as it is represented in the virtual world. These are described in the following section and illustrate the gradation of personal data as it is deposited, merged, transformed and analysed.

3.4 Result: The Model

In the total sample 16 terms were identified, but in this analysis, they are grouped into one of five categories, each of which, it is argued, is distinctive in terms of its origin and construction, and which together form the layers of the model. Each category is named after the term that was considered to be the most representative. The five concepts in the model are:

- *digital footprint*: data descriptive of an individual, laid down by that individual as a result of using, or knowingly being observed by, computing devices;
- *third party digital footprint*: digital footprints created by an unknown computer system, or an individual which are descriptive of another individual (the data subject);
- *digital mosaic*: a collection of digital footprints which can be used to create a picture of a person, a simple digital mosaic consists of a person's own digital

footprints whereas a full digital mosaic is used to describe the collection of both an individual's own, and third party digital footprints;

- *digital persona*: a model of an individual created by the analysis of data from the full digital mosaic, and/or other digital personas, and optionally additional second level data;
- *digitally extended self*: the combination of the above elements to provide the fullest possible digital representation of an individual.

In these definitions, the term *second level data* is used to identify data that are not directly descriptive of an individual, but which provides information about an individual's attributes (e.g. demographic data which is associated with a person's post/ZIP code, or data descriptive of a group to which the individual belongs).

There are several reasons for placing these categories together in a coherent model. Firstly, it provides a vehicle for discussing the issues associated with the collection and use of an individual's data, and in doing so, defines a set of terms thus reducing ambiguity. Secondly, it illustrates where boundaries exist. It is often at the edges where more interesting and difficult decisions have to be made, especially with respect to knowledge and control of an individual's data. Finally, by naming structures in certain ways we affect how they are viewed. In this case, the term 'digitally extended self' has been created to describe the virtual self, not as a separate entity but as an extension of the real self.

Figure 3.1 shows an overview of the model. The basis of the model are the digital footprints created by the data subject, and the third party digital footprints created by other individuals. Combined they form the digital mosaic. This, in turn, is the basis for digital personas that typically exist to profile an individual for some purpose. These personas may also use second level data, and other digital personas as input to the analysis. The whole, is then defined as the *digitally extended self*.

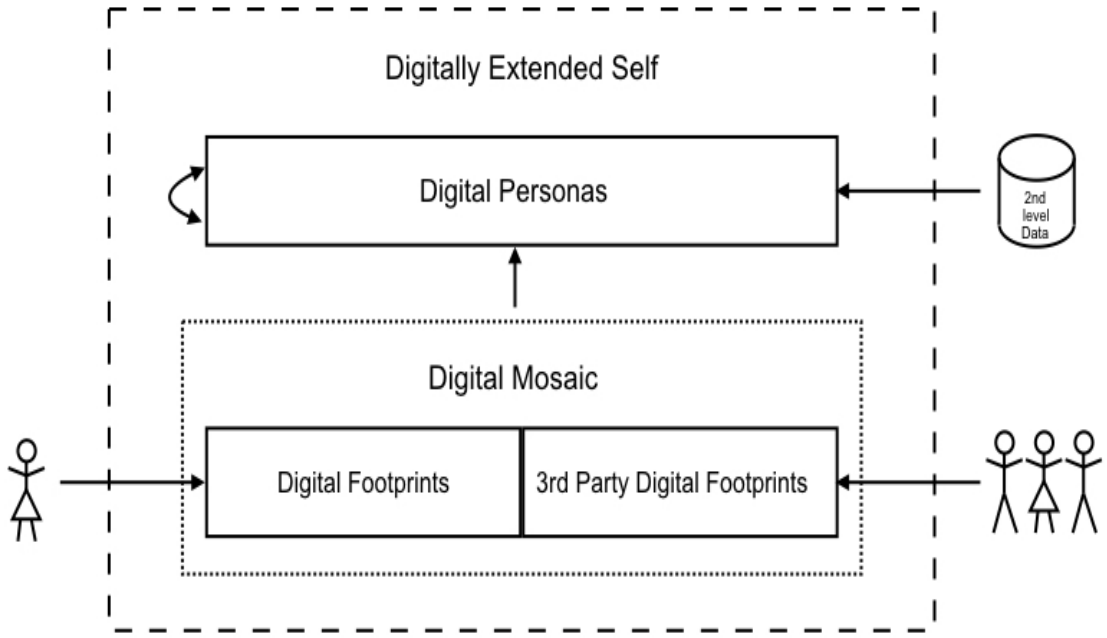


Figure 3.1 The hierarchic model of the digitally extended self – showing the five categories of personal data.

3.5 Validation of the Model Against Terminology

The model serves two purposes. It provides a clear nomenclature which facilitates a cross-disciplinary use of terms, the second is as an overarching data model. The model is therefore validated in two ways. First, to ensure that the model encompasses the existing, highly variable and disorganized, terminology; and second as a second phase of this research in Chapter 4, against actual data.

As a first validation step, the Table 3.2 (and the tables in Appendix BB) show how a range of terms and usages from the purposive sample map to the categories in the model. To create this mapping 45 examples have been selected that provided coverage of the model concepts and where the same terms are used in different senses (for example, Byron (2008) discusses digital footprints in the same way as our model, but Chretien et al. (2009) use the term to describe something that maps to a digital mosaic in the model instead). Within the publications no match for third party digital footprint was found as the phenomena were mentioned but not named, it is therefore omitted from the tables.

With this exception, it was possible to exhaustively map terms found in the literature sample to the categories proposed as a result of the analysis, showing that all the terms

used in the 45 publications, that refer to an individual's data, map to specific parts of the proposed model.

Term in Model	Term from Literature	Usage	Example of Usage
1 Digital Footprint	1.1 digital fingerprints	1.1.1 'data about individuals held in the hands of third parties' p2	Wittes (2011, p. 2)
	1.2 digital footprint	1.2.1 a digital artefact left behind by some activity 'as they 'tread' through the World Wide Web, they leave behind a 'footprint' p1227	Batchelor et al. (2012) Siemens & Long (2011) Greysen et al. (2010)
		1.2.2 'personal information available online' p58	Byron (2008)
		1.2.3 postings on social media (by medical students)	Chretien et al. (2009)
		1.2.4 patterns of internet usage / artefacts	Hankin et al. (2013)
		1.2.5 traces of online presence	Hengstler (2011) Madden et al. (2007) O'Keeffe & Clarke-Pearson (2011)
		1.2.6 pervasive environments and contextual traces	Kapadia et al. (2007)
		1.2.7 results of activity in the virtual world which describes someone	Palfrey & Gasser (2008)
		1.2.8 a group of digital footprints on one site i.e. Facebook	Moore & McElroy (2012)
	1.3 digital persona	1.3.1 an electronic portfolio of work created by a student	Clark (2010)
	1.4 identity	1.4.1 'a trail of data artifacts' p10	Briggs (2013)

Table 3.2 Validation 1: Digital Footprint - mapping of literature to the model.

The validation is shown in table 3.2 for digital footprints, and additional validations in Appendix BB for simple digital mosaic, full digital mosaic, digital persona and digitally extended self, show that all the categories within the model map to phenomena that have been named and discussed in the literature. They also give a sense of the ways in which different labels have been used to express and describe the terms in the model. The second

scenario-based validation in Chapter 4 comes to the model from the other direction, and looks at how the data involved in real case studies map to the model. It will demonstrate that the model covers all of the data in the case studies, and also shows how the distinctions made by the model are useful for discussing data in that particular scenario.

3.6 Relationship to other classification approaches

The personal, or social, point of view is generally used when framing the debate regarding issues of privacy and data descriptive of an individual (Nissenbaum, 2010). However, other perspectives may be adopted. For instance, Pollach (2007) suggests a function-based approach, forming a matrix of data types (e.g. sales data) and data handling methods (e.g. selling) in order to help people better understand the consent that they are giving. However, as a method of classification for all data descriptive of an individual this approach is limiting, due to the constraints of constructing an exhaustive set of types and processes. A similar approach, used in the UK Data Protection Act 1998, considers types of organisation that hold data (e.g., research organisations), but also the use to which the data are put (e.g., domestic purposes). Again, this approach does not provide an exhaustive classification of data descriptive of an individual, and it can be argued that the data covered are in parts unclear (Millard and Hon, 2012). This may be a cause of inconsistences in the categories of data provided that are found in responses by companies to subject access requests under the Act.

Polonetsky et al. (2016) take a more ontological approach and propose a categorisation of personal data based upon degrees of identifiability of an individual. This is a useful contribution to the vexed problem of big data usage and personal privacy, and the approach does provide a complete classification. However, what may appear to be de-identified data today may be identifiable tomorrow due to technical advances such as the use of additional data sets that compromise the level of anonymity of the data. Consequently, the classification of data based on degrees of identifiability may fluctuate and become indeterminate.

An alternate approach to data descriptive of an individual would be through the Data Information Knowledge Wisdom (DIKW) hierarchy, a structural approach to data and its transformational uses. The assumption is that data at the bottom of the hierarchy is transformed through processing into information, which is processed to create knowledge, and knowledge, in turn, yields wisdom (Rowley, 2007). This structural and transformational framing can be used to argue that data by itself offer no threat to privacy unless it can be transformed into information, knowledge or wisdom, each having the

potential to be more threatening to an individual's privacy. Whilst Batra (2014) argues that the advent of data analytics in real time blurs the DIKW distinctions, the classification is still of some interest as not all data are subject to analytics, and those that are can still be classified.

Finally, Palfrey & Gasser (2008) use availability as a classification tool, observing a distinction between data that are publicly available and those that are not. This is used to differentiate between the digital identity (the publicly available) and digital dossier (all data descriptive of an individual). There are two issues with this classification: it may be considered too simplistic a distinction if it were the only observation made; and, more significantly, it does not have clear boundaries. For example, that which is considered available by a computer-literate person would be different to that accessible to others with more limited skills.

In this research, a new approach has been taken based upon the origin, handling, and manipulation of data by various agents associated with an individual. It will be demonstrated that this has the benefit of communicating ways in which personal data are transformed and transported, whilst providing a full categorisation of the domain, and at the same time being readily understandable.

The model not only distinguishes between different types of data, but helps draw attention to the fact that an individual's digitally extended self is not tightly controlled or atomic, but rather exists in graduated layers, with multiple owners, that progressively becomes less direct and more speculative as the data becomes more distant from the individual. In this context of multiple actors and varying gradations of data, which may be considered personal, it is clear why questions of privacy, ownership, and rights of access, are so complex.

To illustrate this: in the UK, there has been an ongoing debate regarding the definition of personal data. The Data Protection Act 1998 defines personal data in Section1(1) as data which relate to a living individual who can be identified from those data, or from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller.

This provides a wide definition of personal data and it can be argued that all elements of the model are covered by this definition. This includes second level data that is ascribed to an individual, by an organization, as a result of analytics, for example that based on the use of a specific item such as a model of iPhone. In the case of Durant v. Financial Services Authority [2003] EWCA Civ 1746, Auld LJ, the judgement limited personal data to that which affects a data subject's privacy, such as the subject's name, address, telephone co-

ordinates, working interests or hobbies. In this interpretation, only the core of the model, the *digital mosaic*, is considered private data. However, following this judgement, the Information Commissioner's Office issued guidelines on the determination of personal data (Information Commissioner's Office, 2012) in order to reconcile the Durant judgment with wider opinion. This lists eight questions, a positive reply to any of which may indicate that the information constitutes personal information. In this case the defined boundaries of personal data expand out from the centre of the model to include the *digital persona*—data that can be used to inform or influence actions or decisions affecting an identifiable individual. The guidelines also include data that is linked to an individual. It therefore can be argued that second level data at the edge of the model is also, under this definition, to be considered personal data. Whilst this topic is more nuanced than is shown here and deserves fuller analysis in further research, it has been demonstrated that the model can be used to illustrate the movement in the debate of what personal data is, and if accepted as a basis for legislation, could be used to define the boundaries of personal data.

3.7 Conclusion

The use of personal data continues to be a question of great interest in a wide range of fields, especially with respect to privacy, informed consent, and the right of access to data, driving a need for well-defined and consistent terms to describe that data. However, at present the terminology around personal data is confusing, comprised of multiple overlapping terms, with little agreement on the underlying concepts and their relationships. The research presented in this chapter investigated whether a model of personal data could be developed that would differentiate between different types of that data in a helpful way. In order to achieve this a lexicological analysis of the terms used to describe personal data has been presented, based on an analysis of 247 papers (taken from an original sample of 64,584). It identified five distinct concepts (labelled footprints, third party footprints, mosaic, personas, and extended self). These come together in a model of the *digitally extended self*. The model has been validated by showing how 45 examples of usage from the literature map to the model (showing that each of the categories appears in the literature, even though the terminology for them is inconsistent).

The first phase of this research has demonstrated, in response to the first research question, that a model can be developed that differentiates between different types of personal data in a way that is useful for describing data held by organisations. The second phase of this research, presented in the next chapter, explores how the model can be further verified by obtaining data from a broad set of organisations, and comparing it to the model whilst

using the classification defined by the model to analyse the quality of the data returned. The third and final phase of this research then seeks to examine the issues that organisations face in returning data at different layers of the mode

Chapter 4: Testing the Model with Real-World Data

4.1 Introduction

The previous chapter highlights the inconsistent vocabulary used when referring to personal data, and some of the issues that this can cause. A solution is recommended in the form of a standard set of terms that are defined and represented diagrammatically in a model of personal data, called the digitally extended self. This nomenclature was validated against 45 examples and found to be robust.

This chapter will describe how the terminology and model developed in Chapter 4 were further tested by applying it in a real-world context: the modelling of the researcher's personal data, requested from a range of organisations. This was achieved by submitting subject access requests referring to the UK Data Protection Act 1998, using a predefined process. The data was then analysed using the categories defined in Chapter 3 further validating the model. The analysis highlighted issues with the data supplied especially when compared to reasonable expectations. In Chapter 5 these findings are explored further through a series of interviews with nine experts with experience in issues of personal data.

In this second phase of the research, two questions are addressed in addition to the validation (which supports RQ1):

RQ2: How feasible is it for an individual to obtain the information, held by organisations, which is descriptive of them?

RQ3: What is the quality of the personal data returned by organisations when it is requested by individuals?

The process of requesting data is defined and measured in terms of effort and cost, in order to gauge the feasibility of an individual retrieving their digitally extended self. How well organisations perform, when providing information is evaluated through analysing the data provided by category and assessing the completeness of the response with the use of a standard scoring table (Appendix J). Further analysis illustrates whether some categories of data are more readily provided than others. Section 4.2, explains the methodology used, including how organisations were selected for inclusion, and the processes involved in data collection and analysis. The results of the data collection are then described in section 4.3 before being analysed in 4.4, and finally discussed in section 4.5. Section 4.6 concludes this chapter.

4.2 Method

This section examines the methodological options available for the collection and analysis of data descriptive of an individual through the use of subject access requests, and their suitability. It then describes how the research was done, examines some of the issues in the selected approach, and what action was taken to mitigate them. The work undertaken is a form of *auto digital ethnography*, auto because the subject of enquiry is the author, digital because all information obtained refers to the digital or virtual domain, and ethnography because it is concerned with the study of a group of people albeit in this instance extensions of one person represented in the virtual world.

4.2.1 Methodological Approaches

The research entailed listing organisations with which the author has or had a relationship and selecting a purposive sample to represent public and private companies across a range of sectors, central and local government and not-for-profit organisations. These organisations were then contacted, using a predefined procedure, requesting information that they held that was descriptive of the author. The process was measured in terms of cost and time, and the responses categorised against the classification and model defined in Chapter 3. The quality of the responses was also assessed using a scoring system based on a subjective evaluation of the completeness of the data provided by the organisations. Where organisations stated that data had been sent to, or received from, other institutions, these secondary organisations were used as a snowball sample. Ethical approval, (reference ERGO/FPSE/23880**),** was obtained with detailed responses kept on encrypted drives and password protected whilst the paper copies were locked away securely.
This process does not readily fit into established methodological approaches. It is not an experiment with a control group, but an investigation in the nature of a social enquiry. The approach is retroductive (Blaikie, 2009) based upon a critical realist understanding of the world (Bhaskar, 1998) where reality is stratified into the *empirical*, *actual* and the *real*. It is the *empirical* which is experienced, whilst events occur in the *actual*, these events are generated by structures and mechanisms in the *real* which can be postulated, it being the role of research to either prove or disprove these mechanisms. The model in this case has been formulated as a result of the classification of that data which are attributes of the individual, as described in Chapter 3. This next phase of my research tests that model against real data provided by organisations, with the intention of making observations about organisational behaviour. The final phase of this research, described in Chapter 5,

explores the background and reasons for the nature of the data provided, and challenges the observations through a series of interviews with selected experts.

This second phase used a mixture of conventional methodologies. Case study, where the case studied was that of an individual's data held by varying organisations. Ethnography as an individual's digitally extended self was the focus of the research, and quantitative and qualitative content analysis as content was analysed and measured before values were assigned. Auto digital ethnography would seem an appropriate description of the process undertaken.

A full description of the methods considered can be found in Appendix A.

As a mixture of methodologies was adopted, the advantages and disadvantages of each is now discussed in order that the strong points of each methodology may be drawn upon whilst recognising and mitigating their individual weaknesses.

Research into the author's digitally extended self, by using subject access requests, can be considered to be an 'exemplifying' case study (Bryman, 2008), the objective being to 'capture the circumstances and conditions of an everyday or commonplace situation' (Yin, 2003, p. 41). It is unlikely to be a representative case but it provided a suitable context as it enabled large volumes of personal data to be collected, without difficult ethical considerations, from a range of organisations. From Yin (2014) this has the following advantages:

- *Stability*; paper responses were scanned, then both scanned and electronic responses were made searchable through OCR processing. The paper resources were archived and the electronic versions secured on an encrypted laptop, backed up to local discs using Apple Time Machine software, and a further backup held on the cloud using Tresorit end to end encrypted services. In this way, the scanned and processed documents can be viewed consistently and repeatedly.
- It is *unobtrusive* in that processes and data are not created as a result of the case study - this was only partly the case. The data existed within target organisations IT systems, and as such was not collected for the purpose of this research, however, it was extracted only for the purpose of fulfilling the data request. The accuracy and completeness was therefore subject to the willingness and abilities of both the organisation and the individual designated to extract and supply the data. As a result, the responses reflect not only the data held within an organisation but also the policies and practices of the organisation, and the capability and disposition of the respondent. These factors may vary over time as a result of internal policies and external influences legal or cultural. Despite this

the data collected fairly reflected that which would be provided to an individual who based their request upon the categorisation model.

- *Specificity*; the data collected was focused, only containing the data required for the research.
- The data has *breadth* in that it contains data from across an organisation that may cover a long time-span, and be across many events and settings. At a higher level the selection of organisations for the study provides insights across a range of market sectors, private, public, governmental and NGOs as well as organisation size judged by turnover, employees and reputation. The disadvantages which need mitigation are:
    - *Retrievability*; data can be hard to find. In this instance this was not an issue for the researcher, although it may be for the intermediary who was extracting the data. As the main focus of this research was to observe and measure provision of data to individuals by organisations this is not an issue, and once obtained and processed the data was easily retrieved.
    - *Biased selectivity*; there is no definitive way to know if complete data was provided or on what basis data was selected; although in some cases omissions were deduced by reference to external data e.g. Tesco, the data owner, omitted to provide any Clubcard data, as did Dunnhumby, the data processor, the omission was evident as the data subject had a Clubcard account and received regular communications from Tesco with respect to it. Any detected bias in selectivity from an organisation thus provided additional data for analysis, whilst undetected bias will of necessity go unnoticed but was reflected in the data provided and analysed. The risk of perceived absence of data in responses from organisations was, however, mitigated during the data collection process though the production of a second Subject Access Request (to the same organisation) which specifically addressed the issue of perceived omission of data within the first response, as described in section 4.2.4.
    - *Access* may be deliberately withheld. Again, refusal to answer subject access requests (as turned out to be the case with Cooperative Energy and Facebook, section 4.3.2) was in itself valuable information for the purpose of this research.
    - *Reporting bias*; emails and letters provided by the respondents were subject to the biases of the authors, however, as above it may be

considered typical of those organisations responses and so was valid data for the purposes of this research. Of greater concern is bias created in the analysis of the data by the researcher, which is discussed below.

With respect to this research being a form of auto-ethnography there are additional considerations. First, there is the difficulty of disclosing aspects of the self in this form of transgressive account, it discloses aspects of the self which may be uncomfortable for the writer and the reader (Denshire, 2014). It must be accepted that to some extent the purposive selection of organisations will have been affected by the conscious or subconscious desires of the writer to present themselves in a way compatible with the persona of a doctoral student. This could have been overcome by the use of random selection but then the sample may not have been as broad. An alternative approach would have allowed another to select from a wider range of pre-categorised data. However, even in this case the data presented may have been skewed to fit the desired persona. This effect has been mitigated through the use of snowball sampling. A total of 32 organisations were chosen for the purposive sample. Organisations said to be providing data to, and obtaining data from, this original sample produced a secondary sample of 59 organisations of which 51 were contacted before the predefined cut off time for the data collection was reached. Section 4.3 provides further details on the sample and its construction.

A second issue associated with ethnography is that of reactivity, the researcher stepping into the environment being researched and as a result changing it (McKechnie, 2008). This risk was realised as Subject Access Request letters sent by Royal Mail, caused events to be recorded on organisations computer systems or else were scanned and then processed through an organisation's work flow systems. Some organisations reported these events as part of the formal reply, others did not. In either situation, the data or lack of it proved illustrative of data held by those organisations.

The research followed a predefined systematic process to obtaining, and analysing data (as described in the next section). The pros and cons of content analysis as a research methodology are discussed by Bryman (2008) and are instructive for the methods deployed in this research. First the advantages

- *Transparency*; the research had predefined coding systems and sampling procedures. It was therefore clear what data has been collected, how it was obtained and how it was analysed, enabling others to judge the quality of the research.
- *It facilitates longitudinal analysis*; this research will be archived in such a way that, if appropriate, a follow up study can use the data and analysis.

- Analysis of documents is *unobtrusive* and therefore often a non-reactive method. For this research, however, documents have been produced in response to formal requests and therefore some reactive effect has to be assumed.
- It is a highly *flexible* method, for instance in its application to news media and elite social groups to which access is difficult to obtain, and in this case to the analysis of responses to subject access requests.

An examination of the disadvantages illustrates pitfalls that were observed or mitigated:

- It is dependent upon the quality of the documents that are analysed. Scott (1990) recommends assessing the documents for:
    - *Authenticity*, that the document is what it purports to be. In this case most responses can be authenticated to be from the target organisations. UK addresses were obtained from either the Data Protection Public Register, or Companies House, and foreign contact details from the organisations websites. The data received was also verified to be from the target organisation and to describe the subject of the research.
    - *Credibility*, whether contents of the documents may be distorted in some way. This is an issue as some organisations may not wish to disclose information (e.g. for reasons of competitive advantage or damage to reputation). Others may lack the internal procedures to bring together all of an individual's information, and finally the production of the responses is subject to problems resulting from errors, carelessness or incompetence. As a result, the research may not be replicable. Credibility of the findings should not be impacted upon however, as the focus of the research is to examine that very accuracy and completeness which may be in question.
    - *Representativeness*, whether the documents are representative of all documents from an organisation. This research examines a single individual's data as it contributes to their digitally extended self. Care will therefore be taken not to make statements that generalise for all instances of individual's data held within an organisation or for the way that all subject access requests are answered.
    - Coding schemes rely on a level of interpretation by the coders. Those producing documents and also those analysing them will be, at least to some extent, affected by the culture within which they operate. Therefore, it is likely that there will be a level of mismatch in interpretation of the

documents. Also, problems will arise when imputing latent as opposed to manifest information. In this research, the data can be categorised into two parts. First, data that is present (or not) and so is not open to interpretation e.g. a record of a digital footprint. Second, covering letters, and judgements regarding the completeness of the response, both are subject to interpretation. The letters may be misinterpreted or not taken at face value, or else personal knowledge, extrapolation from other external events, or personal positioning may affect the analysis (e.g. Amazon deny use of demographics or personal profiling which this researcher finds unlikely). In order to mitigate this effect a second opinion will be sought where possible and the findings will highlight where judgements have been made.

- Content analysis can be seen as atheoretical (lacking a theoretical basis). However, in the case of this research its basis for analysis uses predefined categories from a hypothesised model.
- Finally, it is difficult to answer 'why' questions from content analysis (unless this is explicitly stated), however, this stage of the research is focused on the what, how, when and who rather than why (which is dealt with in the final phase of this research).

In summary the methods deployed should give:

- stability of data, through scanning and archiving;
- focus, on only that data which is required for the research;
- a breadth of data, across organisations;
- access, to personal data;
- transparency, as a result of predefined documented sampling, procedures, coding systems and model.

There are a number of drawbacks which need acknowledging where mitigating action cannot or has not been taken. In some circumstances the issues arise within the system that the research examines. For instance, the receipt of subject access requests changes the target organisation; those providing information can show bias in its selection; access to data can be withheld; and the research is dependent on the quality of the documents held. Each of these situations can perturb the data presented by an organisation to external individuals. In this instance, the research examines the data received by the individual and as such internal organisational issues are not of interest although could form a basis for

future research. The selection of the target organisations is however an issue as bias can be assumed with the researcher seeking to show themselves through a perceived acceptable persona. This has been mitigated by including the full list of organisations from which the sample was taken in Appendix B and through snowball sampling based on the original selection. Representativeness is an issue in that the data refers to only one individual and therefore it will be important not to make sweeping assumptions based on this sample of one. However, the total sample of 82 organisations may be considered more representative and it is this data that has been analysed and forms a basis for analysis, findings, and discussion. The analysis results from predefined coding schemes but is subject to bias. In this case it was not practical to have a second person work on the analysis but the coding exercise was reviewed in order to improve consistency. Finally, there is an ethical issue of personal privacy. This can normally be mitigated by anonymising data. With an auto-ethnographic approach this is not possible and the publication of lists of organisations provides an insight into the researcher that may be uncomfortable for them, and also perhaps for the reader.

4.2.2 Data Collection

There are two objectives of this phase. The first is to test the categorisation model which created a taxonomy for that data descriptive of an individual held in the digital domain. The second is to investigate the issues surrounding an individual's access to their own digitally extended self. This section will describe the process that was used to obtain data in furtherance of these objectives before the findings are discussed.

Initially a list of 440 organisations which had email or web connections with the researcher was created (Appendix B). Categories, and sectors within category, were then constructed to cover the central and local government, non-governmental organisations (charities), private companies (marketing, online shopping, utilities and marketing information), and public companies (banking, credit reference, insurance, internet search, marketing information, online and high street shopping, online shopping only, social networks, supermarkets, and utilities). This is not a comprehensive categorisation of all organisations dealt with, and has not been statistically derived, for example education is omitted, but seeks to represent the major sectors of interaction of the authors email and web history. The categories are high level cultural classes (Glushko, 2013) based on governance that can be used to classify all UK organisations. The sectors were defined from the 440 organisations in the population, again using cultural norms which provide discrete and identifiable groupings. It is understood that the classes and sectors selected will have

implications for the sample populations within each grouping, and also to an extent on the ability to observe differentiated behaviour. This does not affect the first objective of this phase of the research, to test the model. It is, however, acknowledged that an alternate categorisation could change the nature of further analysis of the observed results if it were done only by the categories and sectors selected here. At this stage in the research, the objective was to observe the quality of the responses to subject access requests and therefore the classification was made mainly to ensure a spread of the organisations chosen within the original purposive sample. Further classifications were constructed for use in the analysis of the data collected, for instance differentiating between location of organisation.

4.2.3 Selection criteria

Organisations were selected in order to provide a manageable sample size of 32. In order to ensure a range of organisations across categories and sectors, from small to large, a purposive sample was taken. As discussed in section 4.2.1, it is recognised that this has the weakness inherent with personal selection which could be eliminated by random sampling. However, random sampling from the full population would have been unlikely to provide the full spread of organisations and therefore the original purposive sample was augmented by a subsequent snowball sample. The base sample, which aligned with the categories and sectors discussed above, was selected to cover small private organisations such as Cult Pens, which may be assumed to hold little personal information, to large international companies such as Amazon, which could be assumed to have extensive data resources. Within the charity sector there is the small Open Rights Group, the multinational Amnesty International and One Voice, which operates in Israel. Some organisations were chosen because of their relationships, where they are known to share data, for example Tesco and Dunnhumby. Within central government contact has been with the larger departments, which have therefore been selected, whilst local government is determined by residential status. Finally, in the public company sector representatives range from the early adopters of information technology such as the financial services sector through to the new industries of search, Google, and social networks, Facebook and Twitter. These companies were also selected based on the assumed volume of information held for example United Utilities have a small footprint in the online contacts that have been experienced whereas Waitrose and John Lewis were expected to have a high volume of contacts. The results of the selection process are shown Table 4.1 below.

Category	Sector	Organisation
Central Government	Central Government	UKBA, HM Revenue and Customs, NHS, ONS
Local Government	Local Government	Oxford City Council, SLDC
NGO	Charity	Amnesty, One Voice, Open Rights Group, RSPB
Private Company	Marketing Information	Dunnhumby, Flurry
	Online Shopping	Boden, Cult Pens
	Utilities	Coop Energy
Public Company	Banking	John Lewis Partnership, Lloyds Bank
	Credit Ref	Equifax, Experian
	Insurance	Zurich Life
	Internet Search	Google
	Marketing Information	Acxiom
	Online & High Street	John Lewis, M&S
	Online Shopping	Amazon, Apple
	Social Network	Facebook, Twitter
	Supermarket	Tesco, Waitrose
	Utilities	United Utilities, Vodafone

Table 4.1 Purposive sample by category and sector

4.2.4 Processes for Data Collection

In order to ensure consistency of approach a process, illustrated in Figure 4.1, for contacting each organisation was developed specifying that a standard letter (Appendix E) be sent to each organisation from the purposive sample, replies analysed, and follow up letters written focusing on perceived issues observed through the analysis of the initial

replies. All letters were sent recorded delivery so that the date of receipt could be recorded and the elapsed time for responses noted for later analysis.

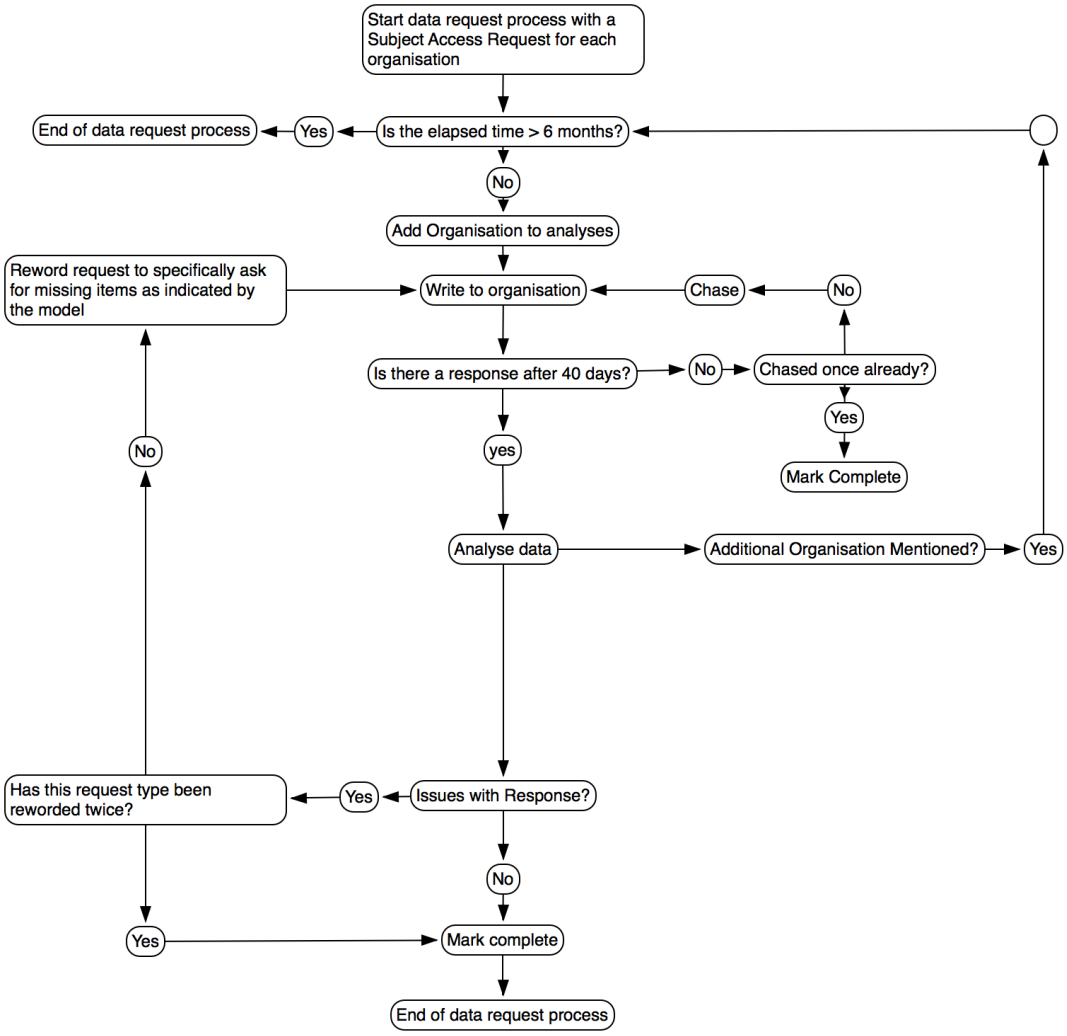


Figure 4.1 Process for data collection

A log was completed for each organisation (Appendix F) in order to record activities and time spent, excepting that time taken scanning documents, as this would not be a necessary part of a data collection process for individuals. This was then summarised onto a spreadsheet showing costs incurred and time spent for each organisation from which data was collected (Appendix G). A second log was kept of the timings of events and whether organisations responded to requests (Appendix H). This data will enable a rough extrapolation for the total cost and time required to explore the authors digitally extended self.

When responses were received from organisations the data was secured. If the response was on paper it was scanned and OCR'd, and the paper copy filed for reference. Electronic

copies were then kept in directories for each organisation on an encrypted laptop, secured to a Tresorit account and backed up to hard disk in 2 locations. The details were then logged and the response analysed. The content was checked for completeness based upon the researcher's knowledge of the data landscape and where appropriate, internet searches. In addition, all documents were examined for evidence of the components defined in the model, and the presence, or absence, recorded. These notes from each organisation were kept in a journal (Appendix I) and the responses assessed according to the predefined criteria (Appendix J) and recorded on a spreadsheet for later study (Appendix K). Where anomalies or deficiencies in the replies were revealed, a follow up request for clarification was sent following the same system as the original request. Where organisations revealed that data was received from, or sent to, other named parties the standard subject access request letters were sent, to those organisations, following the process defined above, thus forming a snowball sample extending out from the original purposive sample.

Both the journal and analysis enabled the responses to be compared against the elements of the categorisation model, thus the model could be assessed for completeness. When starting the journal, notes were made on an illustration of the model depicting areas of the model that were not mentioned in the responses or were either partially or fully considered. Additional entries were noted when data was stored outside the UK, and for those organisations that provided data to, or received data from, the responding organisation. This provided a very visual analysis but was time consuming. Later entries used a standard format with headings for digital footprints, third party digital footprints, digital persona, where data is held, external data from, and external data to.

Once annotated into the journal the responses were analysed by reference to the journal and where necessary to the original documents. For each of the elements, digital footprint, third party digital footprint, digital persona, data from, and data to, the responses were analysed and the results compared to the scoring matrix (Appendix J). Scores were allocated by combining two criteria. First, if data was absent, had been provided in part, or in full. The second, an assessment of whether an organisation held data, and whether the data had been partly or fully provided was in part subjective. The judgement involved personal knowledge of the relationship between the data subject and the organisation, and where available, access to additional data. Vodafone for example provided no details of the digital footprints laid down when calls were made, omitted the third party digital footprints created by others when calling and texting, but provided those created when data was entered by shop staff. In this instance, it was clear that digital footprints had been omitted from the response, and that third party digital footprints had been partially reported. The

second criteria aims to provide some gauge on the level of openness but is more judgemental. It scores disclosure of information that was known to exist, lower than when there was an uncertainty about the data's existence, with the highest score given where there was no evidence known to this researcher that the reported data existed. This process was repeated when responses were received to follow up letters.

This procedure was followed for the duration of the data collection phase, which allowed a four-month period for sending the initial subject access requests and follow up data requests. Some adjustments were made to the format of the journal to improve its structure and to help with the comparison of data received to the model. An analysis of the data will follow illustrating the potential issues for an individual in researching their own digitally extended self, an assessment of the effectiveness of the model, an evaluation of organisation's responses and an evaluation of the issues that arose.

4.3 Results

The analysis of the number of organisations contacted and those that received follow-up enquires is shown below in table 4.2. Communications were initially sent to 32 organisations plus a further 51 from the snowball sample requesting data that they held, which was descriptive of this researcher. Two organisations, John Lewis and Waitrose (being the same company), chose to send a combined reply. As a result, 82 organisations are analysed in this chapter, of which 31 were from the initial purposive sample, and 51 from the snowball sample. There were 19 organisations who did not answer the requests for data, one from the purposive sample and 18 from the snowball sample. Therefore, the number of responses received was 63, of which five provided no data leaving 58 organisations from which some data was received. Of the five organisations providing no data, four were data processors and one was a research organisation. Follow-up communications requesting further information were sent to 29 organisations, 24 from the purposive sample and 5 from the snowball sample. From these 20 replies were received, 18 from the organisations in the purposive sample and two from those in the snowball sample.

Numbers of Organisations Sampled		
The number or organisations in the initial purposive sample		32
Less 1 as John Lewis & Waitrose provided a joint reply		1
Unique organisations in initial purposive sample	(A)	31
Organisations obtained for the snowball sample		59
Less those found after the cut-off date		8
Organisations in snowball sample	(B)	51
Total number of organisations contacted	(A + B)	82
Organisations that did not respond		
Purposive sample		1
Snowball sample		18
Total that did not respond	(C)	19
Total of responses received	(A+B-C)	63
Less organisations that did not provide data		5
Total of organisations that provided some data		58
Number of Organisations with Follow-up Contact		
Follow-up contacts from the purposive sample		24
Less those that did not respond		6
Responses received from follow-up contact with the purposive sample		18
Follow-up contacts from the snowball sample		5
Less those that did not respond		3
Responses received from follow-up contact with the snowball sample		2
Total responses from follow-up requests		20

Table 4.2 Significant numbers in the data collection exercise

The objective of this second phase of research is to examine the effectiveness of the categorisation model in describing an individual's data, and to gather information that can provide insights into the difficulties that an individual can have in exploring their own digitally extended self. This section will start by examining the issues that faced the author when collecting data to validate the categorisation model, defined in Chapter 3, and to investigate his own digitally extended self. Section 4.3.2 then reports on the communications from organisations, whilst section 4.3.3 examines whether the data

provided validated the model presented in Chapter 3. Finally, section 4.3.4 describes the attributes of organisations used in the analysis to be found in section 4.4.

4.3.1 Impact on the Individual

An analysis of one's own digitally extended self provides an insight into a life, and to an extent allows us to see ourselves as others see us. In this way it acts as a restricted form of life logging as described by Bell, Gemmel & Haag (2009). When life logging, collection, curation, and deletion of data is under the control of the individual. This is not the case with that data which makes up the digitally extended self. The individual, although supported by legislation, is dependent upon organisations to provide information.
This section addresses the second research question:

RQ2: How feasible is it for an individual to obtain the information, held by organisations, which is descriptive of them?

The first thing that this individual noted when trying to research their digitally extended self was the time taken. A list of organisations dealt with (Appendix B) had already been constructed and maintained. From this list a purposive sample of 32 organisations was created (Appendix C) and the time taken to research, communicate with and analyse the responses was recorded on time sheets (Appendix F). This data, together with postage, and each organisation's administrative costs were recorded for the original purposive sample and the secondary snowball sample of 51 organisations (Appendix G). From this diary it has been observed that the time taken varied between 0.85 and 6.33 hours per organisation with a mean time of 1.43 hours.
As each response was received, the time for handling the response was recorded. A decision was made however, not to record the time taken to scan and OCR the paper responses. The rationale was that an individual would not necessarily want or need to do this work although it was essential for this research in that it provides portability and security. This was by far the most time-consuming element of the collection process, for example on a standard document scanner (in this case a HP Photosmart 7510) the 242 pages of John Lewis Partnership Card data took 359 minutes (6 hours) to scan, and that does not include paper handling and organising once scanned. This is effectively a full working day for one reply out of the 58 where data was provided.
The costs, which included postage and the £10.00 fee for a Subject Access Request, averaged £12.78 per organisation. Some organisations such as John Lewis, Amazon and Cult Pens returned the £10.00 cheque, one can hypothesis that few organisations have the facility to deal with cheques, or else consider such requests a customer service. The costs

for postage varied depending upon whether a single letter was sent or a second was necessary to raise queries about the response. The costs per letter varied depending on whether they were posted at a public Post Office or private service (such as Mail Box), which charged a higher fee. Letters sent outside of the UK were also more expensive. The costs could have been reduced by the use of first class or even second-class mail but then there would have been no record of delivery. With hindsight, this may have been a better strategy as assumptions on delivery time could have been made. However, follow up letters would not have been able to assert that the first letter was delivered, and also it could be surmised that recorded delivery letters are treated more seriously than others, although no evidence on this has been found.

Organisational responses to this research varied from a single page (e.g. Dunnhumby) to Lloyds Bank who provided 1086 pages in response to the initial request. In all, the paper responses formed a substantial record and measured over 15 inches in height (reduced by double sided printing), see photographs in Appendix L. In addition, several organisations replied by email providing the responses electronically. By the end of the data collection phase the storage directory held 11.3 GB of data.

It was rare to receive details of analytics that had been undertaken by organisations, they may be considered to be outside of the scope of the Data Protection Act 1998, but also not all organisations have analytic capability. When they were provided, it was unsettling to see how inaccurate they could be. For instance, Acxiom assume a 97% probability of employment and zero percent probability of being retired (the author retired over 10 years ago); an 8% likelihood of interest in charities whilst I volunteered for a major charity trust and subscribed to over 8 different charities. On the other hand, the basic data appeared accurate. However, I found it unsettling that others should have such detailed insights into my life, why should others know that on the 9th December 2004 I spent £59.90 in Hove's Othello restaurant?

The data analysed is from the 32 organisations from the purposive sample and the 51 organisations from the snowball sample that were contacted before the time limit in the research was reached. For the full digitally extended self to be researched all 440 of the recorded organisations (Appendix B) would need to be contacted. In addition, organisations that were disclosed as, either providing information to, or receiving information from, these organisations would need to be contacted. The original sample referred a further 59 organisations of which 46 were not already included in the list of recorded organisations. If this were reflected by all of the 440 recorded organisations a further 633 could be expected to hold information describing the subject individual. On the

other hand, this could be an over estimate as reference may be made to the same secondary organisations. However, the 633 referred organisations will have the possibility to refer to further organisations, and they would be able to refer to others etc. Evidence has not been collected which would enable any reasonable estimates to be made regarding the extent of this digitally extended self.

If the assumption of an additional 633 organisations is taken then the total cost of exploring this digitally extended self would be £9,841, and the time taken would be 1605 hours of effort. There would be approximately 142GB of data and the paper record would reach 193 inches in height. Despite this it is unlikely that the full digitally extended self would have been explored. Organisations, (e.g. John Lewis), do not name other organisations who provide data to them, or to whom they send data, neither do they supply information regarding digital personas. Additionally, personnel within organisations may not provide complete data because of ignorance, poor workmanship, or even company policy. Finally, that data that is provided gives only a snapshot in time of an individual's constantly changing digitally extended self.

4.3.2 Responses received

Of the original purposive sample of 32 organisations, responses were obtained from 31, with 1 organisation failing to respond to the initial and follow up requests. Of these 31 organisations, 2 combined their responses (John Lewis and Waitrose), and 24 were asked for additional information. Of these 18 answered with 6 ignoring the follow up request. The snowball sample provided 59 further organisations. Of these 51 were contacted, the remaining 8 were discovered after the time frame for initiating data collection had ended. Of these 33 responded, and 18 failed to answer the requests. Five organisations were asked for additional information with 2 answering and 3 ignoring the follow up request. Organisations varied greatly in size, capability and willingness to respond to subject access requests. This created variances in response times. A minority of organisations ignored their legal obligation to respond (e.g. Cooperative Energy), whilst others are resident outside the UK and take advantage of arguably lax local enforcement within the EU (e.g. Facebook in Ireland), finally others are outside European jurisdiction and either ignore requests (e.g. Ancestry), or construct artificial barriers to information availability (e.g. Mail Chimp 'we can't comment on, divulge information about, or block access to a Customer's account unless you submit to us a valid court order or subpoena from the State of Georgia', Appendix Y). Finally, organisations may go into administration and the administrators fail to respond (e.g. Rogavi).

From the original purposive sample of 32 organisations 31 responded, only Cooperative Energy failed to answer. Figure 4.2 illustrates the mean response time was 41 days, outside the 40 day response time required by UK law under the Data Protection Act 1998. The fastest response was received the same day from Facebook, whilst the longest time taken was by Twitter, 118 days. Follow up communications were sent to 24 of the 31 organisations and 17 replies were received, 7 failed to answer. Note that the follow up requests are not worded as formal subject access requests but referenced the original request. The mean response time to follow-up requests was 26 days. The fastest response was received from Flurry in 1 day, whilst the slowest was from John Lewis Partnership in 71 days.

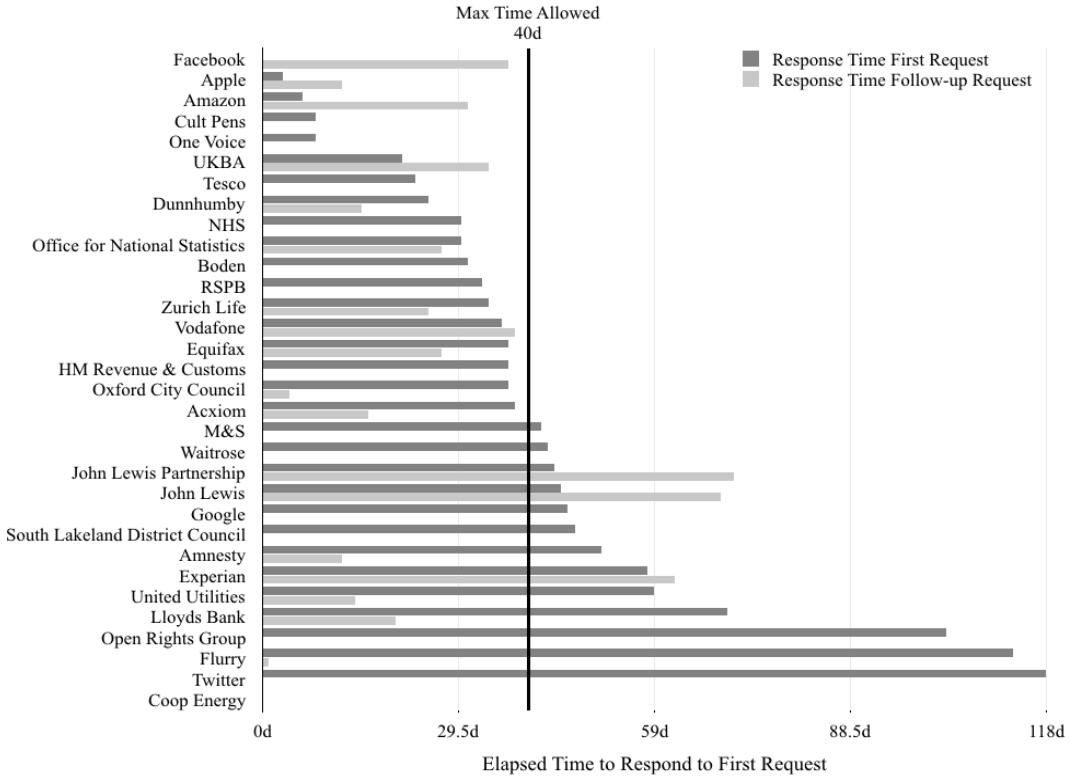


Figure 4.2 Elapsed time to respond to initial and follow-up requests for purposive sample

From these figures, it is clear that the process of investigating one's digitally extended self can be expensive in terms of time and money, but also long in elapsed time. The next sections discuss the data received in comparison to the model, and in terms of completeness.

4.3.3 Did the data fit the model?

In Chapter 4, it was argued that the terminology used to label personal data was inconsistent and a proposed new nomenclature explained. This was then illustrated in the

form of a model, which categorises, and shows the relationships between, data descriptive of an individual. The model was then validated against the literature. This Chapter describes how responses from 64 organisations (the 31 that responded from the original sample, and 33 that responded from the snowball sample) have been classified into the categories defined within the model.

The volume of data is such that it was unfeasible to categorise each data element. For instance, John Lewis Partnership Card, managed by HSBC, replied with 242 pages of information. Some of these contained a bibliography of codes whilst others contained 55 transaction lines. Some organisations provided less pages of data, others more. The method adopted therefore was to read each reply and allocate the groupings of data to the categories in the model. This was recorded in a journal with an entry for each (see Appendix I for an example). During this process, all data items descriptive of an individual were able to be matched to categories within the model.

One aspect of the data however was not fully covered in the model, which looks at an individual's digitally extended self as a whole, whilst an organisation holds a subset of that data. The missing element is the way that organisations move data between themselves. The model catered for the input of digital persona and second level data into the digitally extended self from the outside, but not for the movement, or sharing, of parts of the digitally extended self between organisations.

To illustrate this issue the centric visualisation of the model of the digitally extended self is introduced below. This is constructed with *digital footprints* at the heart of the data, which describes an individual. It is then incrementally extended through the concept of multiple artefacts forming a *digital mosaic*, identified by the inner circle. Next analyses are formed using data from digital footprints and external sources resulting in *digital personas*. The whole within the outer circle is named the *digital extended self*.

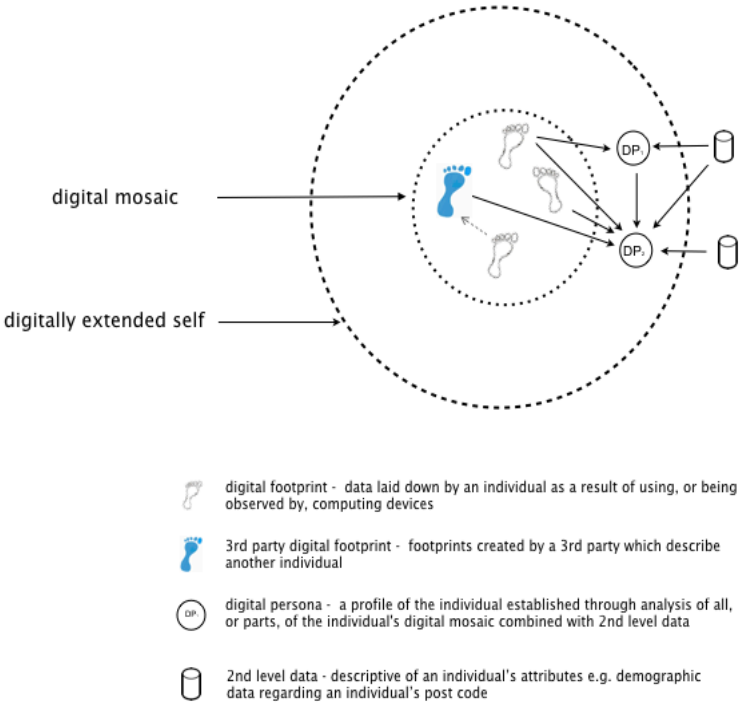


Figure 4.3 Centric visualisation of the model of the digitally extended self

The centric visualisation represents the whole of the digitally extended self; however, data is transferred between organisations. This is shown in figure 4.4 where the digitally extended self is depicted composed of multiple sets of data held by a number of organisations, between which, data may be exchanged.

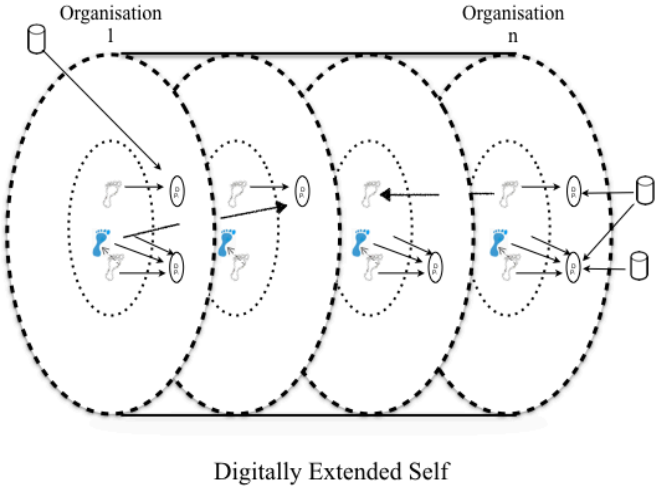


Figure 4.4 Deconstructed centric visualisation of the digitally extended self showing organisational instances

In the centric visualisation figure 4.3 the data categories remain intact and illustrate all of an individual's data. Movement of data between organisational instances is not represented. The location of the data, whilst of interest to the individual, is an attribute of the data rather than of the individual. It answers the question, who knows about me and

where are they, rather than what is known about me. This may be compared to the categorisation of books in libraries that allows for them to be grouped by topic but does not track to whom they may be loaned at any one time. The location of the book is an attribute of the book rather than its contents.

The location and movement information associated with an individual's digitally extended self is important in order that its extent can be discovered, and to help when exploring its spread from organisation to organisation. Unlike a library where location is a matter of knowing where possessions are located, the digitally extended self has a more intimate relationship with people, with the ability to constrain or enhance a life, and as an extension of the mind as we outsource our thinking into the digital world.

4.3.4 Attributes of Organisations Used for Analysis

This document has so far examined the findings relevant to the categorisation model and its validation. The next sections will address the following research question using the data collected during this research phase

RQ3: What is the quality of the personal data returned by organisations when it is requested by individuals?

The data was analysed using the categories in the model; digital footprints, third party digital footprints, digital persona, and second level data used in the construction of digital persona. In addition, the movement of data was also included using the following categories; imported digital persona, data from other organisations (both forms of second level data), exported digital persona, and data to other organisations.

It should be noted that no specific information was received under the category of 'second level data used in the construction of digital persona', this information is not covered by the Data Protection Act 1998 as it is not a direct attribute of an individual but is an attribute of something associated with a person. An example would be demographics associated with a house due to its characteristics such as size and location. As a result, it has not been included in the following analysis. A column was added to cover a similar situation where data such as credit balance is imported and then used in the calculation of a digital persona, it is labelled 'External Data Imported for Use in Digital Persona'. This is in effect a specific type of 'data from other organisations' delineated by its intended use.

In addition to the categories defined within the model, and those associated with the movement of an individual's data, additional characteristics of organisations were included for analysis. Two were associated with location. First, the location of the organisation was noted, this is important as regulation regarding public access to data differs by country. It

was summarised under three categories, in the UK, in the EU but outside the UK, and outside the EU. These categories were chosen in order to provide sufficient granularity but also because requests are covered firstly by UK law, that in turn is subject to EU directives. Of the 82 organisations contacted 66 were in the UK, 4 were in Europe but outside the UK, and 12 were outside of Europe.

The second location-based characteristic, which was part of the data requested from all organisations, was where its data was held. There is no right of access to this information and it was sparsely provided. It is important, however, as local laws dictate how data should be treated. At the time of data collection, there was an expectation that data held in the USA was treated securely under the Safe Harbour agreement, which was replaced in 2016 by the EU-US Privacy Shield. The question was asked because the Snowden revelations cast further serious concerns regarding their worth. Again, the categories used were within the UK, within the EU outside the UK, and outside the EU. In this instance of the 82 organisations contacted 58 replied with some information of which 11 stated that their data was held within the UK, 1 in Europe but outside the UK, and 11 outside of Europe, 35 organisations chose not to answer.

Finally, two other attributes of organisations were defined, category and sector. The categories were used in selecting the original purposive sample and remain valid for the subsequent snowball sample. However, the sectors needed to be expanded for example to include IOS apps, the full category sector list can be found in Appendix M.

The next section presents the major findings from this phase of the research. It comprises four sections; first a high-level view covering all organisations; second a comparison between categories; third by sector and finally by location

4.4 Analysis

4.4.1 Introduction

Given the relatively large amount of information collected (82 columns of 82 rows) a range of analyses is possible. At this stage, no attempt has been made to cross tabulate the data in order to draw out any possible correlations. Nevertheless, the results provide insights into the data that organisations are willing to provide, the advantages obtained from writing follow up letters, variations across differing categories and sectors, and finally the location of organisation and data.

4.4.2 General Findings

Tables 4.3 and 4.4, below, show findings from the responses to the subject access requests analysed in terms of the constituent parts of the model described in Chapter 3. Table 4.3 shows the results from the initial analysis and Table 4.4 includes responses from any follow up letters that were received. The cells of the scoring matrix (Appendix J) are shown for each category of analysis (e.g. Digital Footprints). The table headings in this representation show the degree to which data was provided, whilst the assessment of openness is shown vertically. For example, Table 4.3 shows that 7 organisations did not provide digital footprint data even though my assessment was that they held it, and that this constituted 12% of the 58 organisations who provided information.

Figure 4.5 is derived from these tables and shows the percentage of organisations providing at least some information for each category of analysis, derived from the model of the digitally extended self. The first part of the bar is calculated from the responses from the first communication and the second part shows the increased data provided as a result of any second communication. The elements of the digital mosaic (digital footprints and third party digital footprints) were provided by 60% or more of the organisations whilst information relating to digital persona by less than 29%. The impact of the follow up letters, which were sent to 29 organisations, was most pronounced for digital persona information with a further 9% of all organisations responding with data, a 47% increase.

		Not Provided		Partially Provided		Fully Provided		Totals	
		No	% of all responses	No	% of all responses	No	% of all responses	No	%
Digital Footprints	**Evidence of Data**	7	12%	4	7%	25	43%	36	62%
	Suspicion of Data	4	7%	0	0%	3	5%	7	12%
	No Suspicion of Data	14	24%	0	0%	1	2%	15	26%
	Total	25	43%	4	7%	29	50%	58	100%
3rd Party Digital Footprints	**Evidence of Data**	7	12%	9	16%	15	26%	31	53%
	Suspicion of Data	2	3%	2	3%	4	7%	8	14%
	No Suspicion of Data	16	28%	0	0%	3	5%	19	33%
	Total	25	43%	11	19%	22	38%	58	100%
External Data Imported for Use in Digital Persona	**Evidence of Data**	8	14%	3	5%	0	0%	11	19%
	Suspicion of Data	12	21%	3	5%	3	5%	18	31%
	No Suspicion of Data	24	41%	1	2%	4	7%	29	50%
	Total	44	76%	7	12%	7	12%	58	100%
Computed Digital Persona	**Evidence of Data**	3	5%	3	5%	0	0%	6	10%
	Suspicion of Data	11	19%	0	0%	1	2%	12	21%
	No Suspicion of Data	33	57%	0	0%	7	12%	40	69%
	Total	47	81%	3	5%	8	14%	58	100%
Imported Digital Person	**Evidence of Data**	3	5%	1	2%	0	0%	4	7%
	Suspicion of Data	9	16%	0	0%	4	7%	13	22%
	No Suspicion of Data	36	62%	0	0%	5	9%	41	71%
	Total	48	83%	1	2%	9	16%	58	100%
Exported Digital Persona	**Evidence of Data**	2	3%	4	7%	0	0%	6	10%
	Suspicion of Data	4	7%	0	0%	1	2%	5	9%
	No Suspicion of Data	47	81%	0	0%	0	0%	47	81%
	Total	53	91%	4	7%	1	2%	58	100%
Data from Other Sources	**Evidence of Data**	8	14%	3	5%	1	2%	12	21%
	Suspicion of Data	13	22%	3	5%	6	10%	22	38%
	No Suspicion of Data	13	22%	1	2%	10	17%	24	41%
	Total	34	59%	7	12%	17	29%	58	100%
Data to Other Sources	**Evidence of Data**	7	12%	3	5%	2	3%	12	21%
	Suspicion of Data	16	28%	1	2%	9	16%	26	45%
	No Suspicion of Data	9	16%	2	3%	9	16%	20	34%
	Total	32	55%	6	10%	20	34%	58	100%

Table 4.3 Analysis of responses from the first communication (data providing organisations only)

		Not Provided		Partially Provided		Fully Provided		Totals	
		No	% of all responses	No	% of all responses	No	% of all responses	No	%
Digital Footprints	Evidence of Data	5	9%	4	7%	27	47%	36	62%
	Suspicion of Data	4	7%	0	0%	3	5%	7	12%
	No Suspicion of Data	14	24%	0	0%	1	2%	15	26%
	Total	23	40%	4	7%	31	53%	58	100%
3rd Party Digital Footprints	Evidence of Data	5	9%	7	12%	20	34%	32	55%
	Suspicion of Data	2	3%	1	2%	4	7%	7	12%
	No Suspicion of Data	15	26%	0	0%	4	7%	19	33%
	Total	22	38%	8	14%	28	48%	58	100%
External Data Imported for Use in Digital Persona	Evidence of Data	7	12%	2	3%	1	2%	10	17%
	Suspicion of Data	10	17%	3	5%	5	9%	18	31%
	No Suspicion of Data	24	41%	3	5%	3	5%	30	52%
	Total	41	71%	8	14%	9	16%	58	100%
Computed Digital Persona	Evidence of Data	3	5%	4	7%	0	0%	7	12%
	Suspicion of Data	6	10%	2	3%	3	5%	11	19%
	No Suspicion of Data	33	57%	0	0%	7	12%	40	69%
	Total	42	72%	6	10%	10	17%	58	100%
Imported Digital Person	Evidence of Data	3	5%	1	2%	0	0%	4	7%
	Suspicion of Data	7	12%	1	2%	4	7%	12	21%
	No Suspicion of Data	37	64%	0	0%	5	9%	42	72%
	Total	47	81%	2	3%	9	16%	58	100%
Exported Digital Persona	Evidence of Data	2	3%	4	7%	0	0%	6	10%
	Suspicion of Data	4	7%	0	0%	1	2%	5	9%
	No Suspicion of Data	47	81%	0	0%	0	0%	47	81%
	Total	53	91%	4	7%	1	2%	58	100%
Data from Other Sources	Evidence of Data	7	12%	2	3%	2	3%	11	19%
	Suspicion of Data	11	19%	3	5%	8	14%	22	38%
	No Suspicion of Data	12	21%	3	5%	10	17%	25	43%
	Total	30	52%	8	14%	20	34%	58	100%
Data to Other Sources	Evidence of Data	6	10%	4	7%	3	5%	13	22%
	Suspicion of Data	13	22%	2	3%	9	16%	24	41%
	No Suspicion of Data	9	16%	2	3%	10	17%	21	36%
	Total	28	48%	8	14%	22	38%	58	100%

Table 4.4 Analysis of responses after any second communication (data providing organisations only)

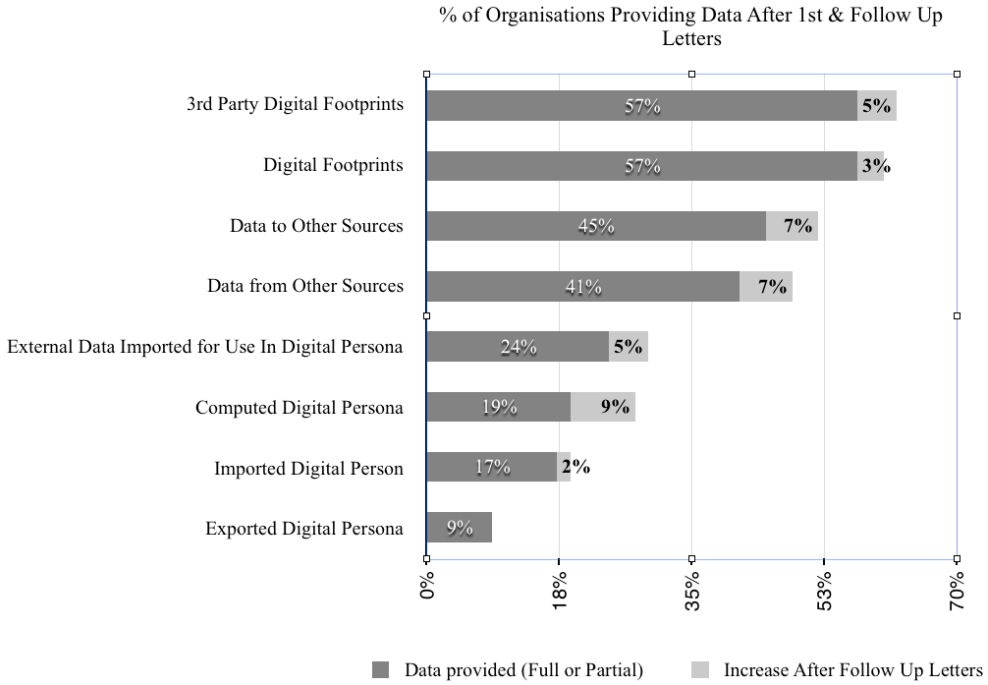


Figure 4.5 Percentage of organisations providing data, by data classification

The next four figures use responses from all organisations contacted. Figure 4.6 shows whether or not data was provided, and Figure 4.7 represents an assessment of the completeness of that data using scores from the matrix in Appendix J. These heat maps reflect the responses with respect to the components of the categorisation models (digital footprints, third party digital footprints, and digital persona). Figures 4.8 and 4.9 show the same representations for data movements.

On each of the four heat maps two groups of organisations can be observed. The first, shown in white, are those 16 (20%) organisations that failed to answer subject access requests. These are predominantly IOS app developers, mostly outside the UK whose locations were difficult to find and who were therefore contacted by email but failed to respond.

The second group of 8 organisations (10%) are shown in grey and are largely those who are exempt from supplying data. For instance, Engaging Networks processed data on behalf of other organisations, for example Open Rights Group and The Refugee Council, who were the data owners and had the obligation to supply data in response to subject access requests.

From Figure 4.6, it would appear that the organisations that provided digital footprint data are more likely to have shared additional information. Indeed, digital personas were almost exclusively provided by such organisations. The exceptions, Acxiom, Equinox, Experian, Call Credit and Zurich Life, are however, instructive. There had been no direct digital contact with any of them and so no digital footprint data could have been provided.

Figure 4.7 presents a subjective view of the likelihood of organisations having held data combined with whether data was provided. The groupings are based on the scoring matrix (Appendix J). A low score of 1 (dark red) indicates that there is evidence of data (e.g. I have laid down digital footprints with the UKBA through facial recognition at airports) but no data has been provided. Where not all of the known data has been provided then a score of 4 (dark blue) results (e.g. South Lakeland District Council provided some data but omitted electoral roll information), and a high score of 9 (light green) represents data being provided where I had no suspicion that data existed (e.g. the digital persona curated by Not On The High Street). In this heat map a score of 3 indicates that no data was expected to be held and none has been provided and is shown as dark grey.

Figure 4.7 shows a similar pattern to that of Figure 4.6 but highlights two groupings. The first in shades of red, where there is evidence or suspicion of data being held, but where no data was provided, and the second where the large areas shaded grey, indicate that though data was not provided none was expected. As this is a subjective measure it may be considered to be as much a measure of my knowledge or scepticism as that of the willingness of organisations to supply data.

The final two figures illustrate the same analysis but applied to categories of data movement; data to other sources, data from other sources, external data imported and used in the creation of digital personas, imported digital personas, and finally exported digital personas.

Figure 4.8 illustrates whether data was provided (without any assessment of the likelihood of data being available for disclosure). As we have seen in Figure 4.5 organisations were less inclined to provide information about the movement of data than about the data itself, and less likely again to provide information about the movement of digital personas, and data associated with them. This information is not covered by the Data Protection Act 1998 and so did not have to be provided, however, it may have been that a smaller percentage of organisations used digital persona than some writers predicted (Baker, 2008), or were unwilling to admit that they did so. Figure 4.8 shows that information relating to the movement of digital personas was only provided by those organisations that also supplied other data movement details.

The heat map in Figure 4.9 takes account of my assessment of the likelihood of data movements taking place. Again, this is subjective and I suggest more likely than previous examples to be prone to error as the movement of data by organisations often leaves no traces that can be observed by the individual. The assessment is based on a number of factors. First, it may be public knowledge that data is exchanged (e.g. UKBA), or an

organisation may state that it exchanges data with another (e.g. Experian revealed that it exchanges data with Joseph Turner). Some organisations declared that they exchanged data but would not name the other organisations (e.g. John Lewis), whilst other businesses depend on information being exchanged as part of their business model (e.g. Equifax). Finally, in a small number of cases it would seem to the author surprising that an organisation did not obtain data from, or send data to, other organisations (e.g. Amazon and Apple). Three groups can be observed. Firstly, organisations which may be expected to be involved in the movement of data, but did not disclose any information relating to the movement of information, (e.g. UKBA, South Lakeland District Council, Facebook). Eighteen organisations (32% of those that provided some data) fall into this category. The second group are those that provided some data but not all that was expected, (e.g. HMRC, Oxford City Council, and Acxiom). There were eight organisations in this grouping, which equates to 14% of those organisations that provided data. The final grouping, where no data was thought to be withheld, consisted of thirty organisations or 52% of those responding with data. There were no central or local government organisations in this group but examples such as Not on The High Street, Orvis, Lloyds Bank and Vodafone stood out. On the face of it there seems to be a polarisation. On the one hand those organisations who chose not to divulge data pertaining to the movement of data in and out of their organisations, and on the other organisations who are happy to share, at least some of, that data. Only a small number of organisations disclosed some data but appeared reluctant to reveal other sources of data movement.

Figure 4.6 Heat map of data provided for model categories

Key: red, no data provided; green, partial or full data provided

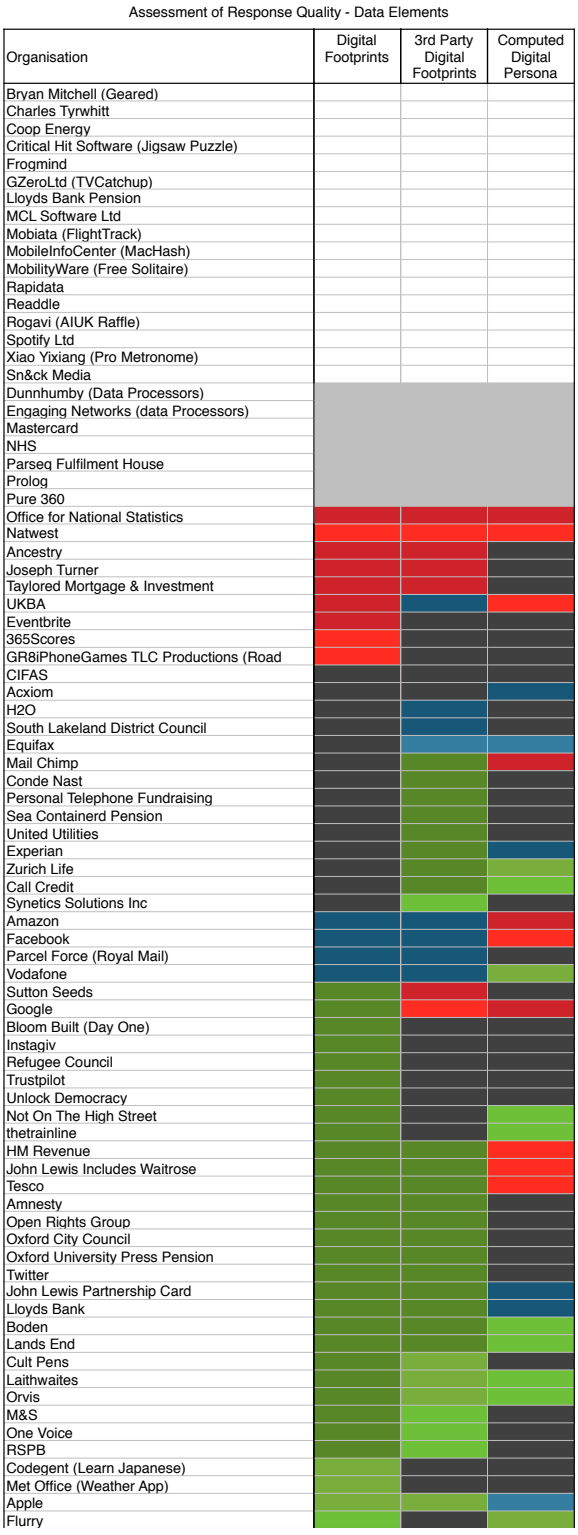


Figure 4.7 Heat map of assessment of response quality for model categories

Key: red, no data provided; black, no data provided nor expected; blue, partial data provided; green, full data provided

Data Provided - Data Movement Elements

Figure 4.8 Heat map of data provided for data movement elements

Key: red, no data provided; green, partial or full data provided

Assessment of Response Quality - Data Movement Elements

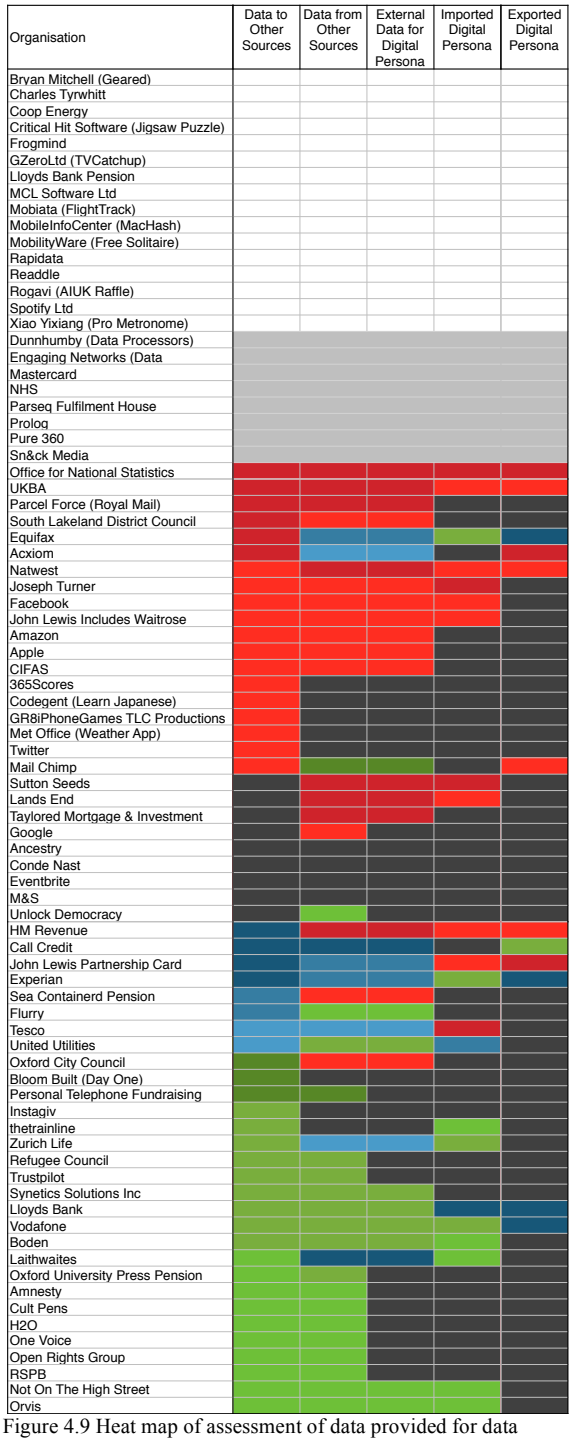


Figure 4.9 Heat map of assessment of data provided for data movement elements

Key: red, no data provided; black, no data provided nor expected; blue, partial data provided; green, full data provided

4.4.3 Findings by Organisation Category

When considering the purposive sample, organisations were separated into 5 categories, central government, local government, public companies, private companies, and NGOs (not-for- profit organisations). The following two tables map the supply of data by organisational category from the 58 organisations that responded to subject access requests by providing some data. The values within each table represent the percentage of organisations, within each category, providing the relevant data elements, the first row indicates the % of organisations that provided partial data and the second row indicates the % of organisations providing full data . Table 4.5 represents the position after the first, and Table 4.6 after the final communication. As previously discussed there is an element of subjectivity within the scoring with respect to whether full or partial data has been provided, as this can only be based on personal knowledge of the data and the organisation. Table 4.5 shows that the performance for central and local government are lower than for other categories, with NGOs outperforming all others. However, it should be noted that each of these categories has a relatively low sample size that may affect the results.

	No	**Overall % of Elements Provided**	**Digital Footprints**	**3rd Party Digital Footprints**	**Computed Digital Persona**	**Imported Digital Persona**	**Exported Digital Persona**	**External Data for Digital Persona**	**Data from Other Sources**	**Data to Other Sources**
All Categories	*58*	*9*	*7*	*19*	*5*	*2*	*7*	*12*	*12*	*10*
		24	*50*	*38*	*14*	*14*	*2*	*10*	*29*	*34*
Central Government	4	3	0	0	0	0	0	0	0	25
		9	50	25	0	0	0	0	0	0
Local Government	2	13	0	100	0	0	0	0	0	0
		6	50	0	0	0	0	0	0	0
NGO	6	0	0	0	0	0	0	0	0	0
		46	83	83	0	0	0	0	100	100
Private Company	21	1	0	0	0	0	0	5	5	0
		30	57	33	24	19	0	19	38	52
Public Company	25	19	16	36	12	4	16	24	24	20
		17	36	36	12	16	4	8	12	12

Table 4.5 Percentage data transparency score by category of organisation after 1st communication for responding organisations

Note: for each category the first row indicates the % of organisations that provided partial data and the second row indicates the % of organisations providing full data.

Table 4.6 shows the position at the end of the data collection process. Central and local government are still the lowest performing categories. NGOs have been overtaken by public companies as a result of the latter supplying data related to digital personas, which it could be assumed NGOs do not possess. The effect of follow up communications has, however, had a greater impact on some categories than others.

	No	**Overall % of Elements Provided**	**Digital Footprints**	**3rd Party Digital Footprints**	**Computed Digital Persona**	**Imported Digital Persona**	**Exported Digital Persona**	**External Data for Digital Persona**	**Data from Other Sources**	**Data to Other Sources**
All Categories	*58*	*10*	*7*	*14*	*10*	*3*	*7*	*14*	*14*	*14*
		28	*53*	*48*	*17*	*16*	*2*	*16*	*34*	*38*
Central Government	4	6	0	25	0	0	0	0	0	25
		9	50	25	0	0	0	0	0	0
Local Government	2	6	0	50	0	0	0	0	0	0
		19	50	50	0	0	0	0	0	50
NGO	6	0	0	0	0	0	0	0	0	0
		48	100	83	0	0	0	0	100	100
Private Company	21	1	0	0	0	0	0	5	5	0
		33	57	38	24	24	0	24	43	52
Public Company	25	22	16	24	24	8	16	28	28	28
		23	40	52	20	16	4	16	20	16

Table 4.6 Percentage data transparency score by category of organisation after final communication for responding organisations

Note: for each category the first row indicates the % of organisations that provided partial data and the second row indicates the % of organisations providing full data.

Figure 4.10, below, takes information from the tables to illustrate the relative responses for data provision. The increase in percentage response is greatest for local government followed by public companies.

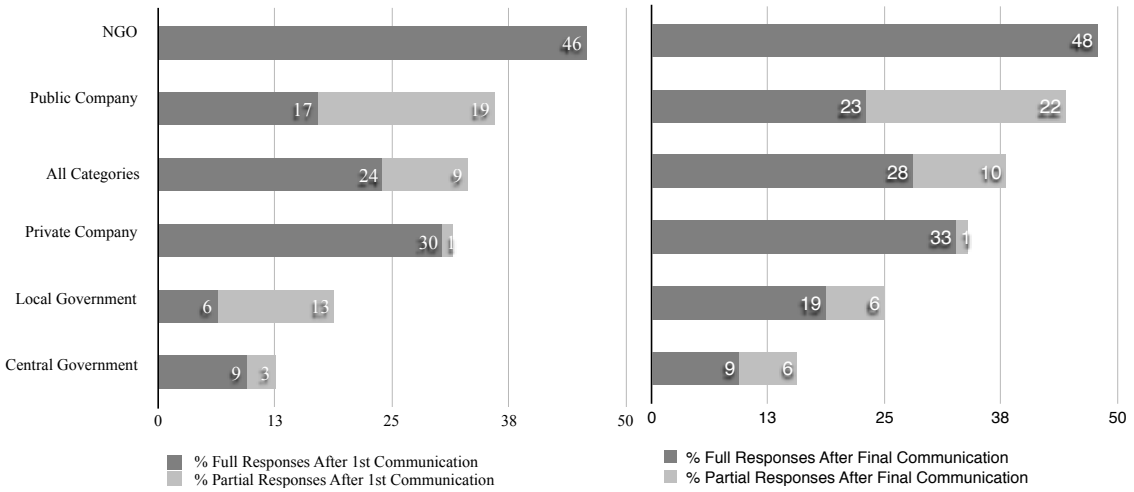


Figure 4.10 Responses with respect to data provision by organisational category for responding organisations

Table 4.7 summarises the responses received from organisations analysed by category and the number of elements provided (i.e. digital footprint, 3rd party digital footprint etc.). It is based on the full sample of 82 organisations and therefore reflects the impact of those that failed to respond.

Category	**Number**	**Total Elements**	**Elements / Organisation**
NGO	6	22	3.7
Public Company	28	68	2.4
Local Government	2	3	1.5
Central Government	5	6	1.2
Private Company	41	47	1.1
Total	82	146	1.8

Table 4.7 Responses received by organisational category and number of elements provided

NGOs provided the most data elements per organisation with over twice the average whilst local and central government, and private companies were below average. Of the 41 private companies, 12 chose not to respond to the requests, the majority of whom were IOS app developers. A further 6 organisations classified themselves as data processors and were therefore exempt from providing information. Of these, data was typically subsequently provided by the data owners. One notable exception was Dunhumby who as data processor refused to provide information (e.g. Clubcard data), as did Tesco the data owner. Having looked at the numerical analysis by category of the data returned by

organisations, the following section describes the same data but from a market sector viewpoint.

4.4.4 Findings by Organisation Sector

The following two tables map performance by organisational sector from the 58 organisations that responded to subject access requests by providing data which is categorised into component parts of the model of the digitally extended self described in Chapter 3, and are in alphabetical order. The values within each table are calculated in the same way as those in Tables 4.3 and 4.4 above. The sectors were chosen as terms with which to group similar organisations hopefully in a way that may be helpful in illustrating different approaches to the provision of data, so credit reference companies are distinguished from finance institutions, even though both may calculate credit limits. As already discussed, there is an element of subjectivity in the allocation of scores, as it is based on personal knowledge of the data available. In addition, the sample size of each sector is low, showing only those organisations that responded to the subject access requests. There were two organisations classified as 'data processing' companies, but as can be seen neither supplied any data, as they acted as data processors for other organisations who were the data owners, and therefore provided the responses.
Table 4.8 shows, that for responding organisations, the shopping, charity, marketing, and credit reference sectors are above average whilst government (local and central), and internet organisations perform the worst. This situation is (in the case of shopping, marketing, and credit reference) due to the supply of data concerned with the digital persona. This makes the charity and charity fund-raising sector results more impressive. The worst five performing sectors on the other hand provided no data related to personal persona. Based on correspondence and a consideration of local government operations, it could be expected that digital personas are not calculated or used. Whilst it could be considered more likely that Google and UKBA do use analyses of personal data, Google for targeted marketing and UKBA for border control and security, the three poorest performers also provide below average information about digital footprints and third party digital footprints. One may conjecture that the lowest performing organisations are either unwilling to provide information, or do not have the competence to do so. This will be addressed in the third and final phase of this research.

	No	Overall % of Elements Provided	Digital Footprints	3rd Party Digital Footprints	Computed Digital Persona	Imported Digital Persona	Exported Digital Persona	External Data for Digital Persona	Data from Other Sources	Data to Other Sources
All Sectors	*58*	*9*	*7*	*19*	*5*	*2*	*7*	*12*	*12*	*10*
	58	*24*	*50*	*38*	*14*	*14*	*2*	*10*	*29*	*38*
Central Government	3	4	0	0	0	0	0	0	0	33
		8	33	33	0	0	0	0	0	0
Charity	6	0	0	0	0	0	0	0	0	0
		46	83	83	0	0	0	0	100	100
Charity Fund	2	0	0	0	0	0	0	0	0	0
		31	50	50	0	0	0	0	50	100
Credit Reference	5	33	0	40	20	0	40	60	60	40
		23	0	40	20	40	20	20	20	20
Finance	6	17	17	17	17	17	17	17	17	17
		19	17	50	0	17	0	17	17	33
Internet	5	0	0	0	0	0	0	0	0	0
		13	40	0	0	0	0	0	20	40
IOS App	8	0	0	0	0	0	0	0	0	0
		19	63	13	13	13	0	0	13	38
Local Government	2	13	0	100	0	0	0	0	0	0
		6	50	0	0	0	0	0	0	0
Marketing	2	19	0	0	50	0	0	50	50	0
		25	50	0	50	0	0	50	50	0
Online & High Street	6	10	0	0	0	0	0	33	33	17
		42	100	83	50	33	0	17	17	33
Online Shopping	7	4	14	14	0	0	0	0	0	0
		34	71	43	29	14	0	29	43	43
Social Media	2	13	50	50	0	0	0	0	0	0
		13	50	50	0	0	0	0	0	0
Utilities	4	22	25	100	0	0	25	0	0	25
		9	0	0	0	25	0	0	25	25

Table 4.8 Percentage data transparency score by sector of organisation after 1st communication for responding organisations

Note: for each sector the first row indicates the % of organisations that provided partial data and the second row indicates the % of organisations providing full data.

The following Table 4.9 shows the position at the end of the data collection process after follow up letters had been sent to 29 organisations, from which 20 responses were received and processed.

	No	Overall % of Elements Provided	Digital Footprints	3rd Party Digital Footprints	Computed Digital Persona	Imported Digital Persona	Exported Digital Persona	External Data for Digital Persona	Data from Other Sources	Data to Other Sources
All Sectors	*58*	*10*	*7*	*14*	*10*	*3*	*7*	*14*	*14*	*14*
	58	*28*	*53*	*48*	*17*	*16*	*2*	*16*	*34*	*41*
Central Government	3	8	0	33	0	0	0	0	0	33
		8	33	33	0	0	0	0	0	0
Charity	6	0	0	0	0	0	0	0	0	0
		48	100	83	0	0	0	0	100	100
Charity Fund	2	0	0	0	0	0	0	0	0	0
		31	50	50	0	0	0	0	50	100
Credit Reference	5	33	0	20	40	0	40	60	60	40
		25	0	60	20	40	20	20	20	20
Finance	6	21	0	0	33	17	17	33	33	33
		25	33	67	17	17	0	17	17	33
Internet	5	0	0	0	0	0	0	0	0	0
		20	40	20	0	0	0	20	40	40
IOS App	8	0	0	0	0	0	0	0	0	0
		19	63	13	13	13	0	0	13	38
Local Government	2	6	0	50	0	0	0	0	0	0
		19	50	50	0	0	0	0	0	50
Marketing	2	25	0	0	50	0	0	50	50	50
		25	50	0	50	0	0	50	50	0
Online & High Street	6	10	0	0	0	0	0	33	33	17
		44	100	100	50	33	0	17	17	33
Online Shopping	7	5	14	14	14	0	0	0	0	0
		36	71	43	29	29	0	29	43	43
Social Media	2	13	50	50	0	0	0	0	0	0
		13	50	50	0	0	0	0	0	0
Utilities	4	25	50	75	0	25	25	0	0	25
		31	0	25	25	25	0	50	75	50

Table 4.9 Percentage data transparency score by sector of organisation after final communication for responding organisations.

Note: for each sector the first row indicates the % of organisations that provided partial data and the second row indicates the % of organisations providing full data.

The following Figure 4.11 illustrates the impact of the follow up communications. This is shown again by sector and so the sample size of the 29 communications, across the 14 sectors is too small to be very helpful. Nevertheless, it is interesting to observe that the Utilities, Finance, Local Government and Internet sectors all increased their scores significantly more than other sectors.

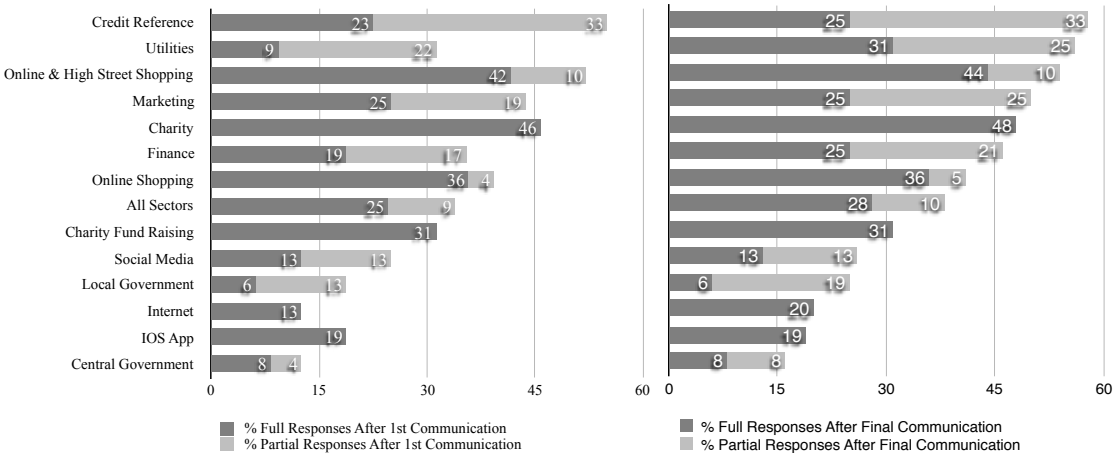


Figure 4.11 Responses with respect to data provision by organisational sector (responding organisations only)

Table 4.10 summarises the final responses received from organisations analysed by sector and the number of elements provided (i.e. digital footprint, 3rd party digital footprint etc.). It is based on the full sample of 82 organisations and therefore reflects the impact of those that failed to respond.

Charities, shopping, and credit reference agencies outperform other sectors. The data processing sector provided the least information but this is explained by the organisations within this sector being data processors rather than data owners, and therefore exempt from replying to subject access requests. Internet-based organisations and IOS app developers are the next worse performing, as a result of not responding.

Sector	No of Organisations	Total Elements	Elements / Organisation
Charity	6	19	3.2
Online & High Street Shopping	7	21	3.0
Credit Reference	4	11	2.8
Online Shopping	7	18	2.6
Utilities	5	12	2.4
Marketing	4	8	2.0
Social Media	2	4	2.0
Finance	12	22	1.8
Charity Fund Raising	3	5	1.7
Central Government	4	6	1.5
Local Government	2	3	1.5
Internet	6	7	1.2
IOS App	18	7	0.4
Data Processing	2	0	0.0
Total	82	143	Mean 1.7

Table 4.10 Mean elements received, analysed by sector across all organisations

4.4.5 Findings by Location of Organisation and Data

Of the 82 organisations contacted, 67 were located in the UK, 4 in the EU outside the UK, and 11 outside the EU. Of the organisations that responded to requests 49 were in the UK, 3 in the EU outside the UK, and 6 were outside the EU (Table 4.11). The sample sizes for organisations outside the UK were therefore low and so the analysis may not be representative of the wider population. Also scores for organisations providing information are dependent on the organisation having that class of information. For example, Cult Pens

who do not own or import data persona are scored the same as Facebook and Google who do not admit to owning or importing data persona. Despite these considerations, it is still of interest to note that for all but one classification the UK performs better than countries outside the EU, and the EU performs better than non-EU countries. The case of digital footprints where the EU not UK, outperforms the UK, is a result of a low sample of 3, where all three organisations are internet based, and have the capability to quickly extract digital footprint data.

	No	**Overall % of Elements Provided**	**Digital Footprints**	**3rd Party Digital Footprints**	**Computed Digital Persona**	**Imported Digital Persona**	**Exported Digital Persona**	**External Data for Digital Persona**	**Data from Other Sources**	**Data to Other Sources**
All Locations	*58*	*10*	*7*	*14*	*10*	*3*	*7*	*14*	*14*	*14*
		28	*53*	*48*	*17*	*16*	*2*	*16*	*34*	*38*
UK	49	11	6	14	10	4	8	16	16	16
		30	55	51	20	18	2	16	37	41
EU not UK	3	13	33	33	33	0	0	0	0	0
		21	67	33	0	0	0	0	33	33
Not EU	6	0	0	0	0	0	0	0	0	0
		15	33	33	0	0	0	17	17	17

Table 4.11 Percentage data transparency score by location of organisation after final communication for responding organisations.

Note: for each sector the first row indicates the % of organisations that provided partial data and the second row indicates the % of organisations providing full data.

All 82 organisations were also asked where they held their data. Of the 58 responding organisations, 23 provided answers to this question despite it being outside the scope of the Data Protection Act 1998. The results are shown in Table 4.12. There may be many reasons for this low response, for instance: organisations not responding may be more secretive; less able to provide this specific data; those answering the requests may not have known the answers as it is outside their normal terms of reference; or as this information is outside the scope of the Data Protection Act 1998 they may have felt free not to answer. Of those who answered this question, 40% of UK based organisations held data outside the EU.

	Location of Data			
Location of Organisation	**UK**	**EU not UK**	**Not EU**	**Unknown**
UK	11	1	8	46
In EU outside UK	0	0	0	4
Outside EU	0	0	3	9
Unknown	0	0	0	0

Table 4.12 Location of organisation compared with location of data

Using the same measures as Figure 4.11, the sector analysis, Figure 4.12 shows percentage response rates for the provision of data for the 58 organisations that answered the subject access requests, analysed by location of their data. Perhaps unsurprisingly organisations who disclose data locations score more highly overall than those who choose not to disclose them (shown as Unknown in Figure 4.12). Discounting the single organisation who held data in the EU outside the UK (Vodafone) there is a clear deterioration in scores with the best performing being in the UK, then outside the EU, and finally those who chose not to divulge where data is held.

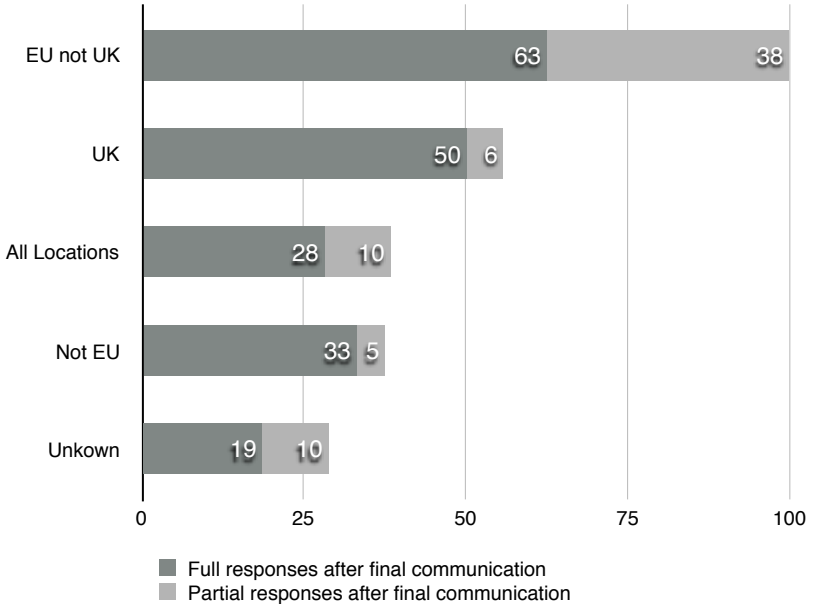


Figure 4.12 Response with respect to data provision by location of data for the 58 responding organisations

4.5 Discussion

The first phase of this research, described in Chapter 3, examined the terminology used to label personal data and proposed a standard set of terms, and a categorisation model, which was validated against a purposeful sample of literature. This second phase took that model of the digitally extended self, and matched it to personal data obtained from a number of private and public companies, NGOs, and governmental bodies. That data was also analysed in order that the behaviour of organisations to requests for personal information may be better understood. In addition, the results shed light on the issues an individual may encounter whilst attempting to better understand their digitally extended self. This section will first argue that the data collected, from subject access requests, validates the categorisation model presented in Chapter 4. It then examines the possibility of an individual retrieving the data that comprises their digitally extended self, before finally assessing variations in the collected data.

4.5.1 Validation of the Model Against Data[3]

To revisit the first research question:

> *RQ1: What are the components of the digitally extended self and how do they relate to one another?*

In Chapter 3 a model was developed based upon terms and usage from personal data and related literature, that categorises the data that is descriptive of an individual. By bringing like things together we enable meaning and understanding of the domain (Svenonius, 2009). In this case, the model of an individual's data is viewed from the aspect of the individual. It is *their* data that is laid down in digital footprints, others describe *them* in third party digital footprints, it is *they* who are described and stereotyped in digital persona, and finally it is *their* data which is moved around. This viewpoint was taken in order to balance the asymmetry in power between the relatively powerless person and the commanding organisation (Rouvroy and Poullet, 2009, Hildebrandt, 2009), which defines 'take it or leave it' terms of usage.

In the first phase of this research, described in the previous chapter, the model was validated by matching highly cited literature, which used terms describing personal data, to the model's categories. This the second phase, seeks to validate the model against real

[3] In part originally published in the Journal of Information Science (Parkinson et al., 2017)

data. Whilst the data represents a single individual, it originates from 58 organisations, and as such includes many data types, all of which map into the categories defined within the model. This section argues that the model is valid, and to illustrate this three case studies are described below.

The case studies are based on the interactions of the author with a UK based bank, an international charity, and a credit reference company. To create these case studies, data were gathered through subject access requests made to the organisation in September 2013, March 2014, and October 2014.

Figures 4.13, 14, and 15 below, show diagrams of what these data were in the case study examples and where they fit within the model. In this instance, centric diagrams have been used, to illustrate *digital footprints* at the heart of the data, which describes an individual. It is then incrementally extended through the concept of multiple artefacts forming a *digital mosaic*, identified by the inner circle. Next analyses are formed using data from digital footprints and external sources resulting in *digital personas*. The whole within the outer circle is named the *digital extended self*.

4.5.1.1 Case Study 1, a UK based bank

The case study shown in Figure 4.13 takes data provided by a UK based bank and maps it against the centric visualisation of the digitally extended self thus revealing the significance of the parts of the digitally extended self *not* under the direct control of the user. In this case, extensive notes and internal records of non-digital interactions made by third party individuals (e.g. account enquiries from branch or telephony agents and bank account customer notes); three separate personas generated for purposes of underwriting, credit scoring, and overdraft scoring; data that are independent of the individual and provided by third party credit reference, fraud and taxation organisations; and finally second level demographic data which describes the individual by inference to the location of their home.

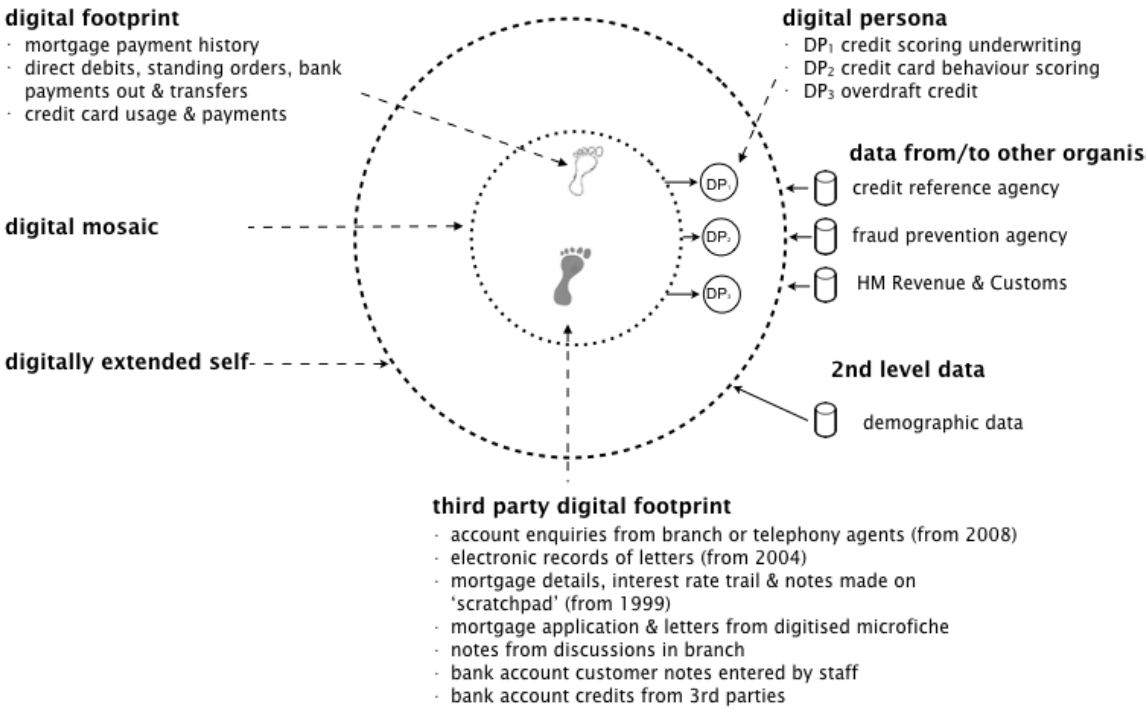


Figure 4.13 The centric diagram showing data (as instances of the model) from case study 1, a UK based bank.

4.5.1.2 Case Study 2, an international charity

The second case study, Figure 4.14, uses data provided by an international campaigning charity and illustrates that even though a minimal level of data was held by this organization, (name, email address, and date of donation), a digital persona, received from an external company, was kept, showing propensities to open and to click on emails from this charity. In addition, the diagram illustrates that data is sent to the charity's offices in three other countries, two of which are approved by the EU for the flow of personal data and one that is not, raising possible privacy concerns.

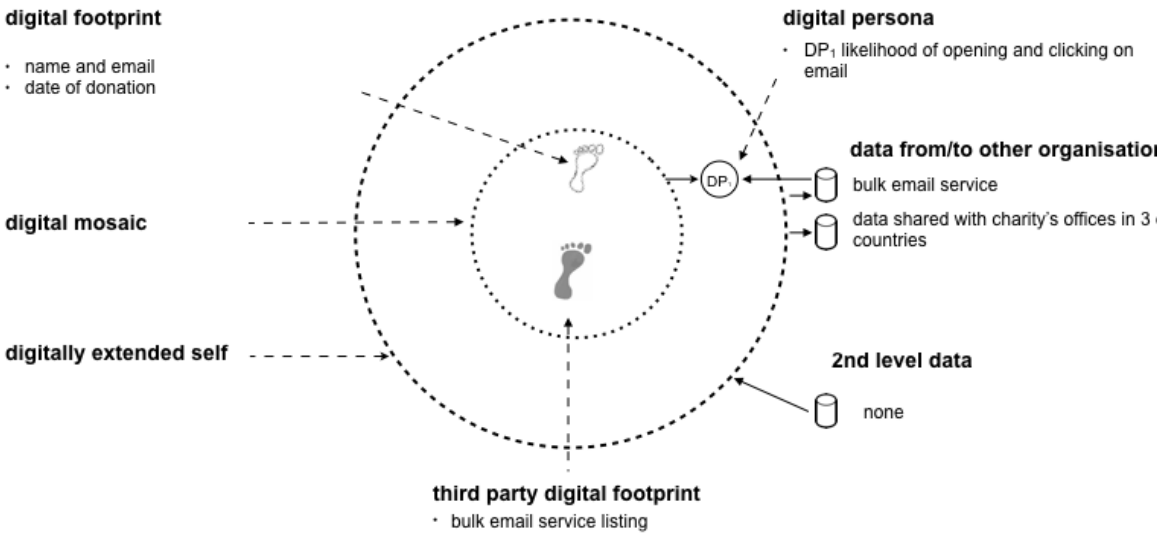


Figure 4.14 The centric diagram showing data (as instances of the model) from case study 2, an international charity.

4.5.1.3 Case Study 3, a credit reference company

The final case study, Figure 4.15, is a credit referencing organization. This private limited company had no direct contact with the data subject but collected third party digital footprints in the form of summary data from financial and mobile phone companies, which it combined with post code, electoral role, fraud, bankruptcy, and court judgment data. This information was used to calculate an over indebted assessment which, together with other data, were provided to financial institutions and mobile phone companies. This collection and dispersal of data descriptive of the author illustrates how sharing of personal data can be used to create a profile which is in turn distributed to other actors, unknown to the subject individual. This case also illustrates how the absence, rather than the presence, of a third party digital footprint can itself be descriptive of an individual. The absence of fraud data, bankruptcy, or county court judgements supports a mosaic (which fortunately showed the author in a positive light).

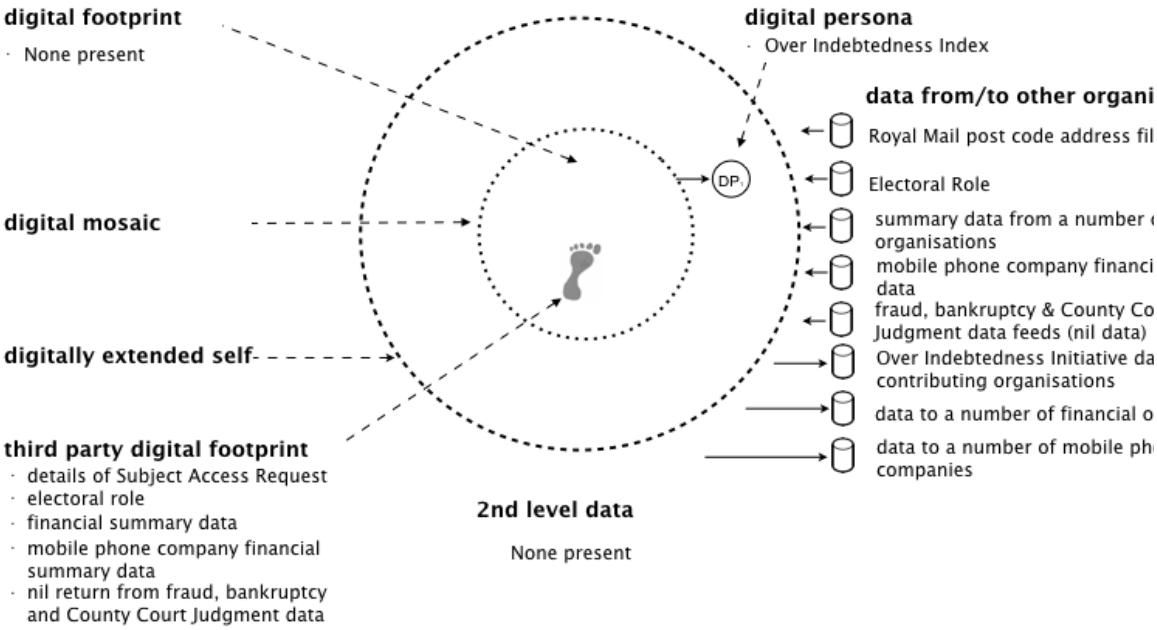


Figure 4.15 The centric diagram showing data (as instances of the model) from case study 3, a credit reference company.

4.5.1.4 General Observations

The model's categories name data that describe an individual, and this data collection exercise confirms that data elements fit into one, and only one, class, that all data fitted the model, and no data was omitted. In addition, the model proved very useful when examining responses from individuals. By comparing the data received against the model, it was possible to see if a category of data had not been provided. Furthermore, by considering the response data as part of a category of like things, it helped to prompt questions about similar items that had been omitted. This is acknowledged as a subjective view and is without measurement. It is also from the individual's perspective as the, perhaps biased, user of the model. However, it may also be useful to look at the model from the organisation's point of view. The subject access request was formatted to ask for data as categorised by the model plus information about data movement. The most complete and also simple response was from Not On The High Street, which placed data into the model's categories and then provided information on data in, and data out. It may be that the researcher's hopes and expectations were met more fully by the use of the model categories in this way, on the other hand the data provided was complete, well-ordered and understandable.

The process used for data collection raised an issue which is pertinent to the view of data that is adopted, in this instance from the individual's perspective. Organisations were asked

about data movement in order that information regarding the propagation of the data could be obtained, and so that the snowball sample could be generated. There are many attributes of the data descriptive of an individual and their importance will depend upon one's focus. For example, a coder would want to know whether the data is numeric or alphanumeric, and an accountant its value. From an individual's point of view, information about movement of data was important, although it was not represented within the model. As Glushko et al. (2013) state, it is the difference between organising things and organising information about things. The analogy of a library illustrates this point. Books are classified into categories within the library, as personal data can be classified into the categories of the model. In addition, the library needs to know location and movement information about the books. The individual also needs to know location and movement information about their personal data.

In summary, the model developed in this work categorises data descriptive of the individual effectively. The hierarchical model defined in Chapter 3 shows a hierarchy of construction of the digitally extended self, whereas the centric visualisation introduced in this chapter illustrates that as one moves away from the centre of the model the individual has less knowledge of and control over the data. Both are representations of the proposed categorisation of personal data, However, information describing the location, or movement of data categories, between organisations, is not an attribute of an individual and so is not explicitly represented by the categorisation model. It is however not inconsistent with the model, and is depicted in the deconstructed centric visualisation (figure 4.4) which illustrates organisations curating subsets of the digitally extended self and moving data between each other. The researcher hypothesised that information contained in the model and diagram may be useful to both the organisation in responding to subject access requests, to legislation in defining personal data, and to people when explaining personal data. This hypothesis is tested in phase three of this research.

4.5.2 Can an Individual Retrieve Their Digitally Extended Self?

The second area to be investigated was the ease with which an individual could obtain their data, the question was framed thus:

RQ2: How feasible is it for an individual to obtain the information, held by organisations, which is descriptive of them?

There are a number of aspects to this issue, which this research has addressed, the time taken, costs involved, and the completeness of the information received.

4.5.2.1 External Issues

When considering the fundamental issue of completeness, there are five factors to consider.

First, when an organisation sends data how does the recipient know it is complete. It was obvious when Oxford District Council provided its data that the electoral roll information was missing. When asked about this in a follow up letter the data protection officer apologised and provided it. This illustrates one issue, that the organisation may make a mistake and omit data. In another situation neither Tesco nor Dunnhumby provided Tesco Clubcard information, Dunnhumby are the customer science company that processes Clubcard data. Follow up letters were sent to both organisations requesting this information but nothing was provided. This could be an error, a legal misunderstanding, or purposeful omission. On Dunnhumby's part, they argued the case that they were data processors only and so did not have to provide data, whilst Tesco did not respond to the follow up letter. Another situation is one of ignorance. The Parcel Force (Royal Mail) replied to the subject access request thus 'Please could you give us an indication of what parts of our business may hold information about you as, due to the size of the business, it is not possible to complete an open-ended search of all records' (Appendix V). Indicating that they did not know where data belonging to an individual may reside, but expecting people to be able to know the answer. A critical realist would argue that individuals can only see the empirical and not the events and structures that lie behind (Mingers, 2004). It is not realistic for organisations to expect individuals to know where data resides.

The second issue is that of location of data, how can an individual know who holds their data. Appendix B contains a large list of organisations that the author has dealt with, and which hold digital records descriptive of him. It cannot be expected that every person holds this sort of index. Given this starting point it is still necessary to discover to which other organisations data moves, and for completeness where it may have come from. Boden receives data from Experian and moves data to them in order to receive marketing

information and to send out emails. Experian confirmed this with Boden being a member of their Club Canvasse. Amnesty move information to a number of organisations one of which was Rogavi. This company is now in receivership, no replies were received from the organisation or administrators and therefore no information regarding the fate of one part of the authors personal data can be obtained. Overall 59% of organisations provided some details of personal data movement, leaving 41% who provided none. As with the provision of data discussed above, a data protection officer may not know where data is sent, they may know but make a mistake and not provide the location, or they may choose not to tell you. John Lewis choose not to provide details of organisations with whom they may have exchanged data, but give a list of 12 categories of organisations one of which is 'traders in personal data'. In this case, we may not know where our data has been sent but can be fairly sure it is being exploited and our transactions mined. John Lewis provided 817 pages of data showing transactions dating back to 2004. Of course, John Lewis do not have to tell people where their data has been sent as this requirement is not covered by any law.

The third issue is that the digitally extended self is constantly evolving. Further visits to Waitrose have been made and the researcher's digital footprints within the John Lewis systems have increased. The author has used new organisations and organisations in turn will have changed their procedures and will have passed the information on to other new destinations. By requesting data, we receive a snapshot of our digitally extended self like some photograph in an album, frozen in time, an aid memoire to what we were, but not a reflection of what we are. From this perspective, we can never know our own data.

The fourth issue is one of comprehension. Once the individual obtains their data, if that was possible, will they be able to understand it? Apple kindly send a glossary of terms with their data one example is 'Guid_Name' which is defined as 'UDID for iOS Devices, MAC address for Macs, and an iTunes created Device ID for Windows/Other'. Even when terms are defined many will fail to understand the data that they have been given. Generally, the information is well presented, Cult Pens chose to provide the information by way of copies of all invoices, which are designed for public consumption and understandable, and each question was answered in clear English.

The fifth, and final, issue for completeness of access to data is the serious issue of organisations failing to reply. In the UK the citizen has a right of access to their data under the Data Protection Act 1998. Similar legislation provides for access within the European Union, but this is not the case when dealing with organisations resident outside the EU. Even within the UK one company from the original purposive sample failed to reply to the recorded delivery subject access request. Cooperative Energy were written to twice at their

address shown on the data protection register, the same address as is shown on their website, but no replies were received. Of the 82 organisations contacted, 58 finally provided data, four did not as they claimed to be data processors, one was a research organisation, one did not because they held no data as the wrong area of the NHS had been contacted, and 18 failed to answer or supply information - 17% of all those contacted. With a 17% failure rate for the companies contacted it is clear that obtaining a comprehensive view of one's own digitally extended self will be challenging. However, it is always possible to chase these organisations more times or in differing ways, and this may enable a fuller view to be obtained. The one organisation in the sample to refuse to supply their data, as opposed to data they claimed to just process, when contacted was the Office for National Statistics. As a research organisation, they do not have to provide data, although they are not forbidden to do so. In their reply, they stated that:

> 'Section 33 of the Data Protection Act does not forbid ONS from answering a request for personal data but it is ONS policy to resist any attempt to force it to disclose personal data for non-statistical purposes even from data subjects.'
>
> Extract from ONS letter 28th March 2014 (Appendix T).

As a major source of information, held on behalf of the people of the UK, it appears unfortunate that it is not available to citizens.

Whilst some organisations are exempt from the Data Protection Act 1998 so is some data. Digital persona is one of the model's categories of data that is effectively exempt, as is how data is moved around between organisations. These areas are not properly supported in current legislation, of digital personas John Lewis stated:

> 'Data Analysis is not personal data; therefore, information around this would not be provided in a Data Subject Access Request' (Appendix Z).

HMRC took a similar stance in respect to digital personas and also movement of data:

> 'You should be aware that the Data Protection Act provides for a number of exemptions to the disclosure of information. HMRC like all registered Data Controllers under the Act, are entitled to apply these exemptions where permitted' (Appendix AA).

The result is that an individual cannot expect to obtain such information, but without it they cannot know where data is held, or how they are considered by organisations. On the other hand, organisations may see provision of this information as a cost without benefit and be reluctant to provide more than they legally have to. Some may find it surprising, however, that governmental institutions staffed by civil servants appear less transparent and helpful than those in the commercial sectors. Nevertheless, current legislation was

enacted 17 years ago, the year that Google was founded, and 6 years before Facebook was formed. It should be of little surprise that it is lacking in some areas.

Since this research was undertaken the General Data Protection Regulation (GDPR) has been passed by the European Parliament (European Union, 2016) and after a two-year preparation period is due to be translated into UK Law from May 25th, as the Data Protection Act 2018. The new act has some exceptions to the GDPR (for example, exemptions for journalists, and researchers who handle personal data), but there is little difference in the data falling under this legislation from the Data Protection Act 1998. Areas of change to be noted are the extension to cover pseudonymised personal data, and identifiers which may be associated with an individual, for instance IP addresses and internet cookies. This enables individuals access to a slightly wider range of data, and also, the charges for subject access requests have been removed, lessening the barrier to discovering one's digitally extended self.

However, it is understood that the data used for digital personas, and details of the algorithms used to create them, will still be outside the Act, although Article 22 creates a right not to be subject to an automated decision if it has a significant effect upon a person (given a number of exemptions). Also, whilst the transfer of data is still restricted to trusted countries, (except for important reasons of public interest), there is still no right to know which organisations, or locations, data has been transferred to, or obtained from. However, the right to erasure of data, which may be requested in a number of circumstances, (including the individuals withdrawal of consent), means that organisations must inform other bodies with which the data was shared, about the erasure of the data, and also let the individual know about these third parties (unless it proves impossible or involved disproportionate effort). The collection of data will now require explicit consent, although given the asymmetry of power between the organisation and the individual, it must be expected that it will not be withheld. However, the transparency associated with his process may have beneficial effects in moderating the behaviour of some organisations. Finally, the ICO will have more powers to levy higher fines on organisations. However, given the ICOs apparently poor track record to date (Interview 8 00:37:43.03) this may not make very much difference. On the other hand, an individual may now appeal to the Tribunal (Information Rights) to obtain an order requiring the ICO to get on with its work and this may have some effect.

In summary, the act would appear to allow access to slightly more data, and to have data erased in some circumstances, which may lead to information being available about data sharing. However, details of digital personas will still be outside the reach of the

individual, and given the effects that these do have on our lives this is disappointing, as is the lack of a right to know with whom one's data has been shared.

Completeness of data is, and may remain, an issue for anyone wishing to investigate their own digitally extended self, not all data is provided, it is difficult to discover where data is moved to, the digitally extended self is forever changing, once obtained the data is not necessarily understandable, some companies do not answer requests, and finally there are significant areas not currently covered by legislation and therefore not included in responses. If this were not enough by way of obstacle to discovering one's own digitally extended self, there are other issues to consider.

4.5.2.2 Personal Constraints

The second area of concern after considering if a person can obtain a full or at least largely complete view of their digital extended self is that of the time involved. From the activity log that fed into the Cost and Time Spreadsheet shown in Appendix G, the mean time taken to contact and obtain data from an organisation was 1.43 hours. This does not take into account time used if the person chose to scan and OCR the data received, but does include a quick browse of the data to match it to the categories within the model. Even so, this equates to 229 working days, of 7 hours for a very roughly extrapolated 633 organisations, which is approximately one working year. Subject access requests may not be resubmitted until a 12-month period has lapsed, and so one could imagine someone continuously enquiring and tracing the extent of their ever increasing digitally extended self in a fruitless quest for completion.

The third aspect to be considered is that of cost. The mean cost incurred for this research per organisation was £12.78. Strategies to reduce this cost have been discussed but on the other hand, if organisations that failed to provide data are pursued greater costs would be incurred. If we assume the rough estimate of 633 organisations then the total cost would be £8,090. In 2013 (the year when the initial subject access requests were submitted) the 50th percentile income in the UK after tax was £18,700 (ONS) the cost of exploring their digitally extended self would therefore have been 43% of income after tax, and for the 75th percentile who earned £28,200, 29%. These are high percentages of income and it could be considered that the costs make this an activity only for the wealthy.

Finally, there is an issue of what happens when data is missing. Normally this is a problem because some element of information has been omitted from a reply, but absence of data is sometimes as important as presence (e.g. CIFAS or a no-fly list). In the case of CIFAS, the Credit Industry Fraud Avoidance Service, absence of data may be interpreted as the person in question having a digital persona as someone who has not defrauded and therefore will

probably not commit a fraud. Similarly, absence from a no-fly list may be taken to suggest that you do not have terrorist tendencies.

Sometimes absence of data is data in itself, although of a weaker strength as it is implied rather than stated. As data has not been sent to the organisations managing these data stores, and people may not know of their existence, it is very challenging to discover this type of 'persona through omission'.

4.5.2.3 Summary

The second research question asked how feasible is it for an individual to obtain the information, held by organisations, which is descriptive of them. The answer would depend upon whether the exercise is to obtain information from one organisation or to discover the extent of their digitally extended self. If we centre on a single organisation then there may well be issues of completeness of data to consider and this is examined further in the following section.

If an individual is wondering about the feasibility of discovering their digitally extended self then the answer from this research would be that it is not possible.

As for the reasons, they are relatively straightforward. The data obtained from known organisations is normally incomplete and occasionally incomprehensible, people are not told where data has moved to, which means that there are organisations with data that people do not know about. In addition, it takes too long and costs too much for most people to even attempt to address the problem, and finally a person's data is always changing and so must be considered unknowable.

4.5.3 Variations in Data Provided

The final research question was:

RQ3: What is the quality of the personal data returned by organisations when it is requested by individuals?

The data sample for this work consists of 82 organisations of whom 58 responded with some data. Of the 24 who did not provide data, four claimed to be data processors, one was a research organisation, one was a non-data holding area of the NHS, and 18 failed to answer requests for information. At a high level, it is clear that organisations as a whole do not perform well with 17% choosing not to respond to a data request.

The next sections will focus on four issues drawing on data from across the phase two findings presented above. They are, the lack of consistency in data provided and the poor performance of some organisations compared with others, whether data provision is seen

as a cost without benefit, whether legislation should be extended to cover more data, and whether location of data should be provided.

4.5.3.1 Lack of Consistency

The first issue is the lack of consistency in the way that organisations respond to subject access requests. At a high level, it can be observed (Table 4.13) that some organisations choose not to answer requests for data. Within the UK and EU, 19% or organisations failed to answer requests for information whilst 50% of those outside the EU failed to respond. The failure rate is higher than predicted and reflects the poor performance of the IOS App development community, as well as problems associated with getting information from organisations outside the EU.

Location		**No**	**%**
UK & EU	No response	13	19%
	Data Processor etc.	5	7%
	Data Provided	52	74%
Outside EU	No response	6	50%
	Data Processor etc.	0	0%
	Data Provided	6	50%

Table 4.13 Failure to answer requests for personal information by location (number of organisations)

The next level of granularity is organisational category. Again, there is an inconsistency of responses. Of the 82 organisations written to, NGOs provided on average information on 48% of possible data elements compared with private companies who averaged 34%, central government 15% and local government 25%. The low score for private companies was again affected by IOS developers, often outside the UK who did not respond to requests for data. Of the 58 organisations that responded to requests for personal information, NGOs performed better after the first communication with public companies slightly outperforming them after follow up data was received. The reason for this may be that public companies were able to provide data relating to digital personas that NGOs do not possess. Central and local government were the worst performing categories. The variations in responses are large, and deserve further research. Perhaps the most striking

differences lay between the NGO category and local and central government. Organisations in these categories all responded to the subject access requests and were located in the UK. The organisations all in theory serve the public interest, rather than that of shareholders or owners, but the quality of their responses was dramatically different. When organisations are grouped by sector a similar pattern emerges, although conclusions by sector must be treated with caution due to the low numbers of organisations in each sector. Nevertheless, of the 82 organisations written to, shopping, charity and credit reference all provided more than 2.8 elements per organisation on average, whilst local and central government provided less than 1.5 with IOS developers providing 0.4. Of the 58 organisations that responded to requests for personal information the online and high street shopping, and charity sectors were the top performers. Central and local government together with internet, social media and IOS app developers clustered as the worst performing. With respect to civil authorities it is interesting to note that individuals believe that the Data Protection Act 1998 increases transparency and accountability (Fanucci, 2008) but on the other hand an audit of local authorities (Information Commissioner's Office, 2014, p. 2) stated 'This clearly shows there is room for improvement in all the organisations we visited'.

The final level of granularity is that of data category, and at this level the constituents of the digital mosaic were the most often provided with 57% of all organisations complying to the request. On the other hand, information regarding data movement was supplied in less than 50% of the cases, and digital personas in less than 24%. Within these figures there is a great deal of variation as can be observed in the heat maps in Figures 4.4, to 4.9 above.

In summary, there was great inconsistency in the replies given by organisations to the subject access requests submitted during this research. Generally, people are uncomfortable with levels of inconsistency (Merritt et al., 2010) however, when people believe that the decision-making process, which leads to inconsistency is sound, they have less regret about the outcome (Pieters and Zeelenberg, 2005). Thus, it may be that even when presented with evidence that civil authorities perform poorly when replying to subject access requests they themselves are comfortable with that situation, as the rules have been followed. Chapter 5 will present the final phase of this research which will seek to investigate the reasons for observed inconsistencies in responses and in particular, the poor performance within the central and local government categories and sectors.

4.5.3.2 A Cost Without Benefit?

The second issue relates to how organisations view their obligations under the Data Protection Act 1998. It is possible that obligations under the act are seen as a cost without benefit, this will be examined in Chapter 5. Nevertheless, the act allows people the right to correct inaccurate data, as for example, it may prevent access to services. However, of the 58 organisations that responded to the subject access requests, none provided a method to correct data or mentioned data correction in their correspondence. This is despite inaccurate data being an issue for organisations. During the researcher's time at Lloyds TSB a team of at least 30 people were employed to correct data, sitting like Canute trying to reverse the tide of inaccuracies. Inaccurate data lowers customer satisfaction, increases costs, and lowers employee satisfaction (Redman, 1998). So, why don't organisations take advantage of the possibility of customer feedback to correct data? This may be due to inertia, organisational difficulties or cost.

4.5.3.3 Variations in Responses by Data Category

The third area is related to that of inconsistency of replies and relates to less data being provided for categories of information not covered by the Data Protection Act 1998. The following Table 4.14 shows the percentage of organisations written to that replied and provided model category or movement data, either in full or in part. The data that is more often supplied is that fully covered by the act with the elements of the digital mosaic provided in more than 60% of the cases. The data not covered by the act, data relating to digital persona is provided in less than 30% of the cases whilst general information about the movement of data in about 50%. It may be argued that there is a much lower number of organisations who hold or calculate digital persona, and that the provision of general movement of data destinations is evidence that organisations are willing to provide information without the force of the law. On the other hand, those elements covered by the act are provided 20% more often than general movement information.

Data Type	**Some Data Provided**
3rd Party Digital Footprints	62%
Digital Footprints	60%
Data to Other Sources	52%
Data from Other Sources	48%

External Data Imported for Use in Digital Persona	29%
Computed Digital Persona	28%
Imported Digital Person	19%
Exported Digital Persona	9%

Table 4.14 The percentage of times data is provided by data category (after final position)

The arguments for data protection are based on data quality, transparency and more recently informational self-determination (De Hert and Gutwirth, 2006). These arguments apply just as strongly to digital persona and the movement attributes of data as they do to the central elements of individual's digital mosaics. If organisations see the provision of any information as a cost, then it must be expected that they would lobby against any changes to legislation that would increase their administrative burden. Nevertheless, the argument for the protection of the individual is strong, especially in Germany where West German's legislation promoting informational self-determination helped in creating free access to Stasi files (Gieseke, 2014).

4.5.3.4 Variations by Location

The final issue is also related to legislation but focuses on location rather than digital personas and movement information. Organisations that are resident in the EU and process their data within Europe provided fuller information than other organisations. In addition, organisations that were willing to state where data is held provided much more information that those that are unwilling to state the location of their data. This may be because if an organisation is unwilling to provide basic information then they will be much less likely to inform people where their data is held and processed. Alternatively, it may be because data protection officers do not know where data is held or processed. Under European law the transfer of data outside Europe is regulated, at the time of writing, under Directive 95/46, and so data may only be sent to approved countries. There is some issue regarding onward transfer especially to territories that have a lower legal standing and may take a role as a data haven or as an onward transfer centre (Blume, 2014). Indeed, it was argued that even safe harbour agreements did not provide an appropriate level of security. On the 30th November 2013 the EU called for action in 6 areas to restore trust in the safe harbour agreement with the USA following the Snowden revelations (von Solmes and van Heerden, 2015). On the 12th July 2016, the EU-US Privacy Shield (European Commission,

2016) was adopted as its replacement, and in General Data Protection Legislation will be enforced from 25th May 2018 (Council of the European Union, 2016). Whether or not there is a belief in the integrity of EU agreements with other states for processing personal information, transparency should bring a greater level of trust and is a topic that would benefit from greater scrutiny.

4.5.3.5 Summary

Research question 2.2 asked how organisations respond to requests from individuals, whether some data is more readily provided than others, whether some types of organisation respond more fully than others and whether any general issues had been observed. From the evidence collected it is clear that there is a lack of consistency in the level of provision across organisational categories and sectors, with central and local government responding to requests but performing poorly and IOS app developers failing to answer requests for information. There is also a lack of consistency within data categories with elements of the digital mosaic more commonly provided than digital personas, or information regarding the movement of data. This may be for a number of reasons, whether the data exists, whether it is known to data protection officers, or because it is covered by the Data Protection Act 1998. The data categories not covered by the Act are provided in response to subject access requests much less frequently than others. This may suggest that responsibilities under the Act are seen as a cost without benefit, especially as the opportunity to have data corrected is not taken up. Finally, questions regarding the location of data are infrequently answered, but organisations that state that their data is held in Europe provide fuller replies than others.

4.6 Conclusions

This research, so far, has aimed to address three questions:

RQ1: What are the components of the digitally extended self and how do they relate to one another?

RQ2: How feasible is it for an individual to obtain the information, held by organisations, which is descriptive of them?

RQ3: What is the quality of the personal data returned by organisations when it is requested by individuals?

The contribution of the first two phases of this research, Chapters 3 and 4, is in the construction of a validated model that categorises data elements descriptive of an individual. An extension of this covers attributes of the data categories that relate to their

movement. This was achieved through an analysis of that literature which used terms descriptive of an individual's data. The terminology was synthesised and like elements grouped into categories which form the basis of the model, as described in Chapter 3. The centric diagram of the digitally extended self, presented in section 4.5.1, is focused around the individual as a consequence of the overall context of personal data. The model shows that as data becomes more distant from the individual (moving from footprints to mosaics to personas) the questions of ownership, access, and control of that data become less clear as it increasingly incorporates data from third parties (both individuals and organisations, in the form of their computer systems). Finally, the model was validated against data from a range of organisations, indicating that it may be robust and perhaps helpful in the understanding and discovery of data, the framing of legislation, and within organisations for the structuring of responses to requests for personal data.

The analysis of data provided by a range of organisations highlighted a lack of consistency in the levels of data received. It was noticeable that charitable organisations provided greater width and depth of information than central and local government. There were also other significant variations in the data, in particular, the apparently poor reporting of digital persona and data movement information, and the lack of information on where data is located and processed. These data lie outside of the scope of the Data Protection Act 1998 and it may be suspected that this is a reason that organisations do not provide the information when it is asked for. From the data collected, it was not possible to determine why these patterns were observed or if there is a view on the usability of the model. This forms part of the third and final stage of this research, which takes the form of interviews with subject area experts, and is described below in Chapter 5.

Chapter 5: **An Expert View**

5.1 Introduction

This research demonstrates, in Chapter 3, that the terms used to describe personal data are used inconsistently. In response a standard nomenclature is proposed (Parkinson et al., 2017) which enables the classification of personal data, and which is illustrated in a model presented in Chapter 3. As discussed in Chapter 4 the second phase of this research illustrates that when organisations answer requests for personal data the information provided is inconsistent and incomplete. As a result, an individual cannot expect to retrieve all of the data held by organisations that is descriptive of them. The analysis of the data collected also raises a number of issues, which lead to the fourth research question:

> *RQ4: What are the reasons for the variations found in the performance of different classes of organisations?*

This chapter reports on the semi-structured interviews used to explore possible explanations for these variations (the interview guide for which is shown in Appendix P). Interviews with experts allow the researcher to drill down into the reasons behind observed events (Barbour and Schostak, 2005). In this case, representatives from public and private organisations and with roles relevant to the research topic are in a position to shed light on the reasons behind differing performances found during the analysis of data from phase two of this research.

The thematic analysis, reported below, indicates that there are two motifs underlying all interviews. First, the willingness of an organisation to provide information, and second the capability of the organisation to carry out its intended approach.

5.2 Method

The objective of the final phase of this research was to obtain and analyse expert opinion on the outcomes of the first two phases. In response to the research question opinion was sought from experts from different classes of organisation (public and private) and with different roles in order to garner opinion from multiple points of view. Since it was the perspectives of ‘experts’ that was of interest, the methodology most suited to this was qualitative. Three approaches were considered - focus group, survey, or interview. A focus group bringing together experts from different disciplines relevant to the subject area would have the advantage of exploring, in a relatively short period of time, those apparently complex phenomena (and opportunities) that were hypothesised from the initial

phases of this research (Powell and Single, 1996). However, given the difficulties encountered in getting diary dates from an expert panel, and the practical difficulties of bringing them together in one place, this approach was judged to be infeasible.

On the other hand, a survey has the general advantage of covering large groups at a low cost, and can provide a level of anonymity, but is normally used to analyse more than one case at a point in time as a part of a cross sectional design (Bryman, 2008). This method answers the what, how often, and to what extent questions well but has weaknesses in helping us understand the how or why of a situation, which this phase of the research is aiming to do.

As a result, it was decided that semi-structured interviews with a purposive sample of experts would be appropriate to illicit possible reasons behind the findings from phases one and two of this research, based on the research questions listed above. This not only enabled the collection of rich and in-depth information from the panel of experts but also allowed for follow up questions to be asked that could probe deeper into the issues raised during the interviews. It allowed a conversational style to be adopted thus diffusing the potential power of the interviewer, whilst balancing it with a need for conformity in an attempt to obtain considered answers to this research question (Dunne et al., 2005).

The detail of the method used is illustrated below in figure 5.1. Potential interviewees were selected from the following groups:

- government legislators, in order to obtain views upon the use of the model in legislation, on the performance of governmental bodies, and on whether legislation should be changed;
- members of think tanks for their views across all the areas of interest;
- data protection professionals, for the expertise in possible use of the model in answering subject access requests, the reasons for variations in responses, whether legislation should be changed, and how answering subject access requests was seen within organisations;
- IT management for their views on subject access requests within organisations and the use of the model.

Individuals were selected where possible, from personal contacts, thus forming an opportunistic sample. This was key in getting access to professionals who would otherwise be very unlikely to take part, for example it was effective in securing interviews with two former cabinet members whose roles included some interest in personal data. and three senior IT professionals, one of whom owned their own company. The data protection officers and the think tank director were interviewed following positive responses after

cold calling a number of individuals and organisations. Of the nine interviewees, four had governmental experience, two as cabinet members and two as civil servants in large government departments. The two civil servants comprised one IT manager who had recently spent 10 years in a senior IT strategy role and one data protection officer who directed the department responsible for responding to large numbers of subject access requests. In addition, one interviewee chaired a hospital trust at the time of interview. This is a small interview set constrained by accessibility (for instance the ICO was contacted on three occasions but declined to respond), and by the time and resource constraints consistent with doctoral research.

The process of conducting interviews is shown in Figure 5.1. All interviews were recorded with two devices and the recordings moved to an encrypted laptop within directories identified as Interview 1 though 9 in order to preserve anonymity. During the interviews care was taken not to use names or descriptions that would easily identify the interviewee, although during the interviews descriptions of events were relayed that could be used to identify the individuals involved.

All interviewees were provided with information sheets (Appendix N), and they were asked to sign a consent form (Appendix O). One data protection officer declined to sign the consent form but stated they were happy to be interviewed and recorded, as this itself was recorded, the interview took place. Interviewees were also offered transcriptions of interviews but all declined the offer.

All interviews were guided by a series of questions supported by prompt statements (Appendix P) which reflected the research questions whilst attempting to produce a flow through the interview. My background is as an experienced interviewer in the context of requirements collection and problem analysis within the IT industry, and I have undertaken research methods training. When solving a person's problems or collecting requirements for a new IT system, the interviewee is in many respects driving the interview due to their own needs. As a researcher, it was my needs that drove the interview and also meant that I had to describe the background to the interview. This required different techniques and whilst I endeavoured to emulate Kvale's (1996) guide for a successful interview, I consider that it was quite a challenge. A noticeable problem is one of interviewer bias, in the way that questions and follow up questions are asked, or elements emphasised. An awareness of this possibility should have moderated the effect but I am aware that there may be some sub-conscious bias in the manner in which the questions were asked or emphasis created in prompts. Also, the interviewee gains an association with the interviewer during the process. So, for example when asked about the worth or usefulness of the model previously

presented by the interviewer all interviewees responded, by and large, in a positive manner. The effects of this will be considered in section 5.4 Analysis.

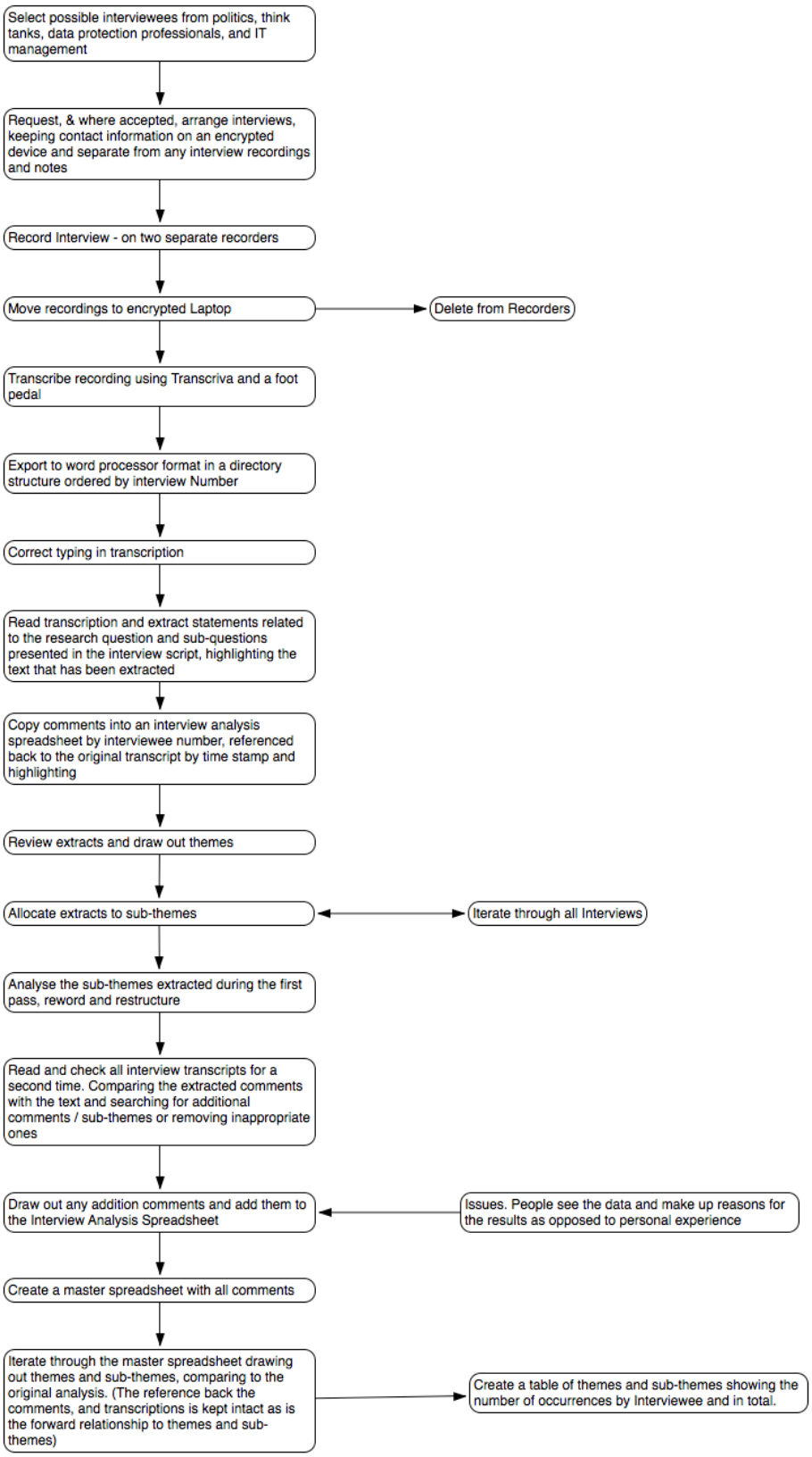


Figure 5.1 Interview method diagram.

It would have been possible to have the interviews professionally transcribed but after some consideration, I decided to transcribe them myself. This enabled the interview technique to be reviewed after each interview and for the responses to be reviewed more closely. In some cases Audacity's filtering capability was used to clean up the recording and reduce background noise.

Following transcription, the interviews were replayed and the transcriptions verified. An example of an anonymised and verified transcription is to be found in Appendix R.

In the final stage the interview transcripts were analysed twice. Each iteration examined the transcripts for recurring topics, similarities and differences between interviews, and linguistic connectors such as 'because' or 'since' to highlight causal connections. This process produced 17 sub-themes. These were then, analysed for groupings with common characteristics, again using a series of iterations until a consistent categorisation was created. The resultant groupings were then labelled resulting in four themes. A weakness in this approach, is again the partiality of the researcher, however, the constraint of this research is that it is performed by a single person. It could also be argued that even if more than one person was involved with the analysis, some level of bias would be present and the results may be no more valid (Armstrong et al., 1997). In order to mitigate some of this effect the transcriptions were analysed a second time, checking for new themes and sub-themes, extracting new quotes, allocating them to the thematic structure, and establishing consistency of this approach. The themes and sub-themes were re-evaluated resulting in new wording, restructuring and some changes. The differences between the first and second thematic analyses are given in Appendix BB, which shows that seven new sub-themes were identified, with 4 being discarded, (making 17 in all).

5.3 Findings

Nine interviews were conducted. They ranged in elapsed time from 25 to 96 minutes with an average length of 66 minutes. The participants were all experts in their different fields, information technology (3), one of whom was an entrepreneur, data protection officers (3), politics (2), and, think tanks (1). Of these four had in excess of 10 years recent central government experience.

From the analysis of these interviews 17 sub-themes were identified relating to aspects of organisations that affect their perceived and actual responses to subject access requests. They are categorised into 4 themes, described below: culture, people, capacity, and governance. These interrelate when viewed from the lens of an organisation's ability to function, but here are used to classify perceived causes for the differing ways organisations

respond to subject access requests. Each theme is briefly described in the following paragraphs in decreasing order of the total number of observations. The summary of the thematic analysis may be found in table 5.1 In this table, each interviewee is identified with an area of expertise which is expanded in the next table 5.2 along with a brief description of them. In addition, the full thematic analysis can be found in Appendix R.

Interviewee Identifier		1	6	7	2	8	9	3	4	5	
Area of Expertise		Ent	IT		DPO			Pol		TT	
Theme	Sub-theme										Total
Culture	Approach to SARs	4	7	1	4	2	7	2	0	0	27
	Transparency	2	0	2	1	0	3	5	1	4	18
	Customer Focus	3	0	0	1	0	0	1	0	1	6
	Protective	0	0	0	0	0	0	1	0	4	5
	Efficiency	0	1	2	0	0	0	0	0	0	3
People	Understanding Personal Data	2	3	4	5	5	4	8	5	9	45
	Knowledge / Training	2	0	0	4	1	0	1	0	2	10
	Trust	0	2	1	0	0	0	0	0	2	5
	Common Requests	1	0	0	0	1	0	0	0	0	2
Capacity	Capability (IT or Otherwise)	0	1	9	0	1	4	1	0	2	18
	Size	6	0	0	2	0	0	1	0	2	11
	Processes	2	0	0	3	2	2	1	0	0	10
	Structure	1	0	0	2	0	1	2	0	0	6
	Competitive Situation	2	0	2	0	0	0	0	0	0	4
Governance	Practice	2	0	1	1	3	1	1	2	0	11
	Mission & Vision	1	1	2	0	0	0	1	0	2	7
	Disposition to DPA & SARs	1	0	0	1	1	0	0	0	0	3
		29	15	24	24	16	22	25	8	28	191

Table 5.1 Summary of thematic analysis

The first theme, **culture**, encompasses views from all 9 interviewees and embraces 5 of the sub- themes. It refers to the beliefs and behaviours which affect the way an organisation responds to subject access requests. The first sub-theme within this theme, and most

observed with 27 instances, is the cultural positioning of the organisation's **approach to SARs**, and how the Data Protection Act 1998 legislation is considered in that context. For example, the information technology entrepreneur stated with respect to subject access requests that 'they are definitely considered a cost' and that 'data protection is a thorn in my side'. **Transparency** is the second most observed sub-theme in this theme, with 18 instances, and covers the cultural attitude to individuals having knowledge of, and ideally access to, data, especially personal data. As a think tank policy director said 'I should know who shares my data across government'. Other sub-themes all with less than 10 examples are in decreasing order of observation; **customer focus** - where giving a customer copies of their personal data is seen as good service; **protective** - where one of an organisation's main considerations is to fend off requests for data; and **efficiency** - the ability, or competence of staff to respond to subject access requests.

The second theme, **people**, was again raised by all interviewees and covers 4 sub-themes relating to the characteristics of an organisations representatives that affect how they behave towards requests for personal information, and how those representatives are viewed by the public. The first, and most observed, of all sub-themes being **understanding personal data**, with 45 observations. As a data protection officer said when referring to another organisation 'they will be exposed to data but they won't recognise it as personal data'. With 10 observations, **knowledge and training** refers to the level of training and expertise of the staff who deal with subject access requests. A board member of the National Association of Data Protection Officers stated 'they are not used to being asked this question so don't understand how to answer it. They won't know where the stuff is'. The final two sub-themes in this theme are people's **trust** of an organisation from outside; and **common requests** - employees supplying only commonly requested data.

The third theme, **capacity**, was raised by 8 of the 9 interviewees and encompasses 5 sub-themes, which focus on the resources, competitive situation of an organisation, and its ability to fulfil its objectives. **Capability** was mentioned 18 times and refers to the proficiency of an organisation to provide the data required in answering subject access requests, for example a Data Protection Officer from a utility observed of their organisation 'there is a known issue, and it wasn't to do with people it was to do with the actual system unfortunately'. The issues relating to organisational **Size** was mentioned 11 times, the Think Tank Policy Director stated that 'for a large company it is a complete nightmare to try to find all the information'. An organisation's **processes** were mentioned on 10 occasions. Organisational **structure** and **competitive situation** were the remaining two sub-themes with 6 and 4 mentions respectively.

Finally, **governance**, reflects 3 sub-themes relating to perceptions of how the processes of high level decision making influence the organisations' responses to subject access requests. **Practice**, covers comments relating to the day to day governance issue for example a Data Protection Officer from the utilities sector suggested that 'a plc and a private company have pretty much the same governance requirements'. The final two sub-themes were **mission and vision** - how the mission and vision of the organisation impact on personal data e.g. for a Magic Circle law firm confidentiality is of upmost importance and all of their data is held in Germany because it has the highest level of data protection in Europe; and the final sub-theme is an organisations high level **disposition to DPA & SARs** access requests.

Interviewee			Interview Duration (mins)
Id	Description	Area of Expertise	
1	Owner of a tech company employing c120 people (Entrepreneur)	Ent	96
2	Member of the Management Group of the National Association of Data Protection Officers (Data Protection Officer)	DPO	85
3	Former Member of the UK Cabinet [1] (Politician)	Pol	60
4	Former Member of the UK Cabinet [2] (Politician)	Pol	25
5	Think tank - policy director (Think Tank)	TT	97
6	Recently retired IT Director of FTSE 100 Company (Information Technology expert)	IT	70
7	Senior IT Professional Magic Circle Law Firm (Information Technology expert)	IT	46
8	Data Protection Officer - Utility (Data Protection Officer)	DPO	59
9	Data Protection Officer - Government (Data Processing Officer)	DPO	56
		Ave	66

Table 5.2 Interview identity coding showing length of interviews

5.4 Discussion

When examining the third research question

> *RQ4: What are the reasons for the variations found in the performance of different classes of organisations?*

The experts considered the observations, presented during the interview, from the first two phases of this research and offered their professional judgements to explain, or contradict, those findings. However, their opinions are split as they speak either from an internal perspective based on the knowledge gleaned from their own organisations operating practices, or as an external observer of organisations of which they have little personal working experience. As a result, this research presents three perspectives on what happens when an individual applies for their data using the subject access request mechanism, the data subject's (in this case, the author), the expert's view as an internal voice, and the expert's view as an observer, so for instance a former member of the cabinet considered government open and transparent whilst outside observers considered it either closed or incompetent.

The analysis of the interviews, described above, suggested 17 sub-themes which were grouped into 4 themes; culture, people, capacity and governance. However, there were also two motifs observed that ran through the interviews and across the observed sub-themes. First, whether an organisation wanted to provide data to individuals, or not, which may be considered as gradations of **willingness**. Second, whether an organisation had the skills and resources to provide the data requested by individuals, which can be classified as gradations of **ability**. These two aspects are exemplified in a quote from the senior IT professional from a magic circle law firm, who had previously worked in central government, '[Y]ou know if you asked MI5 what your information was I would have expected them to give you the bare minimal. If you asked the DWP it is only because they are incompetent'. The opinion being that MI5 would be unwilling although probably able and the Department of Work and Pensions whilst perhaps willing would be unable to provide the data. It is against these two motifs that the sub-themes and themes extracted from the interviews will be considered.

5.4.1 Willingness

Brady and Cronin (2001) argued a willingness to help the customer, and an expectation that the employees will go the extra mile, results in customer perceptions of quality and satisfaction and is inherent in a customer focused organisation. It may be thought that to

the individual submitting a subject access request the issue is whether data is provided, in some respects it is irrelevant whether the organisation is eager to fulfil the request or not. However, with a willing organisation it could be argued that information is more likely to be produced and that a fuller reply may be received. It is also possible to write back politely, offering some direction, and thus obtain the data that is required. In this research the UK Border Agency initially refused to provide data but after a follow up letter they contacted the researcher and provided most of the expected information. One may consider this a willing but unable organisation, and it was later discovered that at the time they had a backlog of over 4,000 cases (Interviewee 9 00:23:22.54). On the other hand, the Office for National Statistics, whose business is data, its storage, analysis and curation, would appear to be a very able organisation but is unwilling to respond to subject access requests. Indeed, under the terms of the Data Protection Act 1998 it is considered a research body and so is exempt, although this does not prevent them from providing to an individual their personal data. They responded thus to a follow up request.

> 'Because of this I am repeating my previous reply that we (as well as all other UK data controllers who process data for only scientific and statistical purposes) are exempted from the section 7 subject access provisions by section 33 of the Data Protection Act and will not process your request.' Letter from ONS 29th March 2014 Appendix U

In addition to these initial observations of the researcher, there are two other perspectives offered from the interviewees, that of those experts who have worked internally to an organisation in question, and those who view the organisation from outside. For example, consider the sub-theme of **Transparency** within the theme of **Culture.** When referring to central government, Interviewee 3, a former cabinet minister stated, when comparing central government to other organisations that 'the government organisation is more transparent.' On the other hand, Interviewee 1, the IT entrepreneur, said with respect to governmental organisations sharing data 'what I have generally noticed about central or local government is their love of the word no'. This disparity may be accurate as the leader of an organisation may be more optimistic about its performance than a more jaundiced view from someone on the receiving end of its service. On the other hand, both may be accurate in that government may be as transparent as they are capable of, which is not very much. As the senior IT professional from a Magic Circle law firm (who had worked for 10 years in a government agency) stated with respect to the same issue '[y]ou can see what they can see, they just can't see very much'. People within central government tended to believe that they are transparent and willing to share data, in general people who have not

worked there consider it to be protective. Indeed, the reverse may also be true. The Data Protection Officer working in government had occasion to submit a subject access request of his own to a private company and found them to be unhelpful (Interview 9 00:47:56.09). He also consistently referred to customers rather than clients suggesting a customer focused culture which would have been consistent with a willing attitude towards providing data.

There is therefore a range of views on each sector, and perhaps even each organisation's, willingness to respond fully to subject access requests which would lie along a continuum from totally unwilling, to wholeheartedly willing. However, willingness to do something delivers little if there is not the ability to carry through with those wishes.

5.4.2 Capability

In this research a willingness to provide data was observed to contrast with an ability to actually retrieve and present the data back to the individual. The Data Protection Officer working in Government illustrated his willingness:

> 'at its peak we had around 4,000 cases that were overdue back in August, September last year, and we have worked really hard through a combination of process reviews, continuous improvement methods, and just kind of influencing customer behaviour in that way as well, doing a bit of triaging, a whole range of things, to help us get completely on top of that' (Interview 9 00:23:22.54).

However, he admitted that 'the most difficult part of that process is printing off the IT records in the first place because of the way that the database is structured' (Interview 9 00:39:25.25). It was for him a problem of capability rather than willingness, in this case his **processes** and **capacity** hindered him. The member of the management group of the National Association of Data Protection Officers observed that ability to deliver was also affected by an organisations structure and **size**:

> 'in a very, very big organisation where your analytics team could be part of a digital team, an IT team, part of the marketing team they could be absolutely anyone you simply wouldn't know where you would start internally getting an answer to the question [of where an individual's personal data was held]' (Interview 2 00:14:06.77).

This was supported by the Think Tank Policy Director who suggested that 'for a large company it is a complete nightmare to try to find all the information' (Interview 5 00:28:02.54). There was some evidence of this from phase two of this research. For example, Royal Mail initially replied:

> 'Please could you give us an indication of what parts of our business may hold information on you as it is not possible to complete an open-ended search of all the records that we hold' Royal Mail Letter 17th April 2014 (Appendix V).

These viewpoints add to our understanding of issues that affect an organisation's ability to provide data. Peppard and Ward (2004, p. 176) helpfully define IT capability as 'the ability to translate the business strategy into long term resourcing plans that enable the implementation of the strategy (i.e. the IT strategy)'. Thus, framing capability against business strategy, which in this instance may be reduced to the willingness or not to provide data in response to subject access requests. This can be seen in Equifax, perhaps a more customer focused (and willing) organisation who took an alternate approach as the main contact sourced replies from areas throughout the organisation:

> 'Please find below the details for the latest Subject Access Request. Please compile all the relevant information and return to the **External Subject Access/uk** inbox **as soon as possible** but no later than **2nd January 2014'** extract from Equifax correspondence 8th January 2014 (Appendix W**).**

Peppard and Ward (2004) also talk about business strategy and in general, this is to support the maximisation of shareholder value. A former member of the UK Cabinet (1) suggested that Government was more nuanced than business, the implication being that in government it is inherently more problematic to achieve an able organisation due to the increased complexity of setting clear goals and mission statements:

> 'Well, as Donald Trump is discovering, government is a lot more complicated than business, and actually you will actually probably have a range of objectives and a hierarchy of objectives in government' (Interviewee 3 00:28:24:95).

On the other hand, comparing governmental organisations to the private sector, the recently retired IT Director of a FTSE 100 company suggested a simpler rationale:

> 'I wonder if historically some parts of private enterprise have got higher investment in computing so that they have just got a lot more modern stuff. Where it is possible to pull it off more easily. Whereas maybe some government bodies are 4 or 5 years behind and so they probably have not got it all together, which probably has an impact.' (Interview 6 01:01:56.59).

As with **willingness** there is a continuum of **capabilit**y to provide data in response to subject access requests, and a number of suggested reasons for any organisations positioning on that continuum. However, willingness and capability can be positioned in relation to each other to form a matrix, as described below.

5.4.3 Willingness / Capability Matrix

Hersey and Blanchard (1988) when arguing for a situational leadership model effectively constructed a willingness / ability analysis to categorise what they described as followers readiness, and would represent this as a matrix for presentational purposes. Here it will be applied to organisations with respect to replying to subject access requests (figure 5.2). The vertical axis represents willingness, whether the organisation has the necessary confidence and commitment. The horizontal axis represents ability, whether the organisation has the necessary knowledge, skills and resources to retrieve and present data.

This phase of the research does not examine the relationship between willingness and capability but reports on observations made by the expert interviewees, for instance indicating that central government has low capability but in the view of interviewee 9 was willing. The research solely places organisations and organisational categories into quadrants within the willingness / capability matrix but provides no insight into whether, or not, there is a causal relationship between willingness and capability. Therefore, it is not possible to report on any linkage or relationship between the two observed characteristics. It may, however, be hypothesised that increased capability for instance, by the provision of improved IT systems, may have a positive effect on staff, or indicate a change in organisational direction, to an extent that the organisation becomes more willing to respond more fully to subject access requests. King and Burgess (2008) when modelling the implementation of Customer Relationship Management systems, report a drop in work quality following systems implementation which later improves as staff become more experienced in the new system. Whilst no research has been identified linking willingness and capability, it may be additionally hypothesised that increases in capability may result in delayed improvements in willingness and vice versa. Such hypotheses require further research of a longitudinal nature and are outside the scope of this work.

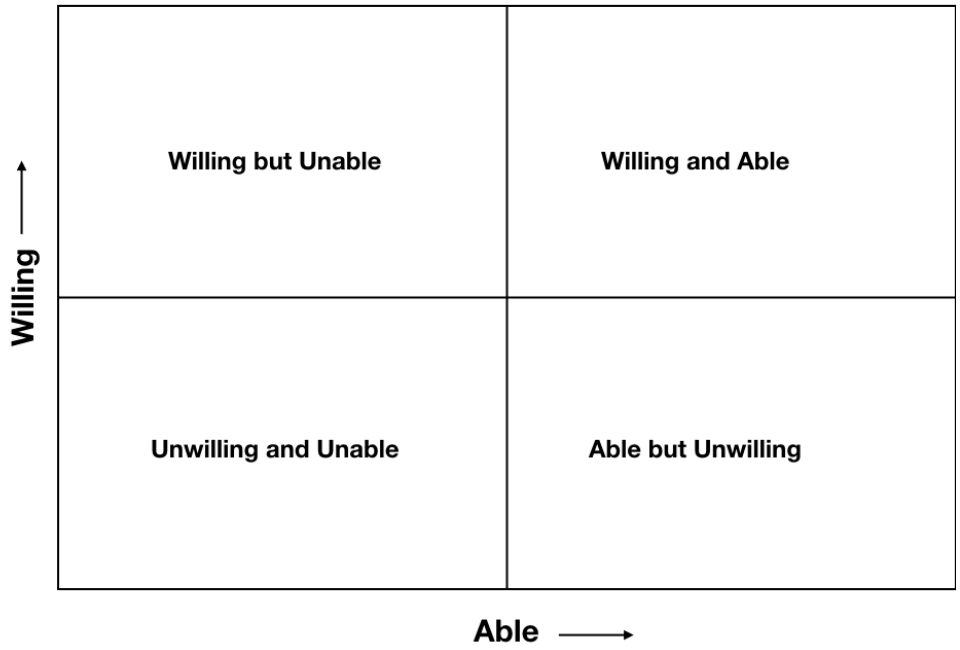


Figure 5.2 Matrix for interview analysis.

Individual organisations or sectors will now be positioned within the matrix, first from the evidence obtained during the interview process (Figure 5.1) and then, by comparison, from the results of the data collection exercise described in Chapter 4. This will provide another perspective for an organisation/sector's performance and also illustrate the differences of opinions between the interviewees and results of the data collection.

Appendix X summarises the research findings with respect to ability and willingness from the viewpoints of the interviewees. It is represented in Figure 5.3 below. It illustrates that, in the view of the interviewees, both local and central government are viewed as having little ability to extract and present data in response to subject access requests although there is a range of opinion regarding their willingness to do so. On the other hand, private and public companies and charities are seen as willing and able with two exceptions. The first is a view put forward by the Think Tank Director that charities may lack ability due to low staffing levels and the second by the Data Protection Officer from Government from their own experience, where one company was not willing to provide data.

The reason for the poor ability of local and central government is considered to be due to large siloed organisations, poor quality staff, and outdated IT infrastructure. In the case of HMRC a resistance to data sharing was also said to be cultural, as a result of the conditions of their foundation charter. On the other hand, the higher performance of public and private companies is considered to be the result of market competition driving customer service resulting in better knowledge of personal data and more effective IT systems. These organisations also tend to have higher compliance requirements and as a result, demonstrate effective processes to respond to subject access requests. Those who judge

charities to be willing to provide data give the culture of openness and their need of data to exist as the reason.

In this analysis Interviewees 3, 7 and 9 had all worked in or run central government departments. Their view of government is highlighted in the next matrix. They agreed with those interviewees from the private sector on the lack of ability within central (and also local) government. The difference arises with respect to willingness, those inside government believe that there is a willingness to provide the subject access request data hampered only by their skills and IT infrastructure. Those outside consider that there is a culture of protectionism and self-ownership in connection with personal data. The contrasting views may not be in conflict. It is possible that government is now relatively transparent compared to the situation 20 years ago, whilst being much less flexible and willing than the private and NGO sectors. In the next section the evidence from the data collection will be compared to the views of the interviewees.

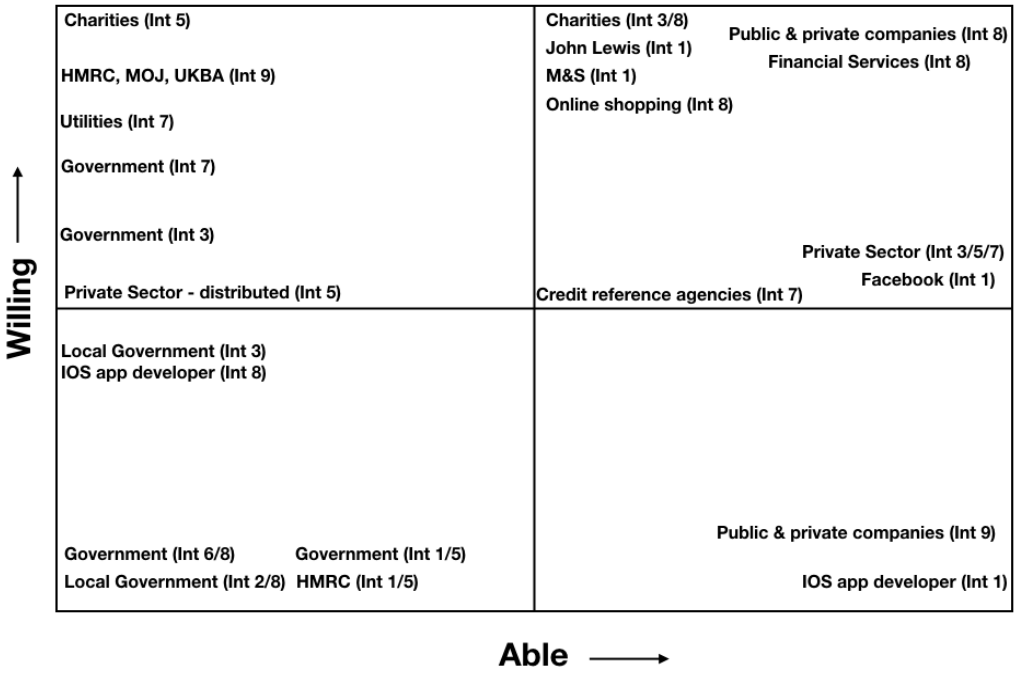
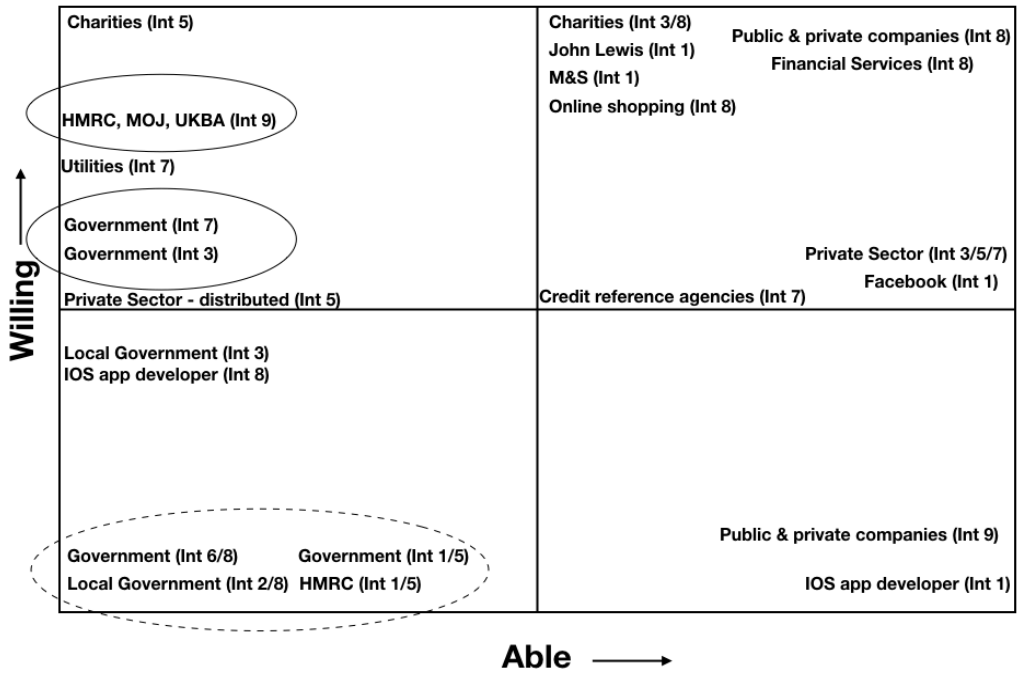


Figure 5.3 Matrix showing willingness and ability positioning derived from the interview analysis.



Key ———— internal view of government

- - - - - - - - external view of government

Figure 5.4 Matrix illustrating governmental performance from two viewpoints.

During this research five government bodies were contacted for information and a summary of their perceived transparency can be found in table 5.3, which is an analysis of their replies, (as opposed to the data that they sent). On the one hand this supports some of the interviewees' explanations. Subject access requests were originally dealt with by Freedom of Information teams in HMRC and The Met Office. This supports statements that Freedom of Information requests have a higher priority than subject access requests, indeed none of the replies in government came from a subject access request team or as is common in the private sector, a customer services team. In addition, the experience with the NHS supports the conjecture that large distributed organisations find it difficult to gather data but also the corollary that an individual finds it difficult to contact the organisation. At the time of writing there is still no one contact point within the NHS to deal with subject access requests.

On the other hand, claims of transparency and a willingness to provide data are not supported. Three of the organisations quoted exemption status from the Data Protection Act 1998, and of the two remaining the NHS could not pass the request to a relevant team, so data was only received from one of the five organisations unencumbered. The Data Protection Act 1998 does not stop organisations sending an individual their personal information, it does provide a pretext behind which an unwilling organisation can hide, and

it would appear, national security issues aside, that this is the case. Indeed, the evidence from the data collection exercise supports the views of those external to government, that it is in the main unwilling and unable to respond fully to subject access requests thus explaining its poor performance presented in Chapter 4.

From the analysis in Table 5.3 we can see that contrasting viewpoints were expressed by the interviewees with respect to the performance of organisations and sectors. In addition, the perceived reasons for the variations in performance have also been examined. The positioning of organisations and sectors within the matrix is helpful in understanding possible high-level reasons for performance of organisations when providing data in response to subject access requests but does not address any possible effects of organisational type upon the findings.

The categories used have been local and central government, private and public companies and NGOs. It is outside the scope of this work to examine the structures of the sample organisations contacted during this research, but it is possible to look briefly at general attributes that have been ascribed to different types of organisation, and consider whether they would impact on responsiveness to subject access requests.

First formalisation, government organisations, and to a lesser degree those bodies who have regular contact with them tend to have high degrees of bureaucracy (Tolbert and Hall, 2009). Offe (2009) states that the classical forms of sovereign state action tend to act on the command/threat, of coercion/obedience, and on the other hand argues that private and civil society allows for spontaneous coordination of action within a social and normalised framework, and suggests that markets are outside the conceptual field of governance as they exist for the private maximisation of gain. It may therefore be conjectured that governmental organisations may be slower to respond and more rules based than companies and NGOs. Whilst not impacting directly on willingness or capability, it suggests that they may be slower to respond to changes in culture and legislation which have led to the provision of personal data under the Data Protection Acts.

Second, it was suggested by interviewees that local government and central government responded poorly due to their complexity and their size. The figures from April 2018 suggest that there are in the region of 5.4 million public sector workers in the UK (Guerin, McCrae and Shepheard, 2018). Of these HMRC employs 56,000 and UKBA 23,500. On the other hand, John Lewis employs 85,500 and Lloyds Bank 75,000, and both provided better quality responses, indicating that size is not necessarily a constraint.

The third issue is complexity, Guerin, McCrae and Shepheard (2018) state that government is becoming more complex and Interviewee 2 also suggested that local government

complexity in providing a wide range of services contributed to the poor subject access request response. Compare this to the John Lewis Partnership which runs two main businesses (John Lewis and Waitrose) and Lloyds Bank which runs 11 (Lloyds Bank, Bank of Scotland, Halifax, Scottish Widows, MBNA, Black Horse (Motor Finance), LEX Auto Lease, LDC (private equity), AMC (agricultural mortgages), Colleys (valuation & Surveying), Birmingham Midshires) each of which is run as an independent entity with their own data protection officers suggesting that complexity may be an issue.

Finally, on compliance and accountability, Interviewee 8 suggested that private and public companies responded better to subject access requests because they operated in a stricter compliance environment. On the other hand Guerin, McCrae and Shepheard (2018) suggest that weak systems within government are compounded by weak parliamentary scrutiny and that where accountability is weak, it can lead to chronic underperformance, poor value for money and outright failure. An examination of the Data Protection Act 2018 shows nine main areas for exemption from responding to subject access requests. Six of these cover government institutions whilst education and research exempt both public and private organisations and journalism is the only exemption without a tie to government. This is not the place to argue the validity of the exemptions but rather to suggest that the poor levels of accountability and exemption from compliance may create cultures less favourable to fully providing answers to subject access requests, impacting on willingness and then also on the provision of capability.

However, if there is a will within the legislature to improve the outcomes for such requests, it may be instructive to return to Hersey and Blanchard (1988) and suggest various courses of action.

Organisation	Quoted DPA Exemption	Initial Contact	Data Sent	Comment
HMRC	Yes	Freedom of Information Act Team	Yes	The data sent was limited to that data supplied by the individual to HMRC in the form of tax returns. No additional data was provided referencing DPA exemptions as the reason.
Met Office	No	FOI Officer	Yes	Met Office IOS app does not collect personal data.
NHS	No	Information Access Officer	No	Contact was made with the wrong department of the NHS (actually a complaints department). They suggested contacting the GP. In order not to take up valuable GP's time this was not progressed. There is still no central NHS body on the Data Protection Register.
ONS	Yes	Legal Services	No	Quoted exemptions under the DPA.
UKBA	Yes, with respect to redactions	Records Management	Yes	Details of journeys in and out of the UK. Note this was not complete due to the e-Borders System not being fully rolled out, and a period where, due to budget cuts, records were not maintained (confirmed by Interviewee 3). No other information was provided. Interviewee 3 said this would be due to issues of national security.

Table 5.3 Analysis of transparency observed from government organisations

For the unwilling and unable, it would be necessary to provide clear legislative direction but also to closely supervise their day to day performance with serious sanctions for non-compliance. For the unable but willing, it should only be necessary to explain the need to comply thus encouraging improvements in performance. The able but unwilling, require a participatory approach with the reasons for compliance explained. Finally, the willing and able, require little input as they will take responsibility for their own compliance. The General Data Protection Regulation 2016 enacted in the UK in April 2018 provides the

individual with wider data access rights and also imposes greater sanctions on organisations which together with the educational initiatives of the Information Commissioner's Office will hopefully increase the willingness of organisations to provide data and also for those lacking capability to address their shortcomings.

5.4.4 Feedback on the Model

The model was presented to the interviewees as an introduction to the findings of the data collection exercise thus providing background and context. All interviewees were asked about the helpfulness of the model and its application. As expected they all said that the model was helpful, but it would have been exceptional if any had said otherwise given that its creator was sat in front of them. However, their observations as to why, do indicate that there is a need for such a model.

Interviewee 3, a former member of the UK Cabinet, indicated that it would have been of use as a vehicle to explain the data that the government wished to collect and retain and that which it did not. This was reflected by the second former cabinet member Interviewee 4 who also stated with respect to the terminology deployed to describe personal data 'you can get lost in the mystification it is unbelievable' (Interview 4 00:01:49.63). It was additionally useful during the interview to differentiate overt surveillance, which creates digital footprints for instance as the individual passes a clearly visible camera, from covert surveillance, creating third party digital footprints when the individual is recorded without their knowledge by a third party. For digital footprints, the agency is with the subject individual but for third party digital footprints it is with another person. As Interviewee 5, the Think Tank Policy Director, stated in order to limit what can or cannot be done with parts of data you have to be able to define it, and this is what the model does. The opinion was that the model should be of use in legislation.

Interviewee 7, the senior IT professional from a Magic Circle law firm with a background in IT strategy and architecture, offered another viewpoint suggesting that whilst the model was not directly useful for data architects, it could be used by information scientists especially when dealing with big data with respect to marketing, targeting and discovery.

Finally, Interviewee 9 the Data Protection Officer from Government, thought that the model would not be useful in their role as they just send data out. Perhaps a sad reflection of the lack of understanding of personal data in some areas or else an indication that although they ran a large team responding to subject access requests, it was someone else who decided what could be seen and what couldn't.

5.5 Conclusion

This Chapter has examined nine interviews which addressed the third research question

RQ4: What are the reasons for the variations found in the performance of different classes of organisations?

The interview transcripts were analysed and indicated 17 minor sub-themes relating to aspects of organisations that affect their perceived and actual responses to subject access requests. They are summarised into four major themes:

- **culture**, the approach to subject access requests, how transparent the organisation is in the provision of data, the level of customer focus, how protective and controlling of data the organisations are, and efficiency (or quality) of staff;
- **people**, their understanding of personal data, the level of knowledge and training that they receive, the trust individuals have with respect to the organisations, and the result of replying only to common requests;
- **capacity**, the capability of the organisation e.g. with respect to IT systems, the size, structure and competitive position, and the presence or lack of processes;
- **governance**, the management direction, vision and mission of the organisation and its disposition towards the Data Protection Act 1998 and subject access requests.

Underlying these are two motifs; the willingness of organisations to provide data in response to subject access requests, and their ability to do so.

The most marked variation in performance arising from the research described in Chapter 4 was that between government (central and local) and the private sector organisations, including NGOs. With four of the interviewees coming from government or having long experience in it, and four from the private sector with the ninth interviewee from a think tank that advises central government, it has been possible to compare internal and external assessments of this situation. Internally, government was thought to be transparent, and willing to provide data although restricted by their organisation and systems. Externally, the view was one of an unwilling organisation, poorly structured with poor systems and poorly trained people.

A third assessment is presented above, using the evidence from the data collection exercise covered in Chapter 4. This supports the consensus opinion from this small sample regarding the low capability of government organisations to provide data compared with the private sector. It also however, points to government organisations that are on the whole unwilling to provide data in response to subject access requests, either for genuine security reasons or more commonly because they feel that they do not have to do so.

This research presents a less than complimentary view of government's attitude and capability to subject access requests and by inference it may be assumed that the private sector and NGOs perform to a much higher level. Whilst this is true to an extent, there are still problems in both of these areas. Some of the smaller NGOs lack the resources, and the private sector encompasses, on the one hand, very small organisations who it is suggested lack knowledge of responsibilities under the Data Protection Act 1998, and on the other, large distributed organisations who find it difficult to consolidate a response.

Chapter 6: **Conclusion**

In this final chapter the four Research questions will be addressed drawing on the overall findings. The limitations of the study design are discussed. The unique contribution to knowledge made by the study is set out and the implications for policy, practice and further research are identified.

The aim of this study is to further the understanding of personal data from an individual's perspective, how that data are described and how they can be accessed. This has been progressed through an analysis of the literature using terms deployed to describe personal data; the collection of the author's own data by submitting subject access requests to a sample of organisations, and an assessment of their responses; and finally, to throw light on the data access issues observed, by interviews with experts in the field.

In the first phase of this research, described in Chapter 3, 16 terms that label personal data, were used to extract 64,584 publications, from which 247 were selected, based on relatively high citation count relative to recency of publication date. It was shown, by analysing the terms used to label personal data, that they are used inconsistently. A new categorisation for personal data was then developed, its whole being labelled the digitally extended self, and a set of terms recommended as a way forward to limit inconsistency in meaning. This has been presented as a model to illustrate the relationships between categories of data, in response to the first research question:

RQ1: What are the components of the digitally extended self and how do they relate to one another?

To obtain a level of validation, the model was compared with 45 extracts from the privacy and personal data literature selected for their wide range of concepts, and in which similar terms are used in different contexts. This indicated that the model was robust. Chapter 4 described the next part of this research, during which 82 organisations were contacted with requests for the author's personal data, of these, 58 responded and provided information. The model was matched with the personal data provided. This indicated that the model was valid and suggested that location could be seen as a useful attribute of each data category. The process of collecting data to test the model also facilitated the second objective of the research, which was to examine whether an individual can access their personal data, and what the issues were associated with that activity, as reflected in the second research question:

RQ2: How feasible is it for an individual to obtain the information, held by organisations,

which is descriptive of them?

In a process of digital auto-ethnography I sent subject access requests to 32 organisations and a further 59 in a snowball sample obtained when organisations were mentioned in the original responses. All of the returns were analysed in terms of the model for completeness and links to other companies, 29 follow up requests were also sent in cases where I believed the data supplied was incomplete. In total 82 companies were contacted, 58 of which returned data. Perhaps unsurprisingly, it is shown that individuals are not able to retrieve or observe all of their digitally extended self, for a range of reasons. Organisations are exempt from responding to requests, others do not provide all the relevant data e.g. Facebook, and other data does not have to be provided e.g. digital personas and the origin and destination of data transfers. In addition, the cost of pursuing this information is in the region of 43% of income after tax for someone on the 50th percentile income, and the time taken to contact the approximate 633 organisations who hold data is equivalent to a year's effort.

By collecting the author's personal data from a selected range of organisations it has been possible to address the third research question:

RQ3: What is the quality of the personal data returned by organisations when it is requested by individuals?

Analysis of the replies received from organisations shows that data at the centre of the centric visualisation of the digitally extended self (see Figure 4.3), were provided more often than data on the outer layers of the model. Digital footprints were obtained from 60% of those organisations willing to provide data, and third party digital footprints from 62%. The categories of data on the outer layers, digital personas and second level data, were provided by 28% and 29% of organisations respectively. There are also differences in the performance of organisations analysed by either category or sector. For those organisations providing data, central government was the worst performing category with a score of 1.25 out of 3.0 (as described in section 4.4.3) for the assessed data provided, with local government the next worse scoring 1.44. Public companies and NGOs were the best performing categories with scores of 1.99 and 1.96 respectively, whilst private companies scored 1.67. When the whole sample is considered and the replies analysed by sector, NGOs performed best, returning on average, information from 3.2 data categories, whilst IOS developers, returning 0.4 data categories, were the worst performing of those expected to provide data. In order for an individual to trace their data it is necessary for them to know where it exists. As part of the data collection exercise, organisations were asked if

they obtained data from outside sources or sent it to other bodies. NGOs were the best performing category scoring the highest in both cases with 3.0 and central government the worst with scores of 1.0 and 1.25 respectively. Private and public companies were in the middle with relatively high scores of 1.90 to 2.05. Possible reasons for these findings were explored during the final phase of this research through interviews with experts from relevant disciplines thus addressing the final research question:

RQ4: What are the reasons for the variations found in the performance of different classes of organisations?

Through the reported experiences of the interviewees, it was possible to examine the disparity between low performing central government and the higher performing groups. Seventeen sub-themes were derived from the analysis of the interview transcriptions, which were grouped into four themes, culture, people, capacity and governance. Throughout the themes and sub-themes two overriding motifs - willingness and capability - were observed to underlie all of the explanations for the differences in performance. Those with experience inside government believed that central government was willing to provide data but did not, in general, have the capability to deliver. Those outside government agreed that there was a lack of capability, but all believed that there was also a lack of willingness within central government to supply personal data. The analysis of the responses to subject access requests supported the view that central government did project an unwilling attitude.

Finally, the interviewees suggested that the model developed in the first part of this research was helpful in the discussions that formed part of the interview, but also in framing legislation and explaining it to the wider public.

6.1 Limitations

The findings of this research are subject to at least two types of limitation. First, the work is subject to the resource constraints of being the work of a single researcher, second, non-random samples were used to select organisations, interviewees, and also for the initial terms used in the literature search described in Chapter 4.

6.1.1 Resource Constraints

This research is limited by resource, because as a thesis it is necessarily the work of a single person. This reduced the available methodological options in the following ways.

First, the analysis of terms and allocation of names for categories of personal data was done by the author. The resulting classification may not be perfect and, in retrospect, the use of a group discussion or survey might have created a different categorisation and labelling. On the other hand, this was mitigated by iterations of the model following discussions with, and challenges by, supervisors. In addition, category names were taken from existing literature thus providing a wider basis for understanding.

Secondly, the selection and categorisation of organisations used for phase two of this research was done by the researcher using cultural categorisations (those in common use) and a purposive sample from the available set of organisations with which the researcher had known interactions. Organisational categorisation and comparison were not the objectives of this research, although variations were observed and examined, and the use of snowball sampling mitigated some of the possible bias inherent in purposive sampling.

Thirdly, the data collected from organisations describes a single individual and may not be representative of the data held by those organisations. It would have been preferable to collect data for a range of individuals. This implies, that in order for it to be effective, people would have been needed, from a variety of backgrounds, who were happy to trust the author with their personal data. The logistical, ethical, and trust issues of obtaining and curating strangers' personal information would have been challenging but might have produced a more robust set of results. However, it was considered impracticable and the time requirements too onerous. The weakness of the chosen approach was reduced by obtaining data from a purposively wide range of organisations.

Fourthly, the selection of interviewees was limited to those who volunteered their time. Other experts were sought from a range of organisations e.g. Information Commissioner's Office, Central and Local Government, across a range of skills. It is acknowledged that the findings from a small interview set may not be robust and would benefit from further research.

Finally, the thematic analysis was completed by the author and it could be argued that input from multiple researchers would have produced more balanced results. Whilst this was mitigated by conducting the analysis twice, it is inevitably produced through the lens of the author.

6.1.2 Non-Random Samples

Samples were used four times during this research and in none of these cases were they randomised or stratified. First, when determining the words used to search the literature for personal data terms (described in Chapter 4). In this case, the population from which to select terms was not large, therefore all the examples from the author's contemporaneous reading were used. There is a possibility that terms and meanings were omitted and that the categorisation and the model would be incomplete or the nomenclature unclear. This risk was addressed by examining publications obtained as a result of the first searches for further labels of personal data and using these in subsequent searches.

The second sample was taken from the 64,584 publications returned as search results by Google Scholar. A random sample may have unearthed some new terms or meanings but may have omitted the more influential uses of the terms. It was decided, therefore, to select a sample based on high citation relative to the year of publication for each term identified.

Third, when selecting which organisations to contact for personal data it would have been possible to contact organisations at random from Companies House, the charity commission, and lists of central and local government bodies. At the end of December 2017 there were 3,993,232 companies registered in the UK (Companies House, 2018), so it was expected that random sampling would have returned too many null returns. Instead, it was decided to sample from the 440 organisations known to have the author's personal data. Again, a random sample could have been taken but, in order to overcome the limitation of a single data subject, it was decided to use a purposive sample, taking organisations from across government, companies and NGOs and from a range of sectors within this. The effect of bias in this sample was reduced by the use of snowball sampling, using organisations named from the original sample.

Finally, interviewees were not selected at random but by opportunistic sampling. This has the disadvantage that the IT experts and high-ranking politicians were known to the author and may therefore be more likely to hold similar views. However, people with this expertise and status are difficult to recruit to interviews so personal contacts were used to secure interviews. No personal contacts were available within the Data Protection Officer or think tank executive community and so letters were written to a number of organisations and interviews arranged with those who responded. No interviewees were rejected and it was difficult to get the 9 interviewees from original target of 12. It was hoped that the four people interviewed who were not personal contacts, provided some balance, and it should

be noted that no obvious difference of opinion was noted between the interviewees known to the author and the others, (in contrast to the opinions from those inside and outside of government).

6.2 Original contribution to knowledge

There are five outputs from this research, the categorisation of personal data, modelled to illustrate their inter-relationships; the validation of the categories from real data, which support the model; the findings from the analysis of the replies to requests for data; the findings from the analysis of the process of obtaining one's own data; and finally, the findings from the content analysis of the interviews. Each of these supports the four original contributions to knowledge described below.

6.2.1 A Model and Categorisation of Personal Data

Chapter 3 identifies that the terminology used to label personal data is used inconsistently and that no move to recommend a standard nomenclature has been identified. The debate surrounding the use and access to personal data is not new (Warren and Brandeis, 1890), and even though the computerisation of personal data caused much discussion as early as the 1960's (Westin, 1967) the issue is very much alive today reflected in the implementation of the General Data Protection Regulations (European Union, 2016) in 2018. Despite these, no standard nomenclature has been agreed although regulations define personal data in what may be considered extended and complex ways. This work defines a standard nomenclature for the first time and recommends its use. It may be argued that each document implies its own definition and categorisation and this may have been common when the discipline was in its infancy, however, a standard terminology has the advantage of creating clarity in debate and dispelling many misunderstandings.

6.2.2 An analysis of organisational performance in response to requests for personal data

In order to validate the model of personal data, 82 organisations were contacted requesting copies of information describing the author that was held by them, and also information regarding the transfer of that data. For the 67 organisations in the UK those requests were presented as subject access requests under the Data Protection Act 1998. It is understood

that the analysis presented in Chapter 4 is the first time that such work has been published. It shows that although the act states that:

> 'an individual is entitled … to be informed by any data controller whether personal data of which that individual is the data subject are being processed by or on behalf of that data controller' (legislation.gov.uk, 2018a, p. Part 2 7.1)

this has not yet become common practice.

6.2.3 Analysis of the process of collecting your own personal data

In addition to analysing the performance of organisations when answering requests by an individual for the personal data held by those organisations, the process of requesting, obtaining and analysing that data was recorded and measured. Again, it is believed that no analysis of this nature has previously been published. It indicates that, at this time, an individual cannot know fully the extent of, or obtain the data that makes up, their digitally extended self.

This is important because if an individual cannot identify and challenge data or the way it is analysed there are a number of implications, some of which are further discussed below. First. access to services may be restricted, for instance when analytics make decisions using opaque algorithms, perhaps with inherent bias, or using inaccurate data. Second, data may be held in insecure areas, or sold to untrustworthy bodies and as an individual does not know where their data is held, they cannot react to security breaches to protect themselves from fraud or identity theft. Third, whilst the preferences for privacy may vary by individual, context and the information provided, there are data that individuals, in the main, want to remain private but don't realise is being transferred across organisations. This may include medical records, financial data, private behavioural practices or some other information sensitive for the individual. People expect this information to remain safely with those that they trust to curate it and may want to know where it is held and to where it may have been sent.

6.2.4 An assessment of, and reasons for, the performance of government organisations in responding to requests for information

The fourth and final contribution to knowledge, is that this research has suggested that lack of willingness and capability are the reasons for the poor performance and lack of transparency, evident in government bodies, when an individual requests details of their personal information held by those organisations. This is based on a relatively small sample of government bodies, and by interviews with only nine people. However, the comparison in levels of performance between government and non-government organisations is marked, and the level of expertise of the interviewees gives credence to their assessments. However, it may be argued that the situation has changed since this research was conducted. An example of this may be found in the Longitudinal Education Outcomes Index (Department of Education, 2018). This contains personal data from five government bodies, which may be helpful, for instance, in examining outcomes for looked after children. Whilst no evidence is available that subject access requests for this data would be unsuccessful, there are indications that this may be the case as, at the time of writing, it is restricted to internal use (government-commissioned use) and not available to external users. This non-availability quotes exemptions under the Data Protection Act 1998. It would appear that some movement is taking place but that government bodies are still unwilling to make some data available at this time.

6.3 Implications for policy and practice

The study may have a number of implications for both policy and practice.

6.3.1 Policy versus practice

This research indicates that Government needs to reflect on whether the data protection legislation is framed in a way to meet its objectives, of protecting personal data but also of allowing individuals access to their own records. This echoes an issue with legislation in general, that of practical enforcement and of it achieving its intended aims. As an ex-cabinet member with experience in personal data stated:

> ‘if you get what you think you need, will it actually be applicable, will it be usable? Will it be if you like accessible, and integrated to a point to what you thought you wanted it for, will it actually turn out to be of any value whatsoever? And quite often that question is left hanging in the air’ (Interviewee 4 00:12:38.60).

6.3.2 Stricter enforcement of data protection legislation

The results described in Chapter 4 indicate that, even given the exemptions in the Data Protection Act 1998, data is not provided when it should be. This may be improved by persuasion and coercion (perhaps when persuasion fails). Given that the Act has been in place for 20 years, it may be appropriate to examine the enforcement mechanisms. There is a view that the Office of the Information Commissioners has, to date, been easy-going in its enforcement activities. As one of the interviewees stated:

> ‘ICO they have been a very relaxed regulator, they are quite in the background, not really looking for conflict, not really looking to make people’s life difficult.’ (Interviewee 8 00:37:43.03).

In some respects that is understandable, in general people do not like to bite the hand that feeds them. The Information Commissioner’s Office is funded by two streams, data protection work is funded by notification fees paid by data controllers, and freedom of information requests are funded by the Department of Culture, Media and Sport (Information Commissioner’s Office, 2018), there would seem to be a conflict of loyalties, on the one hand to the general public and on the other hand to the paymasters. An alternative approach could be for the Information Commissioner’s Office to be funded annually from central taxation based on a formula relative to the number of data controllers and freedom of information requests submitted.

Notwithstanding the above, changes in legislation has now been enacted with the General Data Protection Regulation being introduced into UK law under the Data Protection Act 2018. This gives greater powers to the Information Commissioner’s Office and allows for higher fines for none compliance. Together with the appointment of a new Information Commissioner this may lead to a higher-level enforcement and or punishment for those not adhering to the terms of the act.

6.3.3 Laws on movement of data

People should know where their data is being sent otherwise they cannot exercise their right of access. Under current legislation the individual has no right to know where an organisation holds their data (in terms of location) or to which third party organisations the information has been sent. The law limits the movement of data to approved locations but does not ensure that individuals know to which locations, or organisations, their data has been sent. This lack of transparency does not allow individuals to make value judgements on the safety of their data. If the individual is to have access to their personal data, then they need to know where it is, in which case legislation needs to be amended to allow individuals the right to know where their data is being transferred, in terms of location and organisation. Indications are that the General Data Protection Regulation will move towards this need with Article 15 indicating that the individual has the right to know the recipient or categories of recipients with whom data has been shared. To know that data is shared with, for instance, data brokers, may cause some concern to an individual, but is unhelpful in tracking who is curating and processing an individual's personal data.

6.3.4 Information on digital personas

The legislation provides for scant information regarding digital personas. The data Protection Act 1998 allows for some rights in relation to automated decision making in Part II subsection 12 but as the law is old it does not allow for access to an individual's digital persona, where it is held, when it is used, how it is calculated, and what the implications are for the individual. The General Data Protection Regulation 2016, Article 13 (2) allows for:

> 'the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject' (European Union, 2016, p. 41).

However, this still appears to be limited in its scope as Article 22 refers to cases where decisions are based solely on automated processing which produce legal effects, but has exemption clauses, for instance if it is necessary to enter into or in the performance of a contract between the individual and the data controller.

6.3.5 Fewer exemptions

The Data Protection Act 1998 provides for a number of exemptions in Part IV e.g. regarding the mental health or condition of a data subject. The new Data Protection Act 2018 also provides for a number of exemptions e.g. for journalists, but at the time of writing has not passed into UK law. This research indicates that organisations take full advantage of the exemptions in the legislation to avoid providing data to individuals, for example the Office for National Statistics, HMRC. Indeed, the exemptions seem to be treated more seriously than the inclusions. If organisations are exempt they appear to take this as indication that they should not provide data, and do not. Where organisations are not exempt, data is not always provided. It is understandable in certain circumstances, for instance national security, that information should be withheld, but in others the justification appears less clear. The information in question is that descriptive of a single individual being provided to that same individual. There is no question of breaches in human rights in providing that data, however, in doing so transparency is increased and with that trust (Merlo et al., 2018).

6.3.6 Clear warnings for the public

The movement of data is hidden from the public view but is important in enabling individuals to know where their data is held, and for what it is being used. It may therefore be appropriate to legislate for warnings to be compulsory on web sites. For instance, 'We send your data to other organisations, see section x of the terms and conditions' (with link), or more aggressively 'We sell your data …' could be shown on all websites, publicity and application forms, if data is shared. Provided this was at the top of the website or document in a large type face, it would warn people of the possibility that their data was to be shared and provide for a more transparent approach to the handling of data. There may be objections from industry lobbying groups, and also individuals may start to ignore the warnings as they become habituated to them, however, in a similar way to the labelling of tobacco products it may raise awareness amongst the general population and encourage organisations to behave in a way that they are happy for others to know about.

6.3.7 Use of a standard vocabulary

The use of a consistent vocabulary when discussing personal data would help in an understanding of the different categories of data, what they consist of, and where they originate. The research from Chapter 5 indicates that this would be helpful for politicians when communicating with the public on the topic of personal data. A standard terminology might also bring consistency to academic debate, and finally if used in literature produced by organisations, for instance, terms and conditions, may improve understanding.

6.3.8 Issues for UK Government

This research indicates that government bodies perform poorly in comparison to the public sector and NGOs. The suggested reason is a lack of willingness to provide an individual with their personal data combined with poor systems' capability when compared with other sectors. Whilst the problems within government may be more nuanced than is suggested here, it would seem clear that a move to greater transparency, rather than using broad exemptions, combined with improved computer systems (developed with a view to increased customer focus) is needed.

6.4 Future research

The research conducted in Chapter 4 produced a snapshot of how 82 organisations responded to requests for personal data. Since the research was conducted, the General Data Protection Regulation has been agreed and will move into UK law. A follow-up study, requesting data from the same 82 organisations, would be interesting as it would provide a second set of results that could be compared with that from this research. It would provide an insight into the way the General Data Protection Regulation has affected the level and types of data provided to individuals, but also provide a view of how well organisations were complying with the new legislation. It would also form a basis for any future review.

A second area of interest concerns the model developed in Chapter 3. Whilst the model has been shown to be valid by this research the nature of data that may be ascribed to an individual may change, especially towards the edge of the model. For instance, attributes are being ascribed to an individual based on the associations that they keep (Luo et al., 2017), in all likelihood based on an algorithm which calculates a statistical probability.

However, just because person x associates with a group of people whose behaviour indicates a propensity for crime does not mean that person x is a criminal and it would be wrong to associate this attribute to x, and the same applies for less serious situations. It would therefore be of interest to examine new ways that attributes are being ascribed to individuals and matching them to the model, creating new categories of personal data as appropriate.

Finally, this research has proposed a categorisation of personal data and a model that illustrates how the categories inter-relate. It has also been suggested that this categorisation may be of help in framing legislation and explaining proposed legislation to the public. In light of this, an examination of how current laws relating to personal data, describe or categorise personal information could be instructive, and recommend changes for future laws and government communications.

6.5 Conclusion

This study has generated a model which standardises the categorisation of personal data and has shown how that data are accessed and used by companies, government and NGOs. I hope that it will help individuals, and organisations, better understand the nature of personal data and encourage organisations to be more transparent about the data that they hold, and what they do with it, whilst not discouraging future development.

Appendices

Appendices

A Analysis of methodologies for Phase 2

Methodology	Main Features	Analysis
Archival Research	Primary research where evidence is sought and extracted from original archival records.	Not applicable, the research does not involve investigations in libraries or archives rather a request for others to provide appropriate data based upon their judgement.
Case Study	Research of a case within a real life, contemporary context or setting (Ying-chun, 2009), that may be an event, process, program, or several people (Robert, 1995). The case could be the focus of attention (intrinsic case study) or the issue and the case used to illustrate the case (Robert, 1995).	Applicable - the research examines a case within the virtual world. The case is not an event, process, program or several people, but an extension of a single individual. It is an instance of a class of data that represents the digitally extended self.
Computer Simulation	A mathematical model of some natural system in physics, economics, social sciences etc. Simulation of the system is represented by running the model.	Not Applicable - there is no mathematical model involved.

Methodology	Main Features	Analysis
Content Analysis (also named Quantitative Content Analysis)	'an approach to the analysis of documents and texts … that seeks to quantify content in terms of predetermined categories and in a systematic and replicable way' (Bryman, 2008, p. 274).	Applicable - the research involves a systematic approach to obtaining and analysing documents. The data obtained was then analysed and mapped onto the components of the categorisation model in a way that should be replicable.
Critical Social Research	An analysis of social practice especially those structures which may be considered oppressive for instance, class, gender and race. The method seeks to look beyond the accepted lens to reveal underlying practices.	Not applicable - the research is not concerned with oppression as its main thrust. The responses are not seen from a class, race or gender viewpoint. In this sense the research is analytical as opposed to creating a critique of the situation.

Methodology	Main Features	Analysis
Cross Sectional Design	'collection of data on more than one case … at a single point in time order to collect a body of quantitate and quantifiable data in connection with two or more variables (usually many more than two), which are then examined to detect patterns of association' (Bryman, 2008, p. 44).	Not applicable - The digitally extended self could be viewed as a single case study, however, it is constructed of data from multiple sources. These sources are analysed and compared however at this stage in the research there is no attempt to detect any patterns of association.
Discourse Analysis	Aims at identifying characteristics of the person by analysing their discourse (written or spoken).	Not applicable - it may be used in the interpretation of documents, but the research does not look to identify socio-psychological characteristics of organisations by analysing style /content of the communications, although this would be a possibility.
Ethical Enquiry	An analysis of ethical problems especially with respect to obligation, rights etc..	Not applicable, the research does not examine the ethical issues involved in the collection and retention of personal data.

Methodology	Main Features	Analysis
Ethnographic Content Analysis	Used in the interpretation of documents, however it is a highly reflexive process with concepts emerging throughout the research as a result of the interplay between the investigator, concepts, data collection and analysis (Altheide, 1987).	Not applicable - in this research concepts were defined before the research began although analysis of issues discovered during the research will be refined during the analysis.
Ethnography	A cultural or social group, or a subset of a group are studied, primarily by observations and time spent in the field by the researcher. The ethnographer generally listens to and records the voices of informants with the intent of generating a cultural portrait. (Thomas, 1993, Wolcott, 2008).	Applicable - The research studies an individual (my digitally extended self) based not on observations but on extracted data requested from third parties, however, there is no intent to produce a cultural portrait.

Methodology	Main Features	Analysis
Grounded Theory	A substantive or context specific theory is developed that explains a phenomenon through the development and linking of categories as a result of continuous comparison of data derived from interviews. (Corbin and Strauss, 2014).	Not Applicable - The research does create an abstract analytical schema of a phenomena but not by collecting interview data. Nor does it create the theory as a result of the research but compares the research to a hypothesised model.
Hermeneutics	The theory of text interpretation, now extended to verbal and non-verbal communications. (Bryman, 2008).	Not applicable - a methodology for the interpretation of documents. This research is interested in data categories and content together with an assessment as to completeness and accuracy, rather than any meaning which may be derived from the documents.
Narrative Research	The examination of stories, narrative, or descriptions of events that explain human experiences (Pinnegar and Daynes, 2006).	Not Applicable - the research does not create or analyse narrative, nor analyse human experiences.

Methodology	Main Features	Analysis
Phenomenology	A research methodology which extracts meanings of experiences, topic or concept as commonly understood in order to reduce them to a central meaning or essence. (Moustakas, 1994).	Not Applicable - The research does not attempt to extract the essence from individuals.
Qualitative Content Analysis	Qualitative content analysis goes further than counting words or phrases to the careful examination of language in order to classify large amounts of text into categories that represent similar meanings (Weber, 1990).	Applicable - This technique has been used to analyse the case study responses. Derivatively as the classification model was predetermined, and summatively as the data collected was interpreted and values assigned to the quality of the responses (Hsieh and Shannon, 2005).
Semiotics	The study of language and sign systems that create meanings.	Not applicable - the research is not concerned with how meaning is created.

B List of organisations used by the author

123Drive	Bank of Scotland	Catering Equipment Hire
123Reg	Barclaycard	CD Universe
192com	Base.com	change.Org
247Electrical	Base40.com	CheckATrade
Abbey Life	BatteryCharged	Colwiz
Abebooks	Baumatic	Compost Direct
Above & Beyond	BBC	DynDNS
Absolute Radio	BBC Good Food	Dyson
academia.edu	Bell	Eakers Home Improvement
ACM	Belstar Electronics	Ede Ravenscroft
Acxiom	Billion UK	Electric Shop
Advertise Direct	Birmingham Midshires	Electrical Stock
AERA	Birstall	Ellis Brigham
Aeroplan	Blackspot	Empire Direct
Air New Zealand	BlahDVD	EndNote
Alexa	Blue Squirrel	EON
Alfred	Boden	EPS
Alibaba	Bonhams	eSeeds
Alpkit	Book Depository	Evernote
Amazon	Book Fellas	Eyeplan
Amenity	Bookbrain	Facebook
American Airlines	Booths	Fairport Convention
Ameritrade	Boots	Felco
Amnesty International	BorderFree	Fido
Ancestory.com	Boxers & Briefs	Fig Leaves
Ancestry	Brains Trust	Filofax
Andy Banjos	Brighton Dome	Firebox
AnyScreenProtector	Brighton Taxis	First4Group
Aphrohead	British Airways	First4Hampers
Apliances Online	British Computer Society	Flickr
Apple	British Gas	Flipboard
Applian	Broadband Player	Flowers Unlimited Brighton
Applydea	BT	Flurry Analytics
Argos	Buller	FlyBE
Asdrumark	Burton McCall	Footart
Association of Project Managers	BuyABattery	Footsteps
AudioGo	Buzan	FootTraffic
Audiotranskription	Cahoot	Free Speech Debate
Auditri	Cake Stuff	Freecom
Auto Trader	Camel Removals	FrontGate
Avaaz	Canada Helps	Frontline Club
B&Q Account	Canada Life	Garden4Less
B3ta	Canon	Gardens Cottage
BA Airmiles	Carbon Trust	Garlik
Bakery Bits	Carbonite	Gas

Baltic Air	Card Protection Plan	General Pharmacy
Bamboo	Cards Made Easy	GenieSoft
Banbury Road Medical Clinic	Care Comfort	George Justice
BangCD	Cartridge Shop	Golite
Gorringes	Joseph Turner	Music Label
Goulds Online	Just Champagne	MUST
Great Langdale Road Aces	Just Giving	My Heritage
Green Flag Recovery	Just Handles	MyTights
GreenFingers	Just Ink & Paper	Names & Tapes Direct
GriSoft	Just MOTs	Napster
Grupo Santader	Kaupthingedge	Nationwide Building Society
Guardian	Keens Shoes	National Savings &Investments
GuidoFawkes	Kingston Village Shop	Natures Healtbox
Handles4Doors	Kitchen Doorhandles	Natwest Brokerline
HandleWorld	Klnowles Nets	NCH Software
Harper Collins	Knobs & Knockers	Nectar
Harrod Gardening	Koingo Software	Netgear
Healthy Supplies	Laithwaites	New Scientist
Heatmiser	Lakeland	NHS
Hedgemaster	Landa Tec	Nightingale
Heinnie	Landmark Trust	Nike
Heirloom	Lands End	Norton
Hertz	LCH	Notonthehighstreet.com
Highview Salt	Leekes	Novatech
Historic Newspapers	Legal & General	nPower
Hob UK	Lewis Registry Office	NSI
Homeland Security	Lightbulb Company	Nuance
Homes & Property Shop	Lighting Matters	O2
Homevac Electronics	LiGO Electronics	Oddbins
Hoodless Brennan	LinkedIn	OdSox
Hot Wax Honies	Linksys	Office of National Statistics
House Insurance	Literature & Latte	Omega Music
Houses	Little Machine	OmniFocus
HP	Lloyds TSB	One Voice
Human Rights Watch	Locker Room	Open Rights Group
Humanist Society	Logitech	Opodo
Humyo	Lonely Planet	Orbicule - Witness
Identity & Passport Agency	Lost Cousins	Orvis
Identity & Passport Service	Lowri Beck	OSoClean
Igluski	LTA	Outdoor Warehouse
IKEA	LTSB Registrars	Oxford City Council
iMUST	Mac Upgrades	Oyster
Infabode	MacPaw	Palestinian Solidarity Campaign
ING	Manchester United	Paperstone
Ink Emporium	Marks & Spencer	Paragon Software
Inland Revenue	McAfee	Parallels Desktop
InstaCloud	Memory Card Zoo	Pastorino
Institute of Electrical Engineers	Mendeley	Path

Interparcel	Menkind	Paypal
Ironmongery Direct	Michael Bourne	PC World Business
Jelly Vision	Microsoft	Peace Now
John Lewis	Midlands Memorabilia	Peacemaker
John Lewis Finance	Mobile Fun	Pearson Ed Books
Pension Service (State Pension)	Sebo	The Yatch Shop
Personalised Birthday Cards	Secular Society	Things
Pest Control Shop	Selftrade	Ticketmaster - Spain
Pete Bland Sports	SheilaMaid	Ticketmaster - UK
Petrol Prices	Shoes	TMF
Pinterest	Shop4Tools	TomTom
PlantMeNow	Siemens	Tool Station
Play	Silicone Molds	ToolLine
Poles & Blinds	Simple Note	Total Gardens
Police (National Computer)	Simply Doorhandles	Toyota
Portman Building Society	SimplyMoleskine	Trade Handles
Portsmouth Magistrates Court	SJBDirect	Travel Insurance
Posturite	Sky	Tredz
Power Tools Pro	Skype	Tri Sports
Premier Electronics	Sling Player	TripWolf
PriceMinster	Snapfish	TSO
ProCameraSales	Solutions Inc	TuneUp
Prosoft	Song Bird	TV Licensing
Provide.co.uk	Song Kick	Twiki
Public	Sonos	Twitter
Pure	Sony	Uber
QSR Nivo	Soulmates	UbiSoft
Railcards Online	Sound & Vision	UK Border Agency
Rated People	Southern Water	Ultimate Guitar
RBS World Pay	Speed Awareness	Vehicle Licensing Authority
Research Gate	Spotify	Velux
RHS	Staples	Viagogo
Rightmove	Starbucks	Videre
RingGo	Stay Private	Virgin Media
Ripcaster	Steam	Virgin Trains
RoadAngel	Super Duper	Vodafone
Rohan	Super-fi	WAE
Rossums	Surrey Cricket Club	Waitrose
Royal Horticultural Society	Survey Monkey	Warwick Arts Centre
Royal Mail	Sussex University	Water Companies
RSPB	Suttons Seeds	Watersons
Rubbersole	Symantic	Wayfair
S&N Genealogy Supplies	Synology	Weather Shop
SafeLincs	T and S Architectural	Wembley Stadium
Sage Pay	Taps & Sinks Direct	Western Digital
Sale Shops	Target Neutral	WH Smith
Samsung	tBKS	Wiggle
SatMap	TDNet	Wikicfp

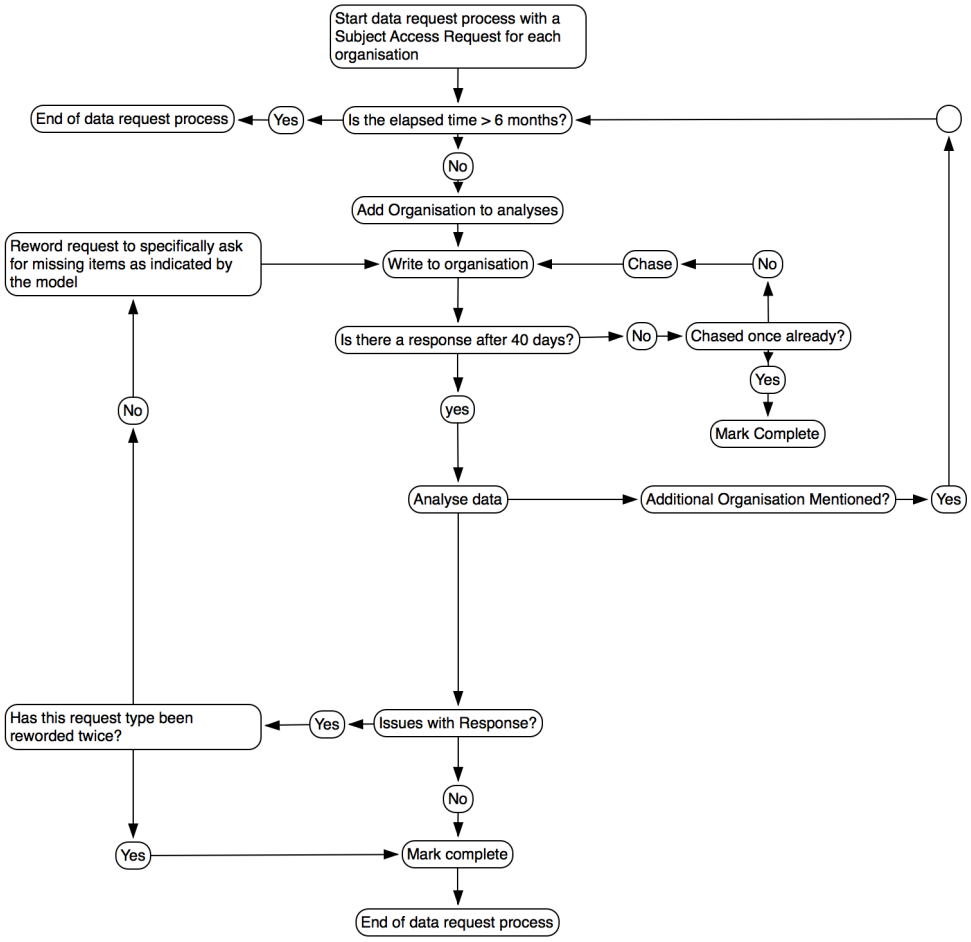
Saunderson Security	Telephone Preference Service	WikiSpaces
Scansoft	Tesco	Wired
Schedule World	The Handle Studio	Wondershare
Scots Plants	The Hut	Wordpress
Scotts	The MP3 Company	World of Mowers
Scribd	The Original Gift Company	World Pay
		Yahoo
		Yeo Valley
		Year of Rock
		Zen
		Zotero
		Zurich Life
		Zyma

C Purposive sample of organisations and their categories

Category	Sector	Organisation
Central Government	Central Government	UKBA
		HM Revenue and Customs
		NHS
		ONS
Local Government	Local Government	Oxford City Council
		SDLC
NGO	Charity	Amnesty
		One Voice
		Open Rights Group
		RSPB
Private Company	Marketing Information	Dunnhumby
		Flurry
	Online Shopping	Boden
		Cult Pens
	Utilities	Coop Energy
Public Company	Banking	John Lewis Partnership
		Lloyds Bank
	Credit Ref	Equifax
		Experian
	Insurance	Zurich Life
	Internet Search	Google

	Marketing Information	Acxiom
	Online & High Street	John Lewis
		M&S
	Online Shopping	Amazon
		Apple
	Social Network	Facebook
		Twitter
	Supermarket	Tesco
		Waitrose
	Utilities	United Utilities
		Vodafone

D Process for data collection



E Sample letters

E1 Internal to the UK

71 Bainton Road
Oxford
OX2 7AG
07767222720
mail@brian.parkinson.name
<date>

<address>

Dear Sir or Madam

Subject Access Request

Brian Laurence Parkinson,

Date of Birth 6th November 1948

Current Addresses 1 The Anvil, The Street, Kingston, Lewes BN7 3PB
2 71 Bainton Road, Oxford, OX2 7AG
3 Lane Ends Barn, Elterwater, Ambleside, Cumbria LA22 9HN

Telephone Numbers 01272 473727
07767 222720
01856 434241
015394 37298

I am a student undertaking a doctorate at Southampton University researching into my own electronic records. I would be grateful if you would supply the information about me that I am entitled to under the Data Protection Act 1998, which is held electronically, relating to:

1. Electronic records which are descriptive of me e.g. name, address, age, transactions, analytic profile, and the purpose for which they are held.
2. Where the data originated, whether from my own actions or elsewhere. If elsewhere, which organisations (name and address of organisation) or individuals (anonymised if necessary) provided the information.
3. Electronic data descriptive of me disclosed to other parties, or which may be disclosed to other parties, whether basic data or the results of analytic profiling. The identity of the other parties (name and address of organisation), and for what purpose.
4. Information used as input to any analytic profiles, which could be used to describe me, and where it came from (name and address of organisation).
5. The names and purposes of any analytic profiles created which could be used to describe me?
6. The country within which each group of data records and analytic profiles are held.

I have attached copies of my driving licence, passport, and a recent utility bill, in order to identify myself, together with a cheque for £10. If you need any more information from me please let me know as soon as possible.

It may be helpful for you to know that a request for information under the Data Protection Act 1998 should be responded to within 40 days.

If you do not normally deal with these requests, please pass this letter to your Data Protection Officer. If you need advice on dealing with this request, the Information Commissioner's Office can assist you and can be contacted on 0303 123 1113 or at www.ico.org.uk

Yours faithfully

Brian Parkinson

Attachments

Cheque for £10.00

Copy of Passport

Copy of Driving Licence

Copy of Cooperative Energy Utility Statement

E2 External to the UK

71 Bainton Road
Oxford
OX2 7AG
07767222720
mail@brian.parkinson.name
<date>

<organisation>

Dear Sir or Madam,

Subject Access Request

Brian Laurence Parkinson,

Date of Birth **6th November 1948**

Current Addresses **1) The Anvil, The Street, Kingston, Lewes BN7 3PB**
2) 71 Bainton Road, Oxford, OX2 7AG
3) Lane Ends Barn, Elterwater, Ambleside, Cumbria LA22 9HN

Telephone Numbers **01272 473727**
07767 222720
01856 434241
015394 37298

Account eMail:mail@brian.parkinson.name

MAC Address **DO:23:DB:1F:EB:AB**

UDID **ff83b3fd4cfc29eea1b89dabd280357b0acce257**

Thanks for creating <*product*> which is used regularly in the UK.

I am a student undertaking a doctorate at Southampton University

researching into my own electronic records. I would be grateful if you would supply the information about me that I am entitled to under the Data Protection Act 1998, which is held electronically, relating to:

1. **Electronic records which are descriptive of me e.g. name, address, age, transactions, analytic profile, and the purpose for which they are held.**
2. **Where the data originated, whether from my own actions or elsewhere. If elsewhere, which organisations (name and address of organisation) or individuals (anonymised if necessary) provided the information.**
3. **Electronic data descriptive of me disclosed to other parties, or which may be disclosed to other parties, whether basic data or the results of analytic profiling. The identity of the other parties (name and address of organisation), and for what purpose.**
4. **Information used as input to any analytic profiles, which could be used to describe me, and where it came from (name and address of organisation).**
5. **The names and purposes of any analytic profiles created which could be used to describe me?**
6. **The country within which each group of data records and analytic profiles are held.**

Yours faithfully

Brian Parkinson

F Example log for each organisation

Year 2 Data Collection Activity Log - **Acxiom**

Date	**Activity**	**Time in Minutes**
6/12/2013	Obtain Data Protection Register - Entry Details	5
6/12/2013	Create Letter and Attachments	17
9/12/2013	Post Letter (12 letters posted time taken 46 mins)	4
11/12/2013	Checked & logged letter received	4
1/3/2014	Analysed data	46
1/3/2014	Follow Up Letter	27
2/3/2014	Post follow up letter (2 letters)	15
20/3/2014	Received Follow Up Letter	3
27/3/2014	Analysed follow up data	26
	Total Hours	2.45

G Log of costs incurred and time spent

Organisation	Category	Sector	Purposive Sample (1) Snowball Sample (2) Too Late for Inclusion (3)	Not in List of Known Organisations (1)	Costs SAR (£)	Costs Post (£)	Total Cost (£)	Total Time (hours)
365Scores	Private Company	Software App Development	2	1	n/a	5.28	5.28	0.83
Acxiom	Public Company	Marketing	1	0	10	5.28	15.28	2.45
Amazon	Public Company	Online Shopping	1	0	ret	5.28	5.28	3.25
Amnesty	NGO	Charity	1	0	10	5.28	15.28	2.9
Ancestry	Public Company	Genealogy	2	0	10	2.64	12.64	1.03
Apple	Public Company	Online Shopping	1	0	ret	9.10	9.1	3.60
Barclaycard	Public Company	Finance	3	0	n/a	n/a	n/a	n/a
Bloom Built (Day One)	Private Company	Software App Development	2	1	n/a	0.00	0	0.93
Boden	Private Company	Online Shopping	1	0	ret	2.64	2.64	1.78

Organisation	Category	Sector	Purposive Sample (1) Snowball Sample (2) Too Late for Inclusion (3)	Not in List of Known Organisations (1)	Costs SAR (£)	Costs Post (£)	Total Cost (£)	Total Time (hours)
Brockbank Syndicate Management	Public Company	Finance	3	1	n/a	n/a	n/a	n/a
Bryan Mitchell (Geared)	Private Company	Software App Development	2	1	10	0.00	10	0.92
Cahoot	Public Company	Finance	3	0	n/a	n/a	n/a	n/a
Call Credit	Public Company	Credit Reference	2	1	10	2.64	12.64	1.65
Charles Tyrwhitt	Private Company	Online & High Street Shopping	2	1	10	2.64	12.64	0.52
CIFAS	Private Company	Credit Reference	2	1	10	2.64	12.64	0.75
Codegent (Learn Japanese)	Private Company	Software App Development	2	1	10	2.64	12.64	0.85
Conde Nast	Private Company	Magazine Distribution	2	1	10	2.64	12.64	0.75

Organisation	**Category**	**Sector**	**Purposive Sample (1) Snowball Sample (2) Too Late for Inclusion (3)**	**Not in List of Known Organisations (1)**	**Costs SAR (£)**	**Costs Post (£)**	**Total Cost (£)**	**Total Time (hours)**
Coop Energy	Private Company	Utilities	1	0	10	4.36	14.36	2.18
Critical Hit Software (Jigsaw Puzzle)	Private Company	Software App Development	2	1	n/a	0.00	0	0.52
Cult Pens	Private Company	Online Shopping	1	0	ret	2.64	2.64	1.28
Dunnhumby	Private Company	Marketing	1	0	ret	5.28	5.28	3.23
Engaging Networks	Private Company	Charity Fundraising	2	1	10	1.72	11.72	0.90
Equifax	Public Company	Credit Reference	1	0	10	5.28	15.28	2.78
Equiniti	Private Company	Finance	3	1	n/a	n/a	n/a	n/a
Eventbrite	Public Company	Event Booking	2	1	n/a	6.28	6.28	0.62
Experian	Public Company	Credit Reference	1	0	10	5.28	15.28	4.22
Facebook	Public Company	Social Media	1	0	10	16.94	26.94	3.72

Organisation	**Category**	**Sector**	**Purposive Sample (1) Snowball Sample (2) Too Late for Inclusion (3)**	**Not in List of Known Organisations (1)**	**Costs SAR (£)**	**Costs Post (£)**	**Total Cost (£)**	**Total Time (hours)**
Flurry	Public Company	Marketing	1	0	10	2.64	12.64	2.67
Frogmind (Badland)	Private Company	Software App Development	2	1	n/a	0.00	0	0.53
Google	Public Company	Internet Search	1	0	10	2.64	12.64	5.17
GR8iPhoneGames TLC Productions (Road Warrior)	Private Company	Software App Development	2	1	n/a	0.0	0	0.72
GZeroLtd (TVCatchup)	Private Company	Software App Development	2	1	10	2.64	12.64	0.57
H2O	Private Company	Utilities	2	1	ret	2.64	2.64	0.88
Halifax Bank of Scotland	Public Company	Finance	3	1	n/a	n/a	n/a	n/a
HM Revenue & Customs	Central Government	Central Government	1	0	ret	5.28	5.28	3.68

Organisation	Category	Sector	Purposive Sample (1) Snowball Sample (2) Too Late for Inclusion (3)	Not in List of Known Organisations (1)	Costs SAR (£)	Costs Post (£)	Total Cost (£)	Total Time (hours)
Instagiv	Private Company	Charity Fundraising	2	1	10	2.64	12.64	0.98
John Lewis	Public Company	Online & High Street Shopping	1	0	10	5.28	15.28	4.25
John Lewis Credit Card	Public Company	Finance	1	0	ret	5.28	5.28	3.75
Joseph Turner	Private Company	Online Shopping	2	0	10	2.64	12.64	0.73
Laithwaites	Private Company	Online & High Street Shopping	2	1	10	2.64	12.64	0.97
Lands End	Public Company	Online & High Street Shopping	2	0	10	1.72	11.72	0.85
Lloyds Bank	Public Company	Finance	1	0	10	5.28	15.28	6.33
Lloyds Bank Pension	Public Company	Finance	2	1	10	5.28	15.28	1.45

Organisation	**Category**	**Sector**	**Purposive Sample (1) Snowball Sample (2) Too Late for Inclusion (3)**	**Not in List of Known Organisations (1)**	**Costs SAR (£)**	**Costs Post (£)**	**Total Cost (£)**	**Total Time (hours)**
M&S	Public Company	Online & High Street Shopping	1	0	10	2.64	12.64	0.88
Mail Chimp	Private Company	Marketing	2	1	n/a	0.0	0	2.55
Mastercard	Private Company	Finance	2	1	10	2.64	12.64	0.73
MBNA	Private Company	Finance	3	1	n/a	n/a	n/a	n/a
MCL Software	Private Company	Finance	2	1	ret	2.64	2.64	0.57
Met Office (Weather App)	Central Government	Central Government	2	1	ret	1.72	1.72	0.77
Mobiata (FlightTrack)	Private Company	Software App Development	2	1	n/a	0.00	0	0.58
Mobile Info Center (MacHash)	Private Company	Software App Development	2	1	n/a	0.00	0	0.55

Organisation	**Category**	**Sector**	**Purposive Sample (1) Snowball Sample (2) Too Late for Inclusion (3)**	**Not in List of Known Organisations (1)**	**Costs SAR (£)**	**Costs Post (£)**	**Total Cost (£)**	**Total Time (hours)**
MobilityWare (Free Solitaire)	Private Company	Software App Development	2	1	n/a	0.00	0	0.38
Natwest Bank	Public Company	Finance	2	1	10	2.64	12.64	1.00
NHS	Central Government	Central Government	1	0	10	2.64	12.64	1.22
Not on the High Street	Private Company	Online Shopping	2	0	10	1.72	11.72	1.00
One Voice	NGO	Charity	1	0	10	2.64	12.64	0.88
ONS	Central Government	Central Government	1	0	10	5.26	15.26	1.98
Open Rights Group	NGO	Charity	1	0	10	2.64	12.64	1.57
Orvis	Private Company	Online & High Street Shopping	2	0	10	2.64	12.64	0.87

Organisation	**Category**	**Sector**	**Purposive Sample (1) Snowball Sample (2) Too Late for Inclusion (3)**	**Not in List of Known Organisations (1)**	**Costs SAR (£)**	**Costs Post (£)**	**Total Cost (£)**	**Total Time (hours)**
Oxford City Council	Local Government	Local Government	1	0	10	2.64	12.64	2.92
Oxford University Press Pension	NGO	Charity	2	1	10	0.00	10	0.87
Parcel Force (Royal Mail)	Public Company	Postal Services	2	0	10	5.28	15.28	1.28
Parseq Fulfilment House	Private Company	Finance	2	1	10	2.64	12.64	0.80
Personal Telephone Fundraising	Private Company	Charity Fundraising	2	1	ret	2.64	2.64	0.70
Play Ltd	Private Company	Online Shopping	3	0	n/a	n/a	n/a	n/a
Prolog	Private Company	Finance	2	1	ret	2.64	2.64	1.03
Pure 360	Private Company	Data Processing	2	1	ret	2.64	2.64	0.72

Organisation	Category	Sector	Purposive Sample (1) Snowball Sample (2) Too Late for Inclusion (3)	Not in List of Known Organisations (1)	Costs SAR (£)	Costs Post (£)	Total Cost (£)	Total Time (hours)
Rapidata	Public Company	Data Processing	2	1	10	1.72	11.72	0.78
Readdle	Public Company	Software App Development	2	1	n/a	0.00	0	0.50
Refugee Council	NGO	Charity	2	1	n/a	0.00	0	0.97
Rogavi (AIUK Raffle)	Private Company (in administration)	Marketing	2	1	ret	5.28	5.28	1.10
RSPB	NGO	Charity	1	0	10	2.64	12.64	1.83
SDLC	Local Government	Local Government	1	0	10	5.28	15.28	2.10
Sea Containers Pension	Public Company	Finance	2	0	10	1.72	11.72	0.68
SN&CK Media Ltd	Private Company	Software App Development	2	1	10	2.64	12.64	0.62

Organisation	**Category**	**Sector**	**Purposive Sample (1) Snowball Sample (2) Too Late for Inclusion (3)**	**Not in List of Known Organisations (1)**	**Costs SAR (£)**	**Costs Post (£)**	**Total Cost (£)**	**Total Time (hours)**
Spotify Ltd	Private Company	Entertainment	2	0	10	2.64	12.64	0.48
Suttons Seeds	Private Company	Online Shopping	2	0	ret	1.72	1.72	0.78
Synetics Solutions Inc.	Private Company	Finance	2	1	10	2.64	12.64	0.73
Taylored Mortgage & Investment	Private Company	Finance	2	1	10	2.64	12.64	1.32
Tesco	Public Company	Supermarket	1	0	10	4.36	14.36	2.82
thetrainline	Private Company	Software App Development	2	1	10	2.64	12.64	1.42
Trustpilot	Public Company	Software App Development	2	1	n/a	7.16	7.16	1.80
Twitter	Public Company	Social Media	1	0	10	5.28	15.28	4.38

Organisation	**Category**	**Sector**	**Purposive Sample (1) Snowball Sample (2) Too Late for Inclusion (3)**	**Not in List of Known Organisations (1)**	**Costs SAR (£)**	**Costs Post (£)**	**Total Cost (£)**	**Total Time (hours)**
UKBA	Central Government	Central Government	1	0	10	5.28	15.28	2.38
United Utilities	Public Company	Utilities	1	0	10	5.28	15.28	2.23
Unlock Democracy	Private Company	Campaigning	2	1	10	2.64	12.64	1.40
Virgin Media	Private Company	Entertainment	3	0	n/a	n/a	n/a	n/a
Vodafone	Public Company	Utilities	1	0	ret	5.28	5.28	4.27
Waitrose	Public Company	Supermarket	1	0	10	2.64	12.64	0.85
Xiao Yixiang (Pro Metronome)	Private Company	Software App Development	2	1	n/a	6.28	6.28	0.80
Zurich Life	Public Company	Finance	1	0	10	5.28	15.28	3.55
Totals for Purposive Sample			32		290	118.88	408.88	45.73

Organisation	**Category**	**Sector**	**Purposive Sample (1) Snowball Sample (2) Too Late for Inclusion (3)**	**Not in List of Known Organisations (1)**	**Costs SAR (£)**	**Costs Post (£)**	**Total Cost (£)**	**Total Time (hours)**
Totals for Snowball Sample			59		240	153.54	393.54	91.10
Grand Total			91	46	530	272.42	802.42	136.83
Mean for Purposive Sample					9.06	3.72	12.78	1.43
Mean for Snowball Sample					4.07	2.60	6.67	1.54

H Log of timings and responses

Organisation	Category	Sector	DPA Act Request	Request Delivered	Request Confirmed	Additional Info Requested	Additional Info Provided	Chased	Response Received	Elapsed Days Taken	Follow Up Sent	Follow up Delivered	Follow Up Response Received	Elapsed Days Taken
Amazon	Public Co	Online Shopping	19/11/2013	20/11/2013	-			-	26/11/2013	6d	9/11/13	11/12/2013	11/1/14	31d
Facebook	Public Co	Social Network	26/11/2013	26/11/13	26/11/2013				26/11/2013	0ms	27/3/14	2/4/14	9/5/14	37d
Apple	Public Co	Online Shopping	6/12/13	16/12/2013	17/12/2013				19/12/2013	3d	25/3/14	27/3/14	8/4/14	12d
NHS	Central Govt		19/11/2013	20/11/2013					20/12/13	30d	X			X
One Voice	NGO	Charity	11/12/2013	12/12/2013					20/12/2013	8d	X			X
Cult Pens	Private Co	Online Shopping	9/12/13	12/12/2013	n/a	n/a	n/a	n/a	20/12/2013	8d	X			X
RSPB	NGO	Charity	18/11/2013	20/11/2013	22/11/13				23/12/13	33d	X			X
Oxford City Council	Local Govt	Local Govt	19/11/2013	20/11/2013	27/11/13				27/12/2013	37d	17/1/14	17/1/14	21/1/14	4d
Dunnhumby	Private Co	Marketing Info	9/12/13	12/12/2013	16/12/2013	16/12/2013	24/12/2013		6/1/14	25d	31/1/14	4/2/14	19/2/14	15d
UKBA	Central Govt	Central Govt	12/12/2013	16/12/2013					6/1/14	21d	17/1/14	21/1/14	24/2/14	34d
Tesco	Public Co	Supermarket	12/12/2013	16/12/2013	19/11/2013	19/11/2013	23/12/2013		8/1/14	23d	31/1/14			
Boden	Private Co	Online Shopping	9/12/13	11/12/2013					11/1/14	31d	X			X
United Utilities	Public Co	Utilities	12/12/2013	16/11/2013	19/12/2013				14/1/14	59d	28/2/14	5/3/14	19/3/14	14d

Organisation	Category	Sector	DPA Act Request	Request Delivered	Request Confirmed	Additional Info Requested	Additional Info Provided	Chased	Response Received	Elapsed Days Taken	Follow Up Sent	Follow up Delivered	Follow Up Response Received	Elapsed Days Taken
Google	Public Co	Internet Search	28/11/2013	29/11/2013					14/1/14	46d	26/3/14	26/3/14		
HM Revenue	Central Govt		9/12/13	11/12/2013	19/12/2013				17/1/14	37d	25/2/14	26/2/14		
Equifax	Public Co	Credit Ref	9/12/13	11/12/2013					17/1/14	37d	26/3/14	28/3/14	24/4/14	27d
Acxiom	Public Co	Marketing Info	9/12/13	11/12/2013					18/1/14	38d	1/3/14	4/3/14	20/3/14	16d
Vodafone	Public Co	Utilities	12/12/2013	13/12/2013					18/1/14	36d	21/2/14		4/4/14	4/4/14
ONS	Central Govt		11/12/2013	19/12/2013	19/12/2013				18/1/14	30d	1/3/14	4/3/14	31/3/14	27d
Zurich Life	Public Co	Insurance	12/12/2013	17/12/2013					20/1/14	34d	4/3/14	7/3/14	1/4/14	25d
M&S	Public Co	Online & High Street	11/12/2013	12/12/2013	19/12/2013				23/1/14	42d	X			X
John Lewis	Public Co	Online & High Street	9/12/13	11/12/2013					25/1/14	45d	25/3/14	26/3/14		
John Lewis Partnership	Public Co	Banking	9/12/13	12/12/2013	15/1/14				25/1/14	44d	26/3/14	28/3/14	7/6/14	71d
Waitrose	Public Co	Supermarket	12/12/2013	13/12/2013					25/1/14	43d	25/3/14			0
Lloyds Bank	Public Co	Banking	18/11/2013	20/11/2013					29/1/14	70d	25/3/14	27/3/14	16/4/14	20d
Amnesty	NGO	Charity	6/12/13	11/12/2013					31/1/14	51d	3/3/14	10/3/14	22/3/14	12d
SDLC	Local Govt	Local Govt	12/12/2013	15/12/2013					31/1/14	47d	3/3/14	6/3/14		

Organisation	Category	Sector	DPA Act Request	Request Delivered	Request Confirmed	Additional Info Requested	Additional Info Provided	Chased	Response Received	Elapsed Days Taken	Follow Up Sent	Follow up Delivered	Follow Up Response Received	Elapsed Days Taken
Experian	Public Co	Credit Ref	9/12/13	13/12/2013	20/12/2013	20/12/2013	30/12/2013		9/2/14	58d	21/2/14	13/3/14	14/5/14	62d
Flurry	Private Co?	Marketing Info	9/12/13	11/11/2013		17/1/14	24/1/14		4/3/14	113d	6/3/14	6/3/14	7/3/14	1d
Twitter	Public Company	Social Network	26/11/2013	26/11/2013				19/3/14	24/3/14	118d	24/3/14			?
Open Rights Group	NGO	Charity	11/12/2013	12/12/2013	15/1/14	24/1/14	25/1/14	19/3/14	25/3/14	103d	X			X
Coop Energy	Private Co	Utilities	19/11/2013	20/11/2013				19/3/14	X		X			x
Total	32		32						31	41d	24	20	17	70

Secondary Organisation	Category	Sector	DPA Act Request	Request Delivered	Request Confirmed	Additional Info Requested	Additional Info Provided	Chased	Response Received	Elapsed Days Taken	Follow Up Sent	Follow up Received	Follow Up Response	Elapsed Days Taken
365Scores	Private Company	iPhone Sports Info	2/4/14	2/4/2014					2/4/14	0ms	X			X
Ancestry	Public Company	Genealogy	14/3/14	19/3/14					X	X	X			X
Barclaycard			X							X	X			X
Brockbank Syndicate Management			X							X	X			X
Bryan Mitchell (Geared)			3/4/14	3/4/14					3/4/14	0d	X			X
Cahoot			X							X	X			X
Call Credit			31/3/14	2/4/14					9/5/14	37d	X			X
Charles Tyrwhitt			31/3/14	2/4/14					X	X	X			X
CIFAS			31/3/14	2/4/14		4/4/14	8/4/14		16/4/14	14d	X			X
Codegent (Learn Japanese)			4/4/14	8/4/14					8/4/14	0d	X			X
Conde Nast Digital (Epicurious Recipes)			5/4/14	8/4/14					13/5/14	35d	X			X
Critical Hit Software (Jigsaw Puzzle)	Private Company	iPhone Gaming	7/3/14	07/03/2014					X	X	X			X
Day One		Journaling App	3/4/14	3/4/14	3/4/14	14/4/14	14/4/14		20/4/14	17d	X			X
Engaging Networks			1/4/14	03/04/2014					7/4/14	4d	X			X
Equiniti			X							X	X			X
Eventbrite			4/4/14	08/04/2014	8/4/14				X	X	X			X

Secondary Organisation	Category	Sector	DPA Act Request	Request Delivered	Request Confirmed	Additional Info Requested	Additional Info Provided	Chased	Response Received	Elapsed Days Taken	Follow Up Sent	Follow up Received	Follow Up Response	Elapsed Days Taken
Frogmind	Private Company	Software	14/3/14	14/03/2014					X	X	X			X
GR8iPhoneGames TLC Productions (Road Warrior)	Private Company	iPhone Gaming	7/3/14	7/3/14					7/3/14	0ms	11/4/14		11/4/14	0d
GZeroLtd (TVCatchup)	Private Company	IOS Apps TV	2/4/14	8/4/14					X	X	X			X
H2O	Public Company	Utilities Infrastructure	28/2/14	4/3/14					12/3/14	8d	X			X
Halifax Bank of Scotland			X							X	X			X
Instagiv	Private Company	SMS Charity Funding	3/3/14	5/3/14					13/3/14	8d	X			X
Joseph Turner			31/3/14	2/4/14					X	X	X			X
Laithwaites	Private Company	Wine Supplier	31/3/14	2/4/14		3/4/14			4/4/14	2d	X			X
Lands End			1/4/14	3/4/2014					9/5/14	36d	X			X
Lloyds Bank Pension			1/4/14	2/4/14		16/4/14	9/9/14		17/9/14	168d	9/10/14			
Mail Chimp	Private Company	Internet Services	13/1/14	13/1/14		3/4/14			3/4/14	80d	27/3/14		12/5/14	46d
Mastercard	Public Company	Financial Institution	2/4/14	8/4/14					8/5/14	30d	X			X
MBNA			X							X	X			X
MCL Software			31/3/14	2/4/14					X	X	X			X
Met Office (Weather App)			1/4/14	3/4/14		9/4/14	9/4/14		9/5/14	36d	X			X
Mobiata (FlightTrack)			5/4/14	14/4/14					X	X	X			X

Secondary Organisation	Category	Sector	DPA Act Request	Request Delivered	Request Confirmed	Additional Info Requested	Additional Info Provided	Chased	Response Received	Elapsed Days Taken	Follow Up Sent	Follow up Received	Follow Up Response	Elapsed Days Taken
MobileInfoCenter (MacHash)			3/4/14	3/4/2014					X	X	X			X
MobilityWare (Free Solitaire)	Private Company	iPhone Gaming	14/3/14	14/3/2014					X	X	X			X
Natwest	Public Company	Financial Institution	3/3/14	5/3/2014					2/5/14	58d	X			X
Not on the High Street			1/4/14	2/4/2014					8/5/14	36d	X			X
Orvis			31/3/14	2/4/2014					8/5/14	36d	X			X
Oxford University Press Pension			1/4/14	1/4/2014					9/5/14	38d	X			X
Parcel Force (Royal Mail)	Private Company	Postal Services	2/4/14	8/4/14		13/5/14			18/6/14	36d	X			X
Parseq Fulfilment House	Private Company	Fulfillment Services	3/3/14	5/3/14					X	X	X			X
Personal Telephone Fundraising	Private Company	Telephone Charity Fund Raising	3/3/14	5/3/14					18/3/14	13d	X			X
Play Ltd			X							X	X			X
Prolog	Private Company	Online Shopping	22/1/14	23/1/14					25/2/14	33d	X			X
Pure 360	Private Company	email campaigns	3/3/14	5/3/14					10/3/14	5d	X			X
Rapidata			1/4/14	2/4/14					X	X	X			X
Readdle		Productivity App	4/4/14	4/4/14					X	X	X			X
Refugee Council			06/04/2014	07/04/2014					29/4/14	22d	X			X

Secondary Organisation	Category	Sector	DPA Act Request	Request Delivered	Request Confirmed	Additional Info Requested	Additional Info Provided	Chased	Response Received	Elapsed Days Taken	Follow Up Sent	Follow up Received	Follow Up Response	Elapsed Days Taken
Rogavi (AIUK Raffle)	Private Company	Raffles - in administration 24/12/2014	3/3/14	Returned				09/10/2014	X	X	X			X
Sea Containers Pension			1/4/14	2/4/14					15/5/14	43d	X			X
Sn&ck Media Ltd (Football Rumours)			4/4/14	8/4/14					X	X	X			X
Spotify Ltd	Public Company	Streaming music servicer	7/3/14	11/3/2014					X	X	X			X
Suttons Seeds			1/4/14	03/04/2014					9/4/14	6d	10/9/14			
SyneticsSolutions Inc.			31/3/14	2/4/14					9/4/14	7d	X			X
Taylored Mortgage & Investment	Private Company	Financial Advisors	6/3/14	10/3/14					11/3/14	1d	X			X
thetrainline	Private Company	rail ticket sales	2/4/14	8/4/14					15/5/14	37d	X			X
Trustpilot	Private Company	Internet Services	13/1/14	15/1/14					28/1/14	13d	3/3/14		X	X
Unlocked Democracy			6/4/14	7/4/14					9/5/14	32d	X			X
Virgin Media			X							X	X			X
Xiao Yixiang (Pro Metronome)	Private Company	IOS Apps - Music	2/4/14	20/04/2014					X	X	X			X
59		SAR Sent	51						33	F U Sent	5		2	O/S

Appendix H Log of Timings and Responses

Summary

Number of Organisations in the Purposive Sample	32
Number of Organisations in the Purposive Sample Contacted	32
Number of Responses Obtained from First Requests	31
Number of Organisations who did not respond	1
Mean Response time	41d
Maximum response time (Twitter)	118d
Minimum response time (Facebook)	0w
Follow Up Communications Sent	24
Follow Up Answers Received	18
Number of Organisations who did not respond	6
Mean Time for Follow Up Responses	29d
Maximum Time for Follow Up Responses (John Lewis Partnership)	71d
Minimum Time for Follow Up Response (Flurry)	1d

Number of Organisations in the Snowball Sample	59
Number of Organisation in the Snowball Sample Contacted	51
Number of Responses Obtained from First Requests	33
Number of Organisations who did not respond	18
Mean Response time	25d
Maximum response time (Lloyds Bank Pension)	168d
Minimum response time (Several App Developers contacted by email)	0w
Follow Up Communications Sent	5
Follow Up Answers Received	2
Number of Organisations who did not respond	3
Mean Time for Follow Up Responses	23d
Maximum Time for Follow Up Responses (Mail Chimp)	46d
Minimum Time for Follow Up Response (GR8iPhoneGames)	0w

I Journal of responses – sample entry

Open Rights Group - 1

Response - outside period (103 elapsed days)

Digital Footprints
- Responses to emails
- conference enrolment

3rd party footprints
- entry of enrolment data and direct debits

Digital Persona
- no analytics undertaken excepting overall statistics for the organisation

External Data From
- Rapidata (third party direct debit provider)
- Engaging Networks t/a Eactivist (third party communications campaign provider)
- Eventbrite (event registration platform) - USA company based in Califirnia - data held shown by ORG

Location of Data
- Eventbrite USA
- Rapidity ?
- EActivist ?

External Data From
- Eventbrite (event registration platform) - USA company based in Califirnia - data held shown by ORG

External Data To
- Rapidata (third party direct debit provider)
- Engaging Networks t/a Eactivist (third party communications campaign provider)

J Predefined criteria for assessment

		Openness		
		Evidence of Data (1)	**Suspicion of Data (2)**	**No Suspicion of Data (3)**
Data Provided	No Data Provided (1)	1	2	3
	Partial Data Provided (2)	4	5	6
	Data Provided (3)	7	8	9
	No Response (or Data Processor)	0	0	0

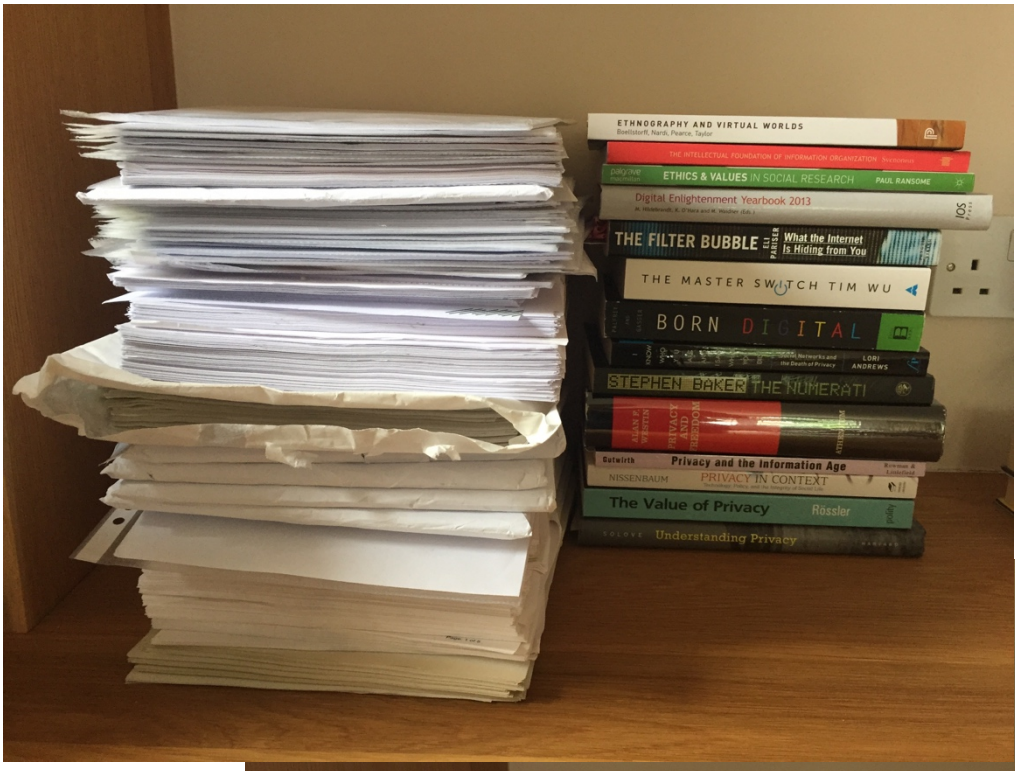
K Spreadsheet for analysis

Organisation	First Reply					Cumulative after Second Reply					Difference between First and Second Replies							
	Digital Footprints	3rd Party Digital Footprints	Data from Other Sources	Data to Other Sources	Digital Persona	Digital Footprints	3rd Party Digital Footprints	Data from Other Sources	Data to Other Sources	Digital Persona	Digital Footprints	3rd Party Digital Footprints	Data from Other Sources	Data to Other Sources	Digital Persona	TOTAL of VALUES	TOTAL of VALUES after 1st Reply	TOTAL of VALUES after 2nd Replies
365Scores	2	3	3	2	3	2	3	3	2	3	0	0	0	0	0	26	13	13
Acxiom	0	0	5	1	4	3	0	6	1	4	3	0	1	0	0	24	10	14
Amazon	4	4	2	2	3	4	4	2	2	3	0	0	0	0	0	30	15	15
Amnesty	1	7	9	9	0	7	7	9	9	0	6	0	0	0	0	58	26	32
Ancestry	1	1	3	3	3	1	1	3	3	3	0	0	0	0	0	22	11	11
Apple	8	8	2	2	2	8	8	2	2	8	0	0	0	0	6	50	22	28
Boden	7	7	8	8	9	7	7	8	8	9	0	0	0	0	0	78	39	39
Bryan Mitchell (Geared)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Call Credit	0	7	4	4	9	0	7	4	4	9	0	0	0	0	0	48	24	24
Charles Tyrwhitt	1	2	3	3	3	1	2	3	3	3	0	0	0	0	0	24	12	12
CIFAS	9	9	2	2	9	9	9	2	2	9	0	0	0	0	0	62	31	31
Codegent (Learn Japanese)	8	0	0	2	0	8	0	0	2	0	0	0	0	0	0	20	10	10
Conde Nast	3	7	3	9	3	3	7	3	9	3	0	0	0	0	0	50	25	25
Coop Energy	1	1	2	2	2	1	1	2	2	2	0	0	0	0	0	16	8	8
Critical Hit Software (Jigsaw Puzzle)	2	3	3	3	3	2	3	3	3	3	0	0	0	0	0	28	14	14
Cult Pens	7	8	9	9	9	7	8	9	9	9	0	0	0	0	0	84	42	42
Day One	7	9	9	7	9	7	9	9	7	9	0	0	0	0	0	82	41	41
Dunnhumby	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Engaging Networks	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Equifax	0	5	5	1	2	0	5	5	1	5	0	0	0	0	3	29	13	16
Eventbrite	1	3	3	3	3	1	3	3	3	3	0	0	0	0	0	26	13	13
Experian	0	5	5	4	4	0	7	5	4	4	0	2	0	0	0	38	18	20
Facebook	4	4	2	2	2	4	4	2	2	2	0	0	0	0	0	28	14	14
Flurry	9	9	9	2	8	9	9	9	5	8	0	0	0	3	0	77	37	40
Frogmind	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Google	7	2	0	3	1	7	2	0	3	1	0	0	0	0	0	26	13	13
GRBiPhoneGames TLC Productions (Road Warrior)	2	0	0	2	0	2	0	0	2	0	0	0	0	0	0	8	4	4
GZeroLtd (TVCatchup)	2	0	0	2	0	2	0	0	2	0	0	0	0	0	0	8	4	4
H2O	0	4	9	9	0	0	4	9	9	0	0	0	0	0	0	44	22	22
HM Revenue	7	7	1	4	2	7	7	1	4	2	0	0	0	0	0	42	21	21
Instagiv	7	3	3	8	3	7	3	3	8	3	0	0	0	0	0	48	24	24
John Lewis	7	7	2	2	2	7	7	2	2	2	0	0	0	0	0	40	20	20
John Lewis Partnership Card	4	7	2	2	2	7	7	5	4	4	3	0	3	2	2	44	17	27
Joseph Turner	1	1	2	2	3	1	1	2	2	3	0	0	0	0	0	18	9	9
Laithwaites	7	8	4	9	9	7	8	4	9	9	0	0	0	0	0	74	37	37
Lands End	7	7	1	3	9	7	7	1	3	9	0	0	0	0	0	54	27	27
Lloyds Bank	7	7	8	8	6	7	7	8	8	6	0	0	0	0	0	72	36	36
Lloyds Bank Pension	1	1	1	1	3	1	1	1	1	3	0	0	0	0	0	14	7	7
M&S	7	9	9	9	9	7	9	9	9	9	0	0	0	0	0	86	43	43
Mail Chimp	0	1	1	2	1	0	7	7	2	1	0	6	6	0	0	22	5	17
Mastercard	1	1	4	4	5	1	1	4	4	5	0	0	0	0	0	30	15	15
MCL Software Ltd	0	3	2	2	2	Follow up	3	2	2	2	0	0	0	0	0	18	9	9
Met Office (Weather App)	8	0	0	2	0	8	0	0	2	0	0	0	0	0	0	20	10	10
Mobiata (FlightTrack)	2	3	3	3	3	2	3	3	3	3	0	0	0	0	0	28	14	14
MobileIntoCenter (MacHash)	2	3	3	3	3	2	3	3	3	3	0	0	0	0	0	28	14	14
MobilityWare (Free Solitaire)	2	3	3	3	3	2	3	3	3	3	0	0	0	0	0	28	14	14
Natwest	2	2	1	2	2	2	2	1	2	2	0	0	0	0	0	18	9	9
NHS	0	0	0	0	0						0	0	0	0	0	0	0	0
Not On The High Street	7	0	9	9	9	7	0	9	9	9	0	0	0	0	0	68	34	34
Office for National Statistics	2	1	2	1	3	2	1	2	1	3	0	0	0	0	0	18	9	9
One Voice	7	9	9	9	3	7	9	9	9	3	0	0	0	0	0	74	37	37
Open Rights Group	7	7	9	9	9	7	7	9	9	9	0	0	0	0	0	82	41	41
Orvis	7	8	9	9	9	7	8	9	9	9	0	0	0	0	0	84	42	42
Oxford City Council	7	7	2	1	9	7	7	2	7	9	0	0	0	6	0	58	26	32
Oxford University Press Pension	7	7	8	8	3	7	7	8	9	3	0	0	0	1	0	67	33	34
Parcel Force (Royal Mail)	4	4	1	1	3	4	4	1	1	3	0	0	0	0	0	26	13	13
Parseq Fulfilment House	3	1	1	1	3	3	1	1	1	3	0	0	0	0	0	18	9	9
Personal Telephone Fundraising	9	7	7	7	9	9	7	7	7	9	0	0	0	0	0	78	39	39
Prolog	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pure 360	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rapidata	3	1	1	3	3	3	1	1	3	3	0	0	0	0	0	22	11	11
Readdle	2	3	3	2	3	2	3	3	2	3	0	0	0	0	0	26	13	13
Refugee Council	9	3	8	8	9	9	3	8	8	9	0	0	0	0	0	74	37	37
Rogavi (AIUK Raffle)	3	3	1	1	3	3	3	1	1	3	0	0	0	0	0	22	11	11
RSPB	7	9	9	9	3	7	9	9	9	3	0	0	0	0	0	74	37	37
Sea Containerd Pension	3	7	2	5	3	3	7	2	5	3	0	0	0	0	0	40	20	20
Sn&ck Media	3	3	3	3	3	3	3	3	3	3	0	0	0	0	0	30	15	15
South Lakeland District Council	9	7	2	1	9	9	7	2	1	9	0	0	0	0	0	56	28	28
Spotify Ltd	1	2	3	3	3	1	2	3	3	3	0	0	0	0	0	24	12	12
Sutton Seeds	7	1	1	3	3	7	1	1	3	3	0	0	0	0	0	30	15	15
Synetics	3	9	8	8	3	3	9	8	8	3	0	0	0	0	0	62	31	31
Taylored Mortgage & Investment	9	1	1	3	9	9	1	1	3	9	0	0	0	0	0	46	23	23
Tesco	7	7	6	6	2	7	7	6	6	2	0	0	0	0	0	56	28	28
thetrainline	7	3	3	8	9	7	3	3	8	9	0	0	0	0	0	60	30	30
Trustpilot	7	3	3	3	3	7	3	3	3	3	0	0	0	0	0	38	19	19
Twitter	7	7	3	2	3	7	7	3	2	3	0	0	0	0	0	44	22	22
UKBA	1	1	1	1	3	1	4	1	1	3	0	3	0	0	0	17	7	10
United Utilities	9	4	3	6	3	3	7	8	6	9	-6	3	5	0	6	58	25	33
Unlock Democracy	7	3	9	8	9	7	3	9	8	9	0	0	0	0	0	72	36	36
Vodafone	1	4	2	2	2	4	4	8	8	8	3	0	6	6	6	43	11	32
Xiao Yixiang (Pro Metronome)	2	3	3	3	3	2	3	3	3	3	0	0	0	0	0	28	14	14
Zurich Life	3	4	4	8	2	3	7	6	8	8	0	3	2	0	6	53	21	32

Key to Scoring

	Evidence of Data	Suspicion of Data	No Evidence of Data
No Data Provided	1	2	3
Partial Data Provided	4	5	6
Data Provided	7	8	9
Not Applicable	0	0	0
	If there is no suspicion of data being held and the organisation says that it does not hold that data, record as data provided. If there is suspicion of data being held and the organisation states specifically that it does bot hold the data, record as data provided unless external hard evidence exists to contradict the organisation. provided		

L Photograph of some of the responses



M Category and sector list

Category	Sector	Organisation
Central Government	Central Government	HM Revenue
		NHS
		Office for National Statistics
		UKBA
	IOS App	Met Office (Weather App)
Local Government	Local Government	Oxford City Council
		South Lakeland District Council
NGO	Charity	Amnesty
		One Voice
		Open Rights Group
		Oxford University Press Pension
		Refugee Council
		RSPB
Private Company	Charity Fund Raising	Engaging Networks (data Processors)
		Instagiv
		Personal Telephone Fundraising
	Credit Reference	CIFAS
		Synetics Solutions Inc
	Data Processing	Pure 360
		Rapidata
	Finance	Mastercard

Category	Sector	Organisation
	Finance (continued)	MCL Software Ltd
		Parseq Fulfilment House
		Prolog
		Taylored Mortgage & Investment
	Internet	Mail Chimp
Private Company (continued)	Internet (continued)	Unlock Democracy
	IOS App	365Scores
		Bloom Built (Day One)
		Bryan Mitchell (Geared)
		Codegent (Learn Japanese)
		Conde Nast
		Critical Hit Software (Jigsaw Puzzle)
		Frogmind
		GR8iPhoneGames TLC Productions (Road Warrior)
		GZeroLtd (TVCatchup)
		Mobiata (FlightTrack)
		MobileInfoCenter (MacHash)
		MobilityWare (Free Solitaire)
		Sn&ck Media
		thetrainline

Category	Sector	Organisation
		Trustpilot
		Xiao Yixiang (Pro Metronome)
	Marketing	Dunnhumby (Data Processors)
		Rogavi (AIUK Raffle)
	Online & High Street Shopping	Charles Tyrwhitt
		Laithwaites
		Orvis
	Online Shopping	Boden
		Cult Pens
		Joseph Turner
		Not On The High Street
Private Company (continued)	Online Shopping (continued)	Sutton Seeds
	Utilities	Coop Energy
Public Company	Credit Reference	Call Credit
		Equifax
		Experian
	Finance	John Lewis Partnership Card
		Lloyds Bank
		Lloyds Bank Pension

Category	Sector	Organisation
		Natwest
		Sea Containers Pension
		Zurich Life
	Internet	Ancestry
	Internet (continued)	Eventbrite
		Google
		Spotify Ltd
	IOS App	Readdle
	Marketing	Acxiom
		Flurry
	Online & High Street Shopping	John Lewis Includes Waitrose
		Lands End
		M&S
		Tesco
	Online Shopping	Amazon
		Apple
	Social Media	Facebook
		Twitter
Public Company (continued)	Utilities	H2O
		Parcel Force (Royal Mail)
		United Utilities
		Vodafone

N Interview information sheet

UNIVERSITY OF Southampton

Participant Information Sheet

Study Title: An investigation into Subject Access Request Responses

Researcher: Brian Parkinson **Ethics number**: ERGO/FPSE/23880
Please read this information carefully before deciding to take part in this research. If you are happy to participate you will be asked to sign a consent form.

What is the research about?
Recent Which research confirms that "sensitive personal and financial data is being traded on a huge scale".
My research also concerns personal data, firstly creating a categorisation for it, and then requesting my own information from 82 organisations. This was then analysed for completeness, before comparing the 'quality' of the responses by organisational type, sector, and location. The outcomes raise issues for data protection officers, IT professionals, and legislators.
This final stage of the research seeks to understand the reasons behind these findings. In particular, the causes for incomplete data; why some types of organisation or market sectors provide more complete responses than others; in which way legislation may be changed; how compliance to the 1998 Data Protection Act is viewed; and finally whether a categorisation of personal data could be of value.

Why have I been chosen?
I am interviewing legislators, people from think tanks, data protection and IT professionals.
You have been selected as an expert in your field who can help me to understand the research findings to date, and suggest ways forward. I also consider that you will be interested in what I have discovered so far.

What will happen to me if I take part?
I will come to visit you at a time and place that is convenient for you, where we will be able to discuss what I have found, and ask for your opinion. I expect that this will take about 20 to 30 minutes.

Are there any benefits in my taking part?
You will have access to my summarised research to date, and have the opportunity to discuss them. You will also be able to access the full findings including the conclusions of this series of interviews when it is published towards the end of the year.

Are there any risks involved?
Not really, you risk losing a small amount of time.

Will my participation be confidential?
Yes. Any notes or recording will be identified only by code e.g. person A and all information will be held on an encrypted device. This is in compliance with the University of Southampton policies and ethical guidelines.

What happens if I change my mind?
You may withdraw from this research at any time.

What happens if something goes wrong?
In the unlikely case of concern or complaint, you may contact the University of Southampton Research Governance Manager (02380 595058, rgoinfo@soton.ac.uk)

Where can I get more information?
Contact me by email at blp1m11@soton.ac.uk or by phone to 07767 222720.

O Interview consent form

CONSENT FORM (1.0)

Study title: An investigation into Subject Access Requests Responses

Researcher name: Brian Parkinson

Ethics reference: ERGO/FPSE23880

Please initial the box(es) if you agree with the statement(s):

I have read and understood the attached information sheet and have had the opportunity to ask questions about the study.	☐
I agree to take part in this research project and agree that a transcript of the interview may be used for the purpose of this study	☐
I understand my participation is voluntary and I may withdraw at any time without my legal rights being affected	☐

Data Protection

I understand that information collected about me during my participation in this study will be stored on a password protected and encrypted computer and that this information will only be used for the purpose of this study. All files containing any personal data will be made anonymous.

Name of participant (print name).....

Signature of participant.....

Date.....

P Interview guide

Background

Perhaps if I begin ... So far in my research I have developed a model of personal data and written to 82 organisations for copies of my own information. Most of the organisations were in the UK. It is the findings from the analysis of the responses, or lack of them, that I wish to explore in this conversation.

Would it be ok for me to record us so that I can concentrate on what you have to say?

If yes, thank you I will not associate your name with the recording but I will get it transcribed and send you a copy if you would like one.

If no then with your permission I will make some notes as we go along.

There are five areas that I would like to cover but if you feel uncomfortable with anything we discuss we can move on or stop the conversation.

Interview Questions and Prompts

Question 1

It may be best if I start with the categorisation of personal data which I developed by examining the terms used to describe personal data.

<concentric model of personal data>

It starts with an individual laying down their own data on the internet or some other place where it is captured digitally.

From there we work out to …

For the next part of my research I wrote to over 80 organisations asking for a copy of the personal data that they held that described me.

I then used the model to analyse the data that was sent, so for example a financial organisation sent these types of data

<Lloyds analysis model from back>

Question 2

In that analysis of data (subject access requests) I found significant variations in the completeness of the information that I was given, do you have any thoughts about what may be the cause of this?

<heat maps>

Prompt: employee skills infrastructure

avoidance of expense

retain competitive advantage

Question 3

Another finding was that some sectors or categories of organisation appear to perform better than others, do you have any thoughts about why that might be?

<category analysis>

<sector analysis>

Prompt: Government (local & central) much worse than Charities Private companies worse than public companies

Question 4

There are two classes of data that are not covered under current legislation for subject access requests. The first is profiles from analytics, and the second is about location, where data is transferred to and where it is or held. I found that some organisations were happy to provide that data but most were not. Do you have any thoughts about why organisations would not be happy to let people have this information?

Why do you think some organisations provide it anyway? Should it be always be available for people to access?

Prompt: Should legislation be amended to allow people access to a wider range of information, such as profiles from analytics?

Data analytics were in their infancy in 1998 when the Data Protection Act was passed

In 1998 data was rarely transferred between counties

Only recently has data harvesting and sale become common due to improved interfacing capabilities

Location of data - where it is from / to / held - laws protecting of personal data varies by country

<response by data category 1st response>

<response by data category final response>

Question 5

In what ways do organisations view subject access requests, are they a cost or a benefit?

Would there be different views within an organisation depending upon role?

Prompt: Has your organisation considered providing (Why don't organisations provide) feedback facilities when data is sent out so that errors can be investigated and corrected?

Question 6

With respect to categorisation that I mentioned at the start of this conversation, can you see where it may be useful - or not of course

Q Interviewee descriptions

Interviewee			Interview Duration
Id	**Description (Expertise)**	**Expertise Code**	**(mins)**
1	Owner of a tech company employing c120 people (Information Technology Expert / Entrepreneur)	Ent	96
2	Member of the Management Group of the National Association of Data Protection Officers (Data Protection Officer)	DPO	85
3	Former Member of the UK Cabinet [1] (Politician)	Pol	60
4	Former Member of the UK Cabinet [2] (Politician)	Pol	25
5	Think tank - policy director (Think Tank)	TT	97
6	Recently retired IT Director of FTSE 100 Company (Information Technology Expert)	IT	70
7	Senior IT Professional Magic Circle Law Firm (Information Technology Expert)	IT	46
8	Data Protection Officer – Utility (Data Processing Officer)	DPO	59
9	Data Protection Officer – Government (Data Processing Officer)	DPO	56
		Ave	66

R Transcriptions

All interview transcriptions can be accessed through the following link

https://web.tresorit.com/l#LEPE7bvMrZjIRZWJSVDhAQ

The password is:- blp1m11

The interviews ranged from 25 to 96 minutes. Interview 7 is shown below as an example it represents a 46 minute interview.

Interview 7

BLP

00:00:46.11

Here is the model this is based on ... I extracted 65,000 documents and I settled on 247. The terminology was all over the place. Digital Footprints, those are data artefacts laid down by the data subject. Third Party Digital Footprints, that is artefacts laid down by somebody else which are descriptive of the data subject. So, notes in a bank's, bank teller notes. Digital Personas which are basically analytics. So, from the bank we used to do collection data, and outside the ring is third party data, demographics. This inner band within the circle which we call the digital mosaic and the whole is the digitally extended self, as it is the digital extension of the self into the virtual. We have used this, I say we, my supervisor has used this at the British Society lecture, and he has also used it for the police to explain what personal data is and what it does. Also, when talking about privacy it is used on masters courses at Southampton. So, I asked two ex-cabinet members, basically I was trying to find about whether it would be of use in any legislative areas, and the answer was yes because most people are totally fuzzy about what private data is, how it is constructed and don't have a set vocabulary to describe it. So, that is ok, but there are issues that I can see because this doesn't differentiate between data and metadata and one of the big issues for legislation is that differentiation, we only want your metadata we don't want your data, honestly. Which in some respects is a little bit hypocritical because the metadata is more useful than data because no terrorist is going to write how do you fancy bombing the Houses of Parliament. What you really want to know is who they are talking to. One the ex-cabinet members was very critical of Amber Rudd with all this stuff about we need to decrypt WhatsApp and they said they don't, they really don't. Anyway, so what happened was this has been published, it went out yesterday in the Journal of Information Science. But one of the reviewers said hang on this should be useful in data science as an overarching data model. I must say, and my reaction, and I am prejudicing you now, my initial reaction was, I don't think so however I will put it in the paper. So, I guess my

question about the model, is does it make sense, but also would it have any applicability for people who are actually systems architects, strategists?

Interviewee 7

00:05:03.19

I think yes, I get what you are saying what you are saying about the model and not understanding the different components of it. I think this is more of a big data thing. I think it is more a what you do with it kind of thing. It is more of an information scientist, I think they call them nowadays, they have invented new titles. So, they are kind of like data architects but not data architects. With data architects are more around how do we avoid duplication of data and how do we avoid that data getting out of sync and how do we share it between the applications whilst staying within the bounds of the law etc. Whereas this is stepping in to a much wider thing of what is it I want to know. It is almost who are the people out there that are likely to want this and who are the people out there are likely to do that. Or who are the people ... and it is more around I think the whole big data thing. It is more around that, marketing targeting and discovering. So, as you quite rightly said find out people what are they likely to like and how can I suggest new things. Or who has this person been associated with and what do they do and what can I actually find out about them. There is more in that space I think than IT architecture inside of systems.

BLP

00:06:43.13

Mainstream IT architecture?

Interviewee 7

00:06:44.90

Yes this is really information sciences as they call it

BLP

00:06:45.23

I think that you have called it right in that, I mean I have not thought of it in those terms but almost certainly the Journal of Information Science targets librarians attached to it and people who are interested in how you classify data. There is a huge amount of work going on now with big data.

Interviewee 7

00:07:14.00

Yes, actually this is kind of the thing, your Digital Footprint, that is the kind of stuff that you want to see and know about for yourself, your Third Party Digital Footprints are things you might want to request. What is it you know about me? What is it you think about me? And that is the only type of data that the individual probably knows, but as you say this whole other stuff around here. Where have you been on Google Maps, what I have eaten, what restaurants I have been in what reviews I have done, all that kind of stuff is the stuff that you are using to build up the personas, the individuals. What it is that they are potentially like which goes beyond IT architecture, it goes into data science, information science and it is the stepping into ... it is like that tool, it can be used for good or it can be used for evil. You know so my phone, I have got location tracking and it is quite scary you know you look at this map and it tells you where you have been for the

BLP

00:08:14.18

You have it turned on?

Interviewee 7

00:08:19.80

Yeh, yeh, you know what I find some of the stuff, and the insights it gives me quite useful. You know, if I am walking, it comes up and says your next trams at such and such, you will be home by such and such. Now if it didn't know that I regularly got the tram backwards and forwards it wouldn't be able to tell me that. You know you are here, you have recently been in these restaurants and one of them round the corner and well that is useful to me but the other side of the coin is what has somebody has hold of all that data. What could they do with it and I am kind of more, and I know that this is probably naive but if you haven't done anything wrong you have nothing to fear unless you live in a kind of state that kind of seizes that data and persecutes people because of it. Which is where the danger is when you start to get into Trump land, you know, what is he going to do with the data, that he gets off the American people? How is he going to use it to expel people from the country, how is he going to use it to do all that kind of stuff? Then it starts to become more sinister. But erm, you are right, I am easy.

BLP

00:09:31.24
I did a lecture at Sussex some time ago and I had only just discovered the Apple tracking. So, I put mine up, downloaded it, and put it on a slide, I put it up for people to see and everyone's face, what! Not a single person in the room who wasn't flabbergasted.

Interviewee 7

00:10:03.15
Yes, it is the same with Google, you look at anyone's data. But you know what it is actually quite good as well. Supposing you become a suspect and the police said to you what were you doing at such and such a night and you are like I don't know. Oh right I went to this restaurant then I had this I staggered home here, here's me walking the 2 and a half miles an hour. So it can be quite useful as well. It also can be quite, this is where you start, I know you are into privacy laws or whatever, when you are starting with crime detection and all those kind of things. Is it a bad thing to be able to identify who is within the area at the time. There are criminals after all and there is evidence should we or shouldn't we be using it for that kind of thing. Now I believe that they do, if someone's a suspect, have the ability to seize people's phones and look at the data. But while they are not suspects or they are not then ...

BLP

00:11:10.20
Apparently, and I haven't read the latest terrorism stats, but generally access to that data is quite restricted and not as you would see on Spooks.

Interviewee 7

00:11:18.94
No, no, I get that. But the other is you know that is the thing about people who think, people do think that what could, could the government ... and I kind of know that they have got a hell of, of a lot of data, and they could do a hell of a lot with it but actually they are not really that good. You know it's like they have got all this stuff and they don't really know how to use it. It is, you may think that Big Brother is watching you, it more a Big Brother may have the tools to be able to watch you but it doesn't know how to use them. And this is the other thing with a link between architecture and this model. Unless you can start to link these things together like Google does when you sign in with Google they are disparate bits of data that are not linked because the data keys are not the same so you may

have footprints all over the place but unless you can match them up to one person they are not really that useful

BLP

00:12:20.49

Hasn't technology changed here because I haven't been working for donkey's years. When I was working a number of companies would come up and say right we can put a customer front end in and we will link into all your back end systems and it will be absolutely fantastic. It never worked because of all the keys being different and the names are spelt differently or whatever. But it strikes me that somehow something has changed or else systems have moved on.

Interviewee 7

00:12:53.99

Yes, some data has been used to good effect. If you look at Microsoft and all these things these are all just Office tools run in the Cloud usually a One-drive etc. and then you have got your connection information who you send emails to what documents you have read and it almost becomes a work place environment like an Amazon or a Facebook. It kind of says you know what three of your colleagues that you regularly communicate with have read this document you might be interested in it. Which is quite good, and then there is all this stuff around being able to search through a pile of information to be able to find something which is of use, or of interest, whilst before it was all in people's mailboxes or it was on people's hard drives or whatever. Now it is in a big pot which you can search through and find the stuff that you are allowed to see. It will show you, this might be of interest and it rates it by if your colleagues have looked at it, how often it has been looked at and all that kind of thing.

BLP

00:13:58.00

So, there is some intelligence at the back of it now. Whereas Ask Sam, if you remember Ask Sam, was pretty basic just looking through the odd word.

Interviewee 7

00:14:12.29

And the other advances that have been made in that kind of technology is around data protection and also making sure that you are ... because we live on the internet, mostly everybody lives on the internet and actually it has almost become impossible to defend yourself by building walls round your data, because people can break in. It is accepting the CIA are in there, the Chinese are in there, the Russian are in there. You know the Chinese have got devices in every router on the internet. You cannot protect data you have got to protect data from an encryption and numbers point of view. So, it makes it too difficult to break into the information rather than the store. So, you can get into the store but you can't get the information.

BLP

00:15:00.92

because it is encrypted

Interviewee 7

00:15:05.05

Because it is encrypted, and what the new stuff is actually, because things like Office 365 are global when you start getting consecutive attacks from friendly or unfriendly forces you know from intelligence agencies, or criminals. You may call intelligence agencies unfriendly as well, many people do. But if you start, they can recognise bad things, because they recognise your patterns of behaviour and say you know what this is out of the ordinary for this guy and wait a minute, a minute ago he accessed that data from Germany and now he is accessing it from Scotland. That is not right because, you know they can shut it down, lock it down. So that kind of information is useful because if they know your habits they can see when something unusual has occurred and they can break it down. Same with credit card transactions today it is kind of, they know what you buy, they know where you go, and they all of a sudden bang, something has happened, they phone you up and go did you make this purchase. So, it is not all bad. I had a phone call someone had bought World of Warcraft stuff on my credit card number and the company phoned me up with 5 transactions through the night with World of Warcraft and it was not me mate and it, it was ok we will take them off your account sir. Just don't worry about them when you see them. You will see them be refunded and whatever. That is good use of data but you have the messy side to as well so it is kind of swings and roundabouts

BLP

00:16:46.20

My American cousin describes it as a transaction. You give people the data, they give you the services. That's how he sees it as good as he doesn't get as much junk mail that is inappropriate he gets more targeted stuff.

Interviewee 7

00:17:01.61

I think it is a good thing too. I am a for open data and as I say very much along the lines of do you know what in our country, I believe that if you haven't done anything you need to hide then you are ok. I don't think we are quite at the Chairman May stage where the Conservative Party are about to look for who ...

BLP

00:17:26.71

Well we will find out in a few weeks ...

Interviewee 7

00:17:26.76

Absolutely. So, yes clearly it is being used for evil purposes. There is that article about effecting outcomes of elections in the Guardian.

BLP

00:17:45.99

I don't know whether you shared it with me or whether Judy's youngest shared it with me

Interviewee 7

00:17:54.35

I sent it to you on messenger or Facebook

BLP

00:17:54.36

It must be you then. Alan works as a policy officer for Uber UK and Ireland and so he knows, is it Susan Schmidt?

Interviewee 7

00:18:10.11

Yes

BLP

00:18:12.99

He knows her, he has had dinner with her

Interviewee 7

00:18:18.45

That is another thing. Uber, I remember when I was in Singapore and a taxi was, you phone them up and they know where you are from your GPS or you just go and use the app and then it tells you what the registration number of the taxi is. How long it is going to be and then bump, you can track it coming towards you. Great

BLP

00:18:44.92

Absolutely, and you can send the details to somebody else and they can track you coming towards them so if you are a woman travelling alone at night in a strange city you can be tracked.

Interviewee 7

00:18:55.28

Yes, but you need to turn on that kind of location tracking and then once it is there what else can you use it for? So, it is kind of ...

BLP

00:19:00.98

Well, I turn it off after I have used it. That is the sort of guy I am I am afraid. I am changing, the more I think about this the more I can see the advantages of it. But I just resent large corporations exploiting me, but I think that I need to come to terms with it.

Interviewee 7

00:19:24.28

Yes, and I think that is quite a telling phrase when you say that they are exploiting you. They don't want to waste your time and piss you off trying to sell you something that you don't want. And they also don't want to waste their time. So, what they are trying to do is you know, who is it that might be actually wanting to get this message and they are going to try and target that. And if you don't you can opt out and things like that. So, it is kind of ...

BLP

00:19:46.06

I think also that there is also there is an age difference a generational issue here. It is not ability with technology it is what technology you are happy using. I think that is the difference. So, like Alan, his Uber and his WhatsApp he doesn't care that Facebook are analysing his data on WhatsApp, whereas I shut it down and don't use it.

Interviewee 7

00:20:18.11

Yes, but if you say you turned it on when you want to use and you turn it off when you don't want to. That in itself can be used in evidence so for example I am back and forwards to work every day, everybody knows where I live, where I come to, so it's not a secret, well I don't mind it, and it's not a secret that I go to this restaurant and it's not a secret that I go to that restaurant so I am quite comfortable with that. If one day I turned my hone off, somebody might go, what was he doing, where did he go

BLP

00:20:43.77

Like in Line of Duty ...

Interviewee 7

00:20:46.64

Yes, absolutely, So, it’s like none evidence can be gaps in evidence can be used as well. So, if you try and cover your data trail that in itself leaves a trail.

BLP

00:21:01.09

Yes, one of the things that I have found is that lack of a digital footprint can also be effectively a digital footprint. So CIFAS, who I didn't know existed, it is a fraud prevention agency that Lloyds use. Have no data describing me on their books which means that I am not a fraud risk.

Interviewee 7

00:21:26.24

OK so that is the flip side. My wife is very private she doesn't use anything on Facebook. She doesn't do this, she doesn't do that. She doesn't have any credit cards, she doesn't, because I have credit cards that she uses and we have joint accounts. So, anything, she doesn't have a data record, she doesn't have anything at all so when she tries to get credit. She can't get credit.

BLP

00:21:49.24

I think this is an issue, I mean it is not to do with this, but there is a levelling about data and the analysis, the analysis the digital personas . They cause a levelling throughout society, and if she's effectively a none person because there aren't any ...

Interviewee 7

00:22:10.76

She is

BLP

00:22:13.49

Will she be able to get into States or will flags start to be raised?

Interviewee 7

00:22:17.51

This is the question isn't it so it is kind of, there is something to be said for not blocking people creating digital personas because if you haven't got a digital debt persona or a digital credit persona then people are going to go why has this person not got one. Are they somebody who has been doing fraud and regularly disappearing and starting up a new persona?

BLP

00:22:45.03

Intelligent people very often will create a persona that they want to be seen by. There is a sociologist called Goffman and he talks about how we play a different role depending upon, you know like, I am a football hooligan when I get to Old Trafford, I am a housewife in Oxford etc. and people will try to create different personas.

Interviewee 7

00:23:07.94

Do they get found out?

BLP

00:23:13.03

Well they will do not but 4 years ago, there wasn't the interconnection

Interviewee 7

00:23:18.91

So here at <names company> once you get offered a job you have to go through a back check. It is the longest hardest back check I have ever seen in my life and you will not get the job if you, if I put on my cv that I was a Head of such and such and I was actually a Deputy Head they will find out and they will not give you the job. They are so thorough, as integrity is everything in this firm. So, there are people who will be in debt and get that information, you know, and if they can't get that information you are not going to get the job anyway. Whether they can prove it or no. If they can't prove it you have had it

BLP

00:24:02.69

So back to this. Information scientists, not IT departments in that organisation.

Interviewee 7

00:24:12.40

I think so, I think it is more about information science than architecture. Architecture is more concerned with creating these things

BLP

00:24:24.18

Not what you do with them? That was my view but my view of IT really is well out of date now. Things have moved on so much. Anyway, the next interesting thing though was having written to all these all organisations, got the data back and I mapped it across this, found out where the holes were, because, I am not quite sure how to describe myself really, a little bit anal. But every organisation that I have had contact with over the net I have got a folder for and there are over 500 of them.

Interviewee 7

00:25:08.99

Gives you something to do in your spare time.

BLP

00:25:10.62

It's what you do when you are retired you have nothing else to do, no I just file things away otherwise I lose them. To be fair I have some software now which organises and sorts my mail for me but I have got into the habit of it and it is useful. It has been useful for my PhD because I can say look here is my persona and I can list 500 different organisations that I have been in contact with. It tells you so much, just knowing that you have been in contact tells you so much about a person it is embarrassing actually. But also, having worked in IT I know roughly what sort of information an organisation will have to have certainly in these terms. I know there will be digital footprints otherwise I wouldn't have been writing to them because I have been dealing with them. But by the nature of the organisation you can tell whether there is going to be third party digital footprints. So, if it is Lloyds there is somebody in a bank branch who is bound to have made a note about me. And you kind of know, or there is an expectation of which sort, or there is an exception of which sort of organisation will have some analytics on you. Certainly Google and Facebook which are two that I have written off to. When they wrote back I did a spreadsheet and analysed against this lot and wrote off to them again asking them for the bits that I thought were missing and quite a lot of them said well, OK mate, and sent it back. Because generally certainly commercial companies are very happy to tell you what data they have. Then I could score them against what they sent me against what I thought they should have sent me. Now that is subjective, but then there is bound to be some fuzzy edges. I analysed it of course on a spreadsheet, and this is the first one. There is no secrets

here, you can see everything that I have got. These are the elements of the model and I also asked them, where they got data from and where they sent it to. As I was interested in, really in the Data Protection Act where everything is covered and in Safe Harbour at the time was up in the air, and actually was seen to be valueless. So, 62% of organisations in the end provided Digital Footprints. So, HMRC sent me back every one of my tax returns, the Digital Footprints. Curiously more organisations sent me back Third Party Digital Footprints than Digital Footprints. This is data from other sources, which is surprisingly high as you don't need to supply that under the data protection act, it is excluded as are Digital Personas and obviously external data going into the personas. So, my first question was, was this a reasonable pattern. Bearing in mind that this is a score not of the organisations that gave me some information about Digital Personas but the % of organisations that I knew, in quotes, had Digital Personas, who gave me information. So, John Lewis has transaction going back to 2004, transaction data, you could say why are they storing that, and there justification actually is that they use all of that data to create digital personas that we have of you, and that is our rationale for keeping it, which is a bit dubious. It was quite chilling, the earliest thing on there was a restaurant bill, in Hove when I went to visit my kids, with my ex-wife. A bottle of wine and some headache tablets on the same day. And that sent a shiver down my spine which I thought was weird as I know about this stuff, but the level of detail that was being kept and the pile of paper that I was sent was surprising. In all I had a pile of paper that high which when scanned and OCRd gave me 16gb of data. But look how poor people are at giving you, your personas as compared to their elements. I think I know why, but do you have any ideas?

Interviewee 7

00:31:19.19

Maybe the way that they analyse data is what gives them competitive advantages and by sharing that they are sharing their competitive advantage. So, if they have a competitor, let me see how you look at people, what, let me see your data on people, but let me see how you view people then that would give me an idea of marketing strategy their retail strategy all that may the reason why they may not be happy to give that data to you. This is the sort of information that will give organisations competitive advantages

BLP

00:32:09.13

And that was my underlying assumption. But also, it is not well covered by the current Data Protection Act.

Interviewee 7

00:32:18.68

It is changing with the GDPR which is limiting the amount of data that you can store about somebody and for how long.

BLP

00:32:32.01

It is probably a good thing, I don't know.

Interviewee 7

00:32:32.94

But what they will do is that they will summarise the data so that it is meta data and then update the meta data with the new data as it arrives. They won't be storing your data they will be storing their view of you. You fit in box A, you fit in box D and then if anything changes of he has moved from box D to box A right so it is kind of, they won't have the data which has put you in the box D to start with but they will know that you have been in Box D.

BLP

00:33:05.56

So, they will be storing the, the output from the analytics but not the input data to the analytics which will restrict future analytic development in some respects because they won't be able to revisit the data.

Interviewee 7

00:33:26.53

But it is kind of, how useful was that stuff 10 years ago. I was in a different place then with a different amount of money, living in a different house.

BLP

00:33:35.98

As an aside have you heard a Radio 4 programme called Digital Identity, this week they were talking about a group of people who were adherents to a film called Shazam. And the researcher wanted to find a copy of it but the film had never been released, never been made. But there is a whole group of people who believe that it had been and that is because their memories had been constructed differently. So, what they remember actually didn't exist. What they thought they were 10 years ago they weren't. So, I think that you are right people change so dramatically. I am an entirely different person now I am quite laid back compared to when I was working. I must have been absolutely terrible to live with. You change so much over time. I also suspect that people who answer Data Protection requests just don't know the data, I wasn't surprised about that result. But I then analysed the responses by category of organisation, so public companies, charities, private companies, local government and central government. Now this is for me the wrong way round. If you were to give me lengths of cardboard and ask me to place them against the different categories I would not have got them right.

Interviewee 7

00:35:29.80

So, what is this score?

BLP

00:35:34.80

The score is based on a scoring matrix which is an amalgam of how well l think they performed against what data I thought they had, against the data that they provided.

Interviewee 7

00:36:16.44

So, kind of this comes back to what I said earlier, if you may think that government have all that data, and maybe they have. They just don't know they have or they just don't realise that they have. And these are the ones who are actively using it so they know exactly what they have got. So maybe that is the right way.

BLP

00:36:37.99

That is a good explanation. My thought was that, but public companies, roughly what you said before, there is a competitive advantage therefore they would keep it to themselves.

Central government are not in competition with anybody so they can just give you the data. That was what was in my head, and actually was my supervisors view when I discussed it with him.

Interviewee 7

00:37:06.73

The other thing is having worked in government I can tell you when you get Freedom of Information requests or you get whatever it is, you think oh shit, and you have got to dig up bits, because it is not connected, you have a whole load of work to pull it together. With these guys probably have it all connected all tied up and just go whoosh.

BLP

00:37:34.19

So, you are saying that these guys have got their act together

Interviewee 7

00:37:38.03

They know how to use data, and are using it, and these guys aren't. They probably have a lot more data than they know they have and they haven't constructed the kind of things that you think they have because they don't have that ability.

BLP

00:37:52.66

So that actually is in alignment with one of my ex-cabinet members who said that Central Government in my opinion is totally transparent.

Interviewee 7

00:38:02.96

You can see what they can see, they just can’t see very much

BLP

00:38:08.23

But it is so not transparent I can’t believe it.

Interviewee 7

00:38:15.67

I think that is more through incompetency than capacity.

BLP

00:38:22.23

That is interesting. Because that is a different view

Interviewee 7

00:38:24.11

I worked in Government for 10 years so I know it is more of an incompetency issue

BLP

00:38:39.53

Well I worked in Central Government for a few years and in those days

Interviewee 7

00:38:45.04

It was all on cards - laughs

BLP

00:38:46.32

Yes, it was all on cards, but information was power so your boss would never tell you anything that he didn't have to tell you and so in my mind these guys were keeping the information like that because knowledge is power and it was a matter of, there is term, governmentability, so having to control the population

Interviewee 7

00:39:04.69

You know if you asked MI5 what your information was I would have expected them to give you the bare minimal. If you asked the DWP it is only because they are incompetent.

BLP

00:39:21.45

I asked the Office for National Statistics, they told me to sod off because they are a research organisation and they didn't have to tell me. So, I wrote back and said just because you didn't have to tell me doesn't mean that you can't, can I have the data. They said sod off - laughs, UKBA, they phoned and apologised and said that due to budget cuts we haven't got a complete list of who came in and out of the country because they didn't record it for a while. You can't believe that can you?

Interviewee 7

00:39:49.75

No, I can

BLP

00:39:50.33

Well I know it is true because one of my ex-cabinet members said I remember that

Interviewee 7

00:39:55.36

Yes, there was a big scandal about it

BLP

00:40:23.62

Well they wouldn't tell me about the no-fly list which apparently is called the Watch Index in the UK. HMRC they just sent me back what I had sent them but I know they have got at least 16 data feeds but as you say it is cock up rather than conspiracy.

Interviewee 7

00:40:24.66

You know, did you go to you know these companies that give you your credit score and all that

BLP

00:40:30.35

Yes Equifax etc.

Interviewee 7

00:40:32.45

Yes, and they'll give you a lot of data, you know where they get most of that data from? Local Government and Central Government because they get all of that stuff and they know what to do with it so it is your voting register stuff, all of that your income, the only thing that they get from somewhere else is the banks but the majority of the stuff you got we give it to them and then, when I say we, I mean the government gives it to them, and then we have to pay to get it back, after they have analysed it, because we don't have that capability. That is why you are getting those low scores.

BLP

00:41:15.89

That is poor, so Local Government they didn't tell me, to be fair it wasn't a normal Subject Access Request because I said can you give me all of the data. They just missed out the electoral role as in well it is not your data. That was a case of not being nasty it was just incompetency

Interviewee 7

00:41:45.82

I am not on the voting thing

BLP

00:41:53.26

And then I analysed it another way. This shows

Interviewee 7

00:41:58.81

You know this doesn't surprise me, it doesn't surprise me at all. Here are the ones who are wanting to know more about you so that they can get money off you, and they are the hottest at analysing or whatever. Once you start with credit reference agencies and blah blah blah, once you start getting down here, you know what, they are not trying to sell you anything,

BLP

00:42:23.25

Yes, the IOS app people are basically and or two guys in a lounge they don't do data generally. Social media are bit naughty

Interviewee 7

00:42:36.82

And they are getting naughtier though aren't they

BLP

00:42:40.69

Well, that was Facebook and Twitter. Mailchimp were a bit naughty they told me that I had to get a subpoena from the State Court of Georgia to get any data

Interviewee 7

00:42:51.93

Yes, but they are not under the UK or EU regulations

BLP

00:42:57.99

No they don't have to give you any data. They did in the end because I wrote back to them and said but according to this journal article your boss has published you are analysing all the emails that you have sent and making profiles of people. Then they just sent the data back straight away because they say they never do that sort of thing but their boss had written a journal article on it. So, there is no surprise there.

Interviewee 7

00:43:36.94

The only surprise here is the utilities one. You know it is kind of it depends ...

BLP

00:43:45.65

Basically my expectations of the utilities were very low, because why would they create a persona on you.

Interviewee 7

00:43:52.93

So, in a sense, they gave you everything that you expected them to give you, that is OK right.

BLP

00:43:56.49

They didn't give me very much but it was everything that I would expect

Interviewee 7

00:44:03.42

That's why it sticks out

BLP

00:44:05.51

Which is why it sort of sticks out but first off, they were pretty poor, they were down here, so wrote off to them and said what about x,y,and z and they said OK.

Interviewee 7

00:44:15.43

Again, they probably don't know what data they have got and only when you pointed that out, I think that they are probably more aligned to the government in. You know what they have a lot of data and they don't really do anything with it.

BLP

00:44:28.49

I think your right actually because actually most of them were government nationalised industries who have probably not moved that far. <Person X> did actually go and work for British Energy you should hear his stories about nuclear reactor safety

Interviewee 7

00:44:44.29

Oh, don't even go there

BLP

00:44:47.92

I think that is probably it, I think that is where we are. This was interesting, who gave me most data, EU organisations.

Interviewee 7

00:44:59.02

Absolutely

BLP

00:45:00.81

It is what you would expect really isn't it. There is no surprise there

Interviewee 7

00:45:01.18

EU compliant to EU legislation, UK compliant to EU legislation uh ho!

BLP

00:45:12.70

Basically, continental Europe are much more personal data savvy than we are. Germany because of the Stasi, Merkels grew up under the Stasi, what do you expect

Interviewee 7

00:45:20.46

We have got all of our data in Germany because they have the highest level of data protection in Europe.

BLP

00:45:31.78

I don't use Dropbox, I use an organisation called Tresorit, they are a German organisation. They keep their data in Switzerland. It encrypts on my laptop. The encrypted data is then sent to their data centres and that is it. They don't know what my data is. It is fantastic security.

Interviewee 7

00:45:58.62

Dropbox?

BLP

00:46:00.08

Yes, no privacy at all. But that is no surprise is it?

Interviewee 7

00:46:09.47

EU countries at the top

BLP

00:46:12.71

I thought that actually getting information back from some of the American and Israeli companies would be difficult but they gave me back more than I expected. However, the guy in China didn't respond at all. That is about it.

Interviewee 7

00:46:39.21

Did you get what you want from me?

BLP

00:46:39.64

I did thank you very much, can I give you a lift?

S Interview thematic analyses

Ref	Comment	Sub-theme
1.23	01:16:42.16 'they are definitely considered a cost' *wrt SARS*	Approach to SARs
1.28	01:24:24.89 'What I question is what the importance is of people being able to request their own information, that is the fundamental bit of this, what does that gain you as an individual or what does that gain in society either end of it the individual or the organisation. It shows your ability to gather and share data records it doesn't show your ability to protect an individual from it.'	Approach to SARs
1.4	00:12:20.31 'data protection is a thorn in my side'	Approach to SARs
1.5	00:13:19.46 'that wouldn't be primary or secondary or even tertiary role it would come pretty near the bottom of the list'	Approach to SARs
2.12	00:30:51.66 'SARs are a bit of an irrelevance really because if someone writes to you and says he wants personal information, a member of staff on £16 grand a year has never heard of and doesn't understand' 'I would be much more concerned about the letter mentioning court cases etc.'	Approach to SARs
2.20	01:09:22.44 (1) 'yes, they are an extra cost'	Approach to SARs
2.21	01:09:22.44 (2) 'you are questioning our integrity, so yes, they do by and large all view them as though they are negative' *They are seen negatively within organisations as they are seen to be questioning the organisations integrity*	Approach to SARs
2.7	00:14:06.77 'I guess the other thing that I can think of and this is very likely they do have the information to hand and fully well know that this is the answer to it they know that very few people complain when it comes to answers to subject access requests'	Approach to SARs
3.14	00:51:58.56 'we used to worry a lot more about Freedom of Information requests that we ever worried about subject access requests,'	Approach to SARs
3.15	'they don't think it is very important they don't have the systems set up to access it'	Approach to SARs

Ref	Comment	Sub-theme
6.10	00:34:26.73 'the idea that you say that John Lewis have got data from you from 2004 and the excuse to keep it for so long is that they are keeping it for their big data analysis. It seems completely unreasonable in my opinion'	Approach to SARs
6.11	00:43:33.10 'when you are in a company trying to service those it is a complete pain in the arse because you know, supposing for example you have got 400 databases that might contain Brian Parkinson, you are not thinking about how do we satisfy this request in the most complete and rational way you are thinking about I have got £15 how can I get away with it.'	Approach to SARs
6.15	01:03:15.19 'They were privatised government bodies' *with reference to utilities poor performance on the first request*	Approach to SARs
6.2	00:11:08.44 'The trouble is when you say that you can't be trusted because you will accidentally do it' *referring to organisations gathering data for one purpose only but using it for something else*	Approach to SARs
6.4	00:14:15.11 'Yes, and the other thing is that, and don't write any of this down, then you will start to look at potential fines the probability of being found out, it may not be found out on your watch because there is a probability that you will have been promoted to another job so someone else will have to sort it out'	Approach to SARs
6.6	00:17:15.80 'I think the trouble is you are walking into a tsunami of personal data which is getting thicker and thicker and bigger and bigger every year and it is probably, even if legislated and defined it is probably a bit unstoppable because it is very hard to police it'	Approach to SARs
6.9	00:34:26.73 'where compliance is in conflict with the rational world of business everyone is trying to drive round it'	Approach to SARs
7.9	00:37:06.73 'The other thing is having worked in government I can tell you when you get Freedom of Information requests or you get whatever it is, you think oh shit, and you have got to dig up bits, because it is not connected, you have a whole load of work to pull it together. With these guys probably have it all connected all tied up and just go whoosh. ' *Comparing the government's attitude to FIO requests as opposed to SARs*	Approach to SARs
8.12	00:31:33.52 (2) 'probably lack of staffing so in terms of resources that tends to delay. They may eventually reply to you but it might be like 6 months later in complete disregard of the legislation and etc.' *wrt Central & Local Government poor responses*	Approach to SARs
8.13	00:31:33.52 (3) 'I feel that the Government entity might actually rely on other legislation to get out of doing subject access requests on the basis that we can't disclose any information for such and such reasons etc. and it is more likely that they have template letters for that than to have a template letter to reply to a Subject Access Request'	Approach to SARs

Ref	Comment	Sub-theme
9.10	00:23:22.54 'at its peak we had around 4,000 cases that were overdue back in August, September last year, and we have worked really hard through a combination of process reviews, continuous improvement methods, and just kind of influencing customer behaviour in that way as well, doing a bit of triaging, a whole range of things, to help us get completely on top of that' *wrt focus on improving SAR response times*	Approach to SARs
9.12	00:37:35.88 'as a matter of course now we will look at it as a two stage approach where we provide the IT stuff first, and then if the customer comes back to say that they are not satisfied with that, that there must be more information, and they want everything that is held on the Home Office files, we will respect that and we will do that. So, that is how we see it. We see that as a preliminary stage and then we will put something through the full process if the customer comes back and says where is the rest'	Approach to SARs
9.18	00:49:44.11 'they got me mixed up with the same name who had not been providing their self assessments. Somebody who had been running a laundry business or something like that. And as a way of applying penalties they had amended the tax code to recoup that money, and I obviously wasn't very happy with the state of affairs, and asked them to provide copies of all the information that they held on me. To compound matters it came through in the post a couple of weeks later in a ripped envelope, A4 ripped envelop, ripped all down one side, so that I had absolutely no idea, you know, if what landed on my door mat was a complete set of the documents that they had sent out, or just part of it. So, you know it was a catalogue of errors really one thing, if it wasn't bad enough to kind of mix me up with somebody else. Then when they send the data through the way in which it lands with me doesn't fill me with much confidence that it is handled with respect shall we say.' *an illustration of how poorly HMRC handled a request for information*	Approach to SARs
9.4	00:07:22.02 'And we did look at that last year to see if there was a more effective way of doing this. If you like influencing customer behaviour because for a long time it has been my belief that a one size fits all approach doesn't really work and that there must be something that we can do in that space. With that in mind that is why we brought out our fast track service which incentivises customers to narrow down the scope of their data request. We can respond to those much quicker and we pass on the time and money savings to the customers because we don't charge for those. So, we have been doing a bit to see if we really zone in on what matters most to customers so that we can provide a slightly more tailored service which then enables us to focus our resources in the best way possible and get through the requests. *Refers to a system of providing a small amount of commonly asked for data as a first response in order to cut down time of response and cost*	Approach to SARs
9.5	00:11:25.00 'I can't help but think that for perhaps the majority the process is seen a convenient cheap copying service' *legal representatives ask for a copy of a client's file rather than for a client's data - a possible breach of the act*	Approach to SARs

Ref	Comment	Sub-theme
9.8	00:20:10.19 'I suppose many of the customers have an axe to grind and have perhaps had an unhappy experience and that a lot of the time it will be just wanting to know when am I going to receive my biometric resident's permit or whatever.' *wrt SARs normally being a part of the complaints process*	Approach to SARs
9.9	00:21:05.80 'We don't see a huge number of FIOs, my section deals with the bulk of <organisation> SARs there is another section which deals with cross cutting cases, ones which may involve other government departments and ones which come from current or former employees. But the bulk would come to my section. So, whilst I am part of UKBA Immigration we would deal with any SARs that relates for example to Border Force or Immigration Enforcement matters as well. And we deal with around 22,000 a year.' *compare with other interview that suggests that Govt departments focus on FIO requests rather than SARs*	Approach to SARs
3.16	00:52:08.44 'I suspect that for public companies they can get it quite easily it is not as complex to get hold of it'	Capability (IT or Otherwise)
5.10	00:27:04.77 'I think that people have good intentions but in general the data is one of capacity, a large company will have people who understand data protection' *wrt NGO responding to SARs*	Capability (IT or Otherwise)
5.9	00:26:08.13 'Within NGOs the main problem is the lack of capacity'	Capability (IT or Otherwise)
6.14	01:01:56.59 'I wonder if historically some parts of private enterprise have got higher investment in computing so that they have just got a lot more modern stuff. Where it is possible to pull it off more easily. Whereas maybe some government bodies are 4 or 5 years behind and so they probably have not got it all together, which probably has an impact.'	Capability (IT or Otherwise)
7.1	00:11:18.94 'the government ... and I kind of know that they have got a hell of, of a lot of data, and they could do a hell of a lot with it but actually they are not really that good. You know it's like they have got all this stuff and they don't really know how to use it'	Capability (IT or Otherwise)
7.10	00:37:38.03 'They know how to use data, and are using it' *with reference to public companies*	Capability (IT or Otherwise)
7.11	00:37:38.03 'They probably have a lot more data than they know they have and they haven't constructed the kind of things that you think they have because they don't have that ability.' *WRT Central Government having data but not knowing the what to do with it*	Capability (IT or Otherwise)
7.13	00:38:15.67 'I think that is more through incompetency than capacity.'	Capability (IT or Otherwise)

Ref	Comment	Sub-theme
	in reference to central government providing low amounts of information but with a background of working for the government for 10 years	
7.14	00:39:04.69 'You know if you asked MI5 what your information was I would have expected them to give you the bare minimal. If you asked the DWP it is only because they are incompetent. '	Capability (IT or Otherwise)
7.15	00:40:32.45 'the government gives it to them, and then we have to pay to get it back, after they have analysed it, because we don't have that capability. That is why you are getting those low scores.' *talking about credit agencies such as Equifax*	Capability (IT or Otherwise)
7.18	00:44:15.43 'Again, they probably don't know what data they have got and only when you pointed that out, I think that they are probably more aligned to the government in. You know what they have a lot of data and they don't really do anything with it.' *WRT the Utility Sector*	Capability (IT or Otherwise)
7.2	00:11:18.94 'a link between architecture and this model. Unless you can start to link these things together like Google does when you sign in with Google they are disparate bits of data that are not linked because the data keys are not the same so you may have footprints all over the place but unless you can match them up to one person they are not really that useful'	Capability (IT or Otherwise)
7.8	00:36:16.44 'you may think that government have all that data, and maybe they have. They just don't know they have or they just don't realise that they have. And these are the ones who are actively using it so they know exactly what they have got.'	Capability (IT or Otherwise)
8.1	00:04:08.99 'there is a known issue, and it wasn't to do with people it was to do with the actual system unfortunately,' *Explains how issues with SARs can arise*	Capability (IT or Otherwise)
9.11	00:28:25.01 'HMRC and I think MOJ both have a similar experience in terms of matching the resources available to the advancement of customers and I am aware that HMRC have a unique challenge in that a lot of their data goes back many, many years. The records go back to the 1960s I think some of which is held on microfiche which must be quite difficult to manage.'	Capability (IT or Otherwise)
9.13	00:39:25.25 'the most difficult part of that process is printing off the IT records in the first place because of the way that the database is structured' *wrt Government computer systems and other comments about being 5 years behind*	Capability (IT or Otherwise)
9.14	00:40:37.02	Capability (IT or Otherwise)

Ref	Comment	Sub-theme
	'It is one of those age old problems really the system itself is quite an old one. It is going to be replaced at some point by a new all singing all dancing thing which hopefully will have the functionality that we need but what is the point in the meantime in investing developers time on something that is going to be replaced' *wrt Government computer systems and other comments about being 5 years behind*	
9.18	00:49:44.11 'they got me mixed up with the same name who had not been providing their self assessments. Somebody who had been running a laundry business or something like that. And as a way of applying penalties they had amended the tax code to recoup that money, and I obviously wasn't very happy with the state of affairs, and asked them to provide copies of all the information that they held on me. To compound matters it came through in the post a couple of weeks later in a ripped envelope, A4 ripped envelop, ripped all down one side, so that I had absolutely no idea, you know, if what landed on my door mat was a complete set of the documents that they had sent out, or just part of it. So, you know it was a catalogue of errors really one thing, if it wasn't bad enough to kind of mix me up with somebody else, then when they send the data through the way in which it lands with me doesn't fill me with much confidence that it is handled with respect shall we say.' *an illustration of how poorly HMRC handled a request for information*	Capability (IT or Otherwise)
1.2	00:10:35.86 'it is natural that any client record system predominately comes from digital footprints'	Common Requests
8.7	00:22:04.68 (2) 'people when they are logging subject access requests all they want is copies of mostly call centre recordings, account notes, it is because no-one logs a request because they are happy.'	Common Requests
1.15	50:58.93 (1) You don't have a choice whether you use HM Customs and Revenue' *they are not customer focused*	Competitive Situation
1.18	00:59:47.66 'online and high street shopping need your persona to be effective particularly online in marketing because it is just their very nature that online intelligence is everything to them' *so they have personas whereas Local & Central Government don't?*	Competitive Situation
7.16	00:41:58.81 'Here are the ones who are wanting to know more about you so that they can get money off you, and they are the hottest at analysing' *explaining why retailers and online shopping performs better*	Competitive Situation
7.6	00:31:19.19 'Maybe the way that they analyse data is what gives them competitive advantages and by sharing that they are sharing their competitive advantage'	Competitive Situation
1.16	*00:50:58.93 (2)* 'These people are only interested in protecting their own job'	Customer Focus
1.17	00:52:54.65	Customer Focus

Ref	Comment	Sub-theme
	'A really good example of that is in Local Government whether you can have access on your computer to another web site the default answer is no, and then on exceptions they will unlock access to a web site. So, you might well find that in Local Government and local police force they have no access to YouTube unless they make a specific request to have access. So, they run a white list everything is no and they only turn it on to add it to a white list. Whereas in the public and private sector we tend to run a blacklist, everything is yes until they exclude it.' *Local & Central government tend to use white lists for web security (i.e. the answer is no by default unless you can prove you need access to say YouTube), public and private companies tend to use a blacklist (i.e. the answer is yes unless something gets banned) - so the governmental approach is safety first*	
1.26	01:22:29.84 'they didn't see it as a conversation' *wrt error correction from SAR requests*	Customer Focus
2.9	00:18:20.91 'SARs, most of them are about people being really annoyed about something the bank did and so the SAR element to it, they are using the tool to extract information, to argue their point so most of it ends up back at their corporate customer service to resolve the complaint'	Customer Focus
3.17	00:52:08.44 'let's just give it to him to stop him moaning' *In relation to Public Companies providing data. <I guess that the corollary is that Central Government doesn't care if you are upset>*	Customer Focus
5.22	1:01:49.10 'modern marketing from the 1930s started and created the idea of you needed to please the customer and in order to please the customer you needed to know the, the desires before they even know them it is the perfect butler no? But then of course how do you know peoples desires, you start to hold information and the thing is that if you look at the process from that perspective it is a great mass of things something quite not denied, but you need to try to be the butler, we are doing this for you not because we want to know about you for any nefarious purposes but like your butler we want to give you a better service and that argument is still being made in many areas not that we need data particularly the online advertising and this is where it comes from I think that the novelty is now wearing off. People say maybe I don't want you to know my other side and also it is annoying so from having a nice butler you know it is moving to an overbearing'	Customer Focus
1.20	01:12:02.92 'you are interested in the data you are not particularly interested in keeping a record of where it came from because once you've got it you've got it'	Data Location
1.21	01:12:58.16 'it probably gets passed around in non-legal ways, and therefore keeping a breadcrumb of where it came from is incriminating, isn't it?'	Data Location
1.22	01:14;52.65	Data Location

Ref	Comment	Sub-theme
	It is only natural though, if you were to think that, if you go back to the 18th, 19th Century and that whole part about if your name became discredited you had a daughter who married out of wedlock, not marry out of wedlock, had a child out of wedlock. Your name would go through the mud, in Downton Abbey times and - but actually even in those times with very poor routes of communication you would be excluded from court or people would give you the cold shoulder at church etc. so it went through communities. It only takes one person to create a piece of information about you, a rumour, and traditionally if it is something of interests, it would branch out quickly and it would be everywhere overnight *Not knowing where 'rumours' or pieces of damaging information come from or reside is normal*	
1.8	00:24:03.88 'personally, we don't get much information from other organisations'	Data Location
2.13	00:34:11.40 'The expectation, what everyone thinks the big classic one is DWP actually does hold a lot of data but that is a relatively new phenomenon HMRC used to hold all of that data and now the last 10 years or so it is the DWP who really hold it' 00:36:47.98 Right so the Home office actually doesn't hold that much personal data, it is the DWP agencies that hold them, the DFE holds the national pupil data base apart from that it doesn't really hold that much so I would say that if you sent requests off to government by and large the answers are that we don't have any.' *Interviewee doesn't believe that government departments hold much data excepting for the DWP*	Data Location
5.13	00:36:19.34 'HMRC and DWP are the main data holders and everyone else wants to access their data'	Data Location
5.23	01:21:00.04 'you may want to look at data portability requirements in the GDPR there is this thing about data generated in the course of a contract' 'if you want to have this information at some point then GDPR you want may to look definitions around the data portability which I can't remember right now but there is a very clear distinction in the type of data you are expected to get from a company when you are requesting in 1 year's time.'	Data Location
6.12	00:52:49.91 'you might think that you can hold the wave back but you can't. You can do what you like but you can't hold it back.'	Data Location
6.13	00:59:57.31 'by outsourcing it 2 or 3 times because you know if you have layers and layers of outsourced organisations it is very difficult to pin down what had been going on.'	Data Location
6.16	01:08:00.67 'I think the thing that worries me more is personal data, which might be used by third parties in an unattractive way. Just using to market to me doesn't really matter too much.'	Data Location
6.7	00:19:25.62	Data Location

Ref	Comment	Sub-theme
	'look at the legal ramifications of how these organisations are arranged you can't even chase the money let alone chase things like big data'	
7.20	00:45:20.46 'We have got all of our data in Germany because they have the highest level of data protection in Europe.' *references data location for Interviewees own organisation*	Data Location
8.17	00:55:17.85 'It doesn't have to be in the black market it is actually the open market as well and it is justified by existing marketing regulations that say it acceptable to sell a marketing list. Surely there are some regulations around it surely you need to be able to check the source. But it is acceptable that marketing lists are sold and bought.'	Data Location
8.18	00:57:02.73 'in a couple of year's time or less they won't be able when the GDPR is here. They won't be able and nobody else will be able, none of us will be able, to just justify generically like that because the people in office will have to be very specific of the names of all the third parties that they will be using or might be using or may be processing the data. So, you can't get away with that any longer either when the GDPR comes in to play.' *wrt generic statements regarding the destination of data*	Data Location
8.4	00:18:17.48 'so therefore, it won't have gone to other third parties that tells us something.' *Compare with 8.5 21:47:99 when the interviewee states that information goes to third parties*	Data Location
8.6	00:22:04.68 (1) 'I might know on that basis who we send it to but would I actually when you file this Subject Access Request would I actually go and say ooh but also let me mention that it has been sent'	Data Location
8.8	00:22:04.68 (3) 'if I was to receive this in these terms obviously I would apply them as they are and I would have to go and look whether the data has actually gone and be very specific when I respond to you and say we do also send your data to da da da, in the performance of your contract da da da. Which of course then you would have to go to Experian if you wanted to know how they are managing your data etc. etc.' *Compare to 8.4 00:18:17:48*	Data Location
9.15	00:46:35.97 'Certainly, if I had any concerns about how my, about responsibly an organisation is using my data, then I think one of my questions would be I think where my information is being drawn from who else might have access to it. What exactly is going to be done with it. And I think my questions would extend to what is your data retention policy. How is it destroyed, when, and all those kind of questions.' *Note also a comment about data retention policy, matches a comment made by another Data Protection Officer*	Data Location
9.16	00:47:28.04 And which country it is held in? Interviewee 9	Data Location

Ref	Comment	Sub-theme
	00:47:30.31 Yes, yes, absolutely	
9.17	00:47:56.09 'I had cause to contact a private company and you know just sort of ask what sort of information they hold on me and what they are doing with it, and how long they intend to hold on to it, and all those kind of things and it is the only time I have ever put in anything purporting to be a Subject Access Request of my own. An interesting thing was that in the initial discussion with the online operator, they sort of said something like, well you know that you can follow this process and this is what you have to do but please be aware that by doing this you are not going to find out the reasons why we have made this particular decision in this case. And I am thinking under the Data Protection Act I am entitled to know you know what your intentions are towards me. But the point I, that I am coming to, is that this is a situation where it was far from clear where the information would be held, because it is a company that has a presence here, a company that has a presence I think in Gibraltar, and therefore it opens up questions about which legal regime you're, you are talking about. And therefore, which Information Commissioner you would want to go to if you had a complaint.'	Data Location
1.24	01:19:17.64 'key thing is preventing my staff from sharing the information' 01:20:15.35 'the biggest threats that most organisations have is losing that information into the public domain' *the focus is on keeping data secure, not having it passed where it should not be or to be hacked*	Disposition to DPA & SARs
2.1	00:00:02.68 'it is difficult to get the government to really care about this very much … it really does appear that no-one is really interested.' *it appears that no-one is really interested in the protection of personal data - unless they can get more data and do something with it.*	Disposition to DPA & SARs
8.9	00:30:23.08 (1) 'I think the feeling of requirements to comply with legislation so you would definitely expect a public company or an NGO to be absolutely firm and strict in replying immediately.' *Explaining why public companies and NGOs perform so well*	Disposition to DPA & SARs
6.8	00:31:08.76 'Because the quality of their staff is so much worse' 'the competitive element the pure business element of life has not really struck them' *on why Central Government performs worse*	Efficiency
7.13	00:38:15.67 'I think that is more through incompetency than capacity.'	Efficiency

Ref	Comment	Sub-theme
	in reference to central government providing low amounts of information but with a background of working for the government for 10 years	
7.17	00:43:36.94 'The only surprise here is the utilities one' *on Utilities performing well after a follow up as expected them to perform badly as Central Government, reason being was that I had low expectations e.g. no persona*	Efficiency
1.19	01:04:38.20 'Most app developers are either single, single individuals or collaborative, like hacking organisations' - *hence they are unaware of DPA responsibilities*	Knowledge / Training
1.9	00:28:37.97 *Creates categories e.g. Leader, Manager but does not consider them a Persona or Analytics*	Knowledge / Training
2.11	00:24:38.66 *The ICO give a lot of information to companies, with you must do this you must do that ... so people can't go round saying that they don't know anything about it*	Knowledge / Training
2.17	00:58:04.65 'you need to tell us what you want and where we can likely find it and if you don't do that they sit here in their ghettos there is some evidence of that' *As an explanation for Local Government performing poorly*	Knowledge / Training
2.18	01:00:30.89 'citizens advice bureau as the data on property basket issues with payments of benefits of certain kinds is more up to date than the governments there analytics are absolutely incredible I have never seen anything like it so I wandered over to see what they were doing I sat down and I had a two hour session with the head of analytics but I lost the thread and he really, really even he was really clear a very good communicator and I sort of turned around and he had stopped and I said that I'm dead now I can't do this anymore.' *Analytics are too hard to understand so tend not to be known about or summarised for a SAR request or customer*	Knowledge / Training
2.5	00:14:06.77 'they are not used to being asked this question so don't understand how to answer it. They won't know where the stuff is'	Knowledge / Training
3.12	00:51:00.82 'simply don't hold it simply in one area' *In reference to Central Government not providing good data*	Knowledge / Training
5.6	00:21:03.23 'even if they wanted to give you information do they understand what they have, and do they know what you mean … because the more sophisticated concepts around data constructs maybe harder for them' *wrt analytics*	Knowledge / Training
5.8	00:21:58.97	Knowledge / Training

Ref	Comment	Sub-theme
	'They should be trained yes' *staff trained for the normal SARs but not for digital personas or data transfers.* *Additionally, the language of the request may be different that that normally used in compliance circles making it less likely that distal personas were provided (check terminology against bill). This is a good reason for common terminology to be adapted.*	
8.2	00:06:12.36 'In large organisations it is also that there are different phases when you have a contact centre. So, the most likely be staff who are employed in a contact centre are not fully trained in Data Protection jargon. They may not recognise the words. Or they may be skilled in or trained in Data Protection jargon, but the customer isn't so the customer may use a different a different set of words so that is not recognised as a Data Protection request.'	Knowledge / Training
1.18	00:59:47.66 'online and high street shopping need your persona to be effective particularly online in marketing because it is just their very nature that online intelligence is everything to them' *so they have personas whereas Local & Central Government don't?*	Mission & Vision
3.11	00:51:00.82 'I can understand the UK Border Agency not wanting to tell you information that may fall under the category of security '	Mission & Vision
5.16	00:44:58.45 'My perception is that there is an element of control, power through control' Interviewee 5 00:45:01.45 'Yes absolutely'	Mission & Vision
5.17	00:47:08.68 'the question of whether they hold information on behalf of the population or whether the state is something that is separated from the population'	Mission & Vision
6.16	01:08:00.67 'I think the thing that worries me more is personal data, which might be used by third parties in an unattractive way. Just using to market to me doesn't really matter too much.'	Mission & Vision
7.16	00:41:58.81 'Here are the ones who are wanting to know more about you so that they can get money off you, and they are the hottest at analysing' *explaining why retailers and online shopping performs better*	Mission & Vision
7.20	00:45:20.46 'We have got all of our data in Germany because they have the highest level of data protection in Europe.' *references data location for Interviewees own organisation*	Mission & Vision
1.6	00:13:46.73 'Data Protection is about what we are taking in not what we are sharing out'	Practice

Ref	Comment	Sub-theme
1.7	00:15:56.30 'a lot of the time that you are looking at your data you don't really know the source'	Practice
2.14	00:46:05.88 'who can demand that the data is deleted, so if you can delete all the data and you haven't breached the contract, you are the controller. Now if I'm commissioning you to do a bunch of work right and we give you £100 a month and you delete all the data the chances are that you have broken the contract. So, you are probably a processor. I am on the other hand the owner and that makes sense to me can I phone you up right now and say transfer all the data to this other random third party would you have to do it,' *Defining the difference between the Data Controller and the Data Processor*	Practice
3.8	00:35:21.06 'I think the categorisation of data would be helpful but then alongside that the matrix if you like is nature of data / use of data So people might be willing, you know I am perfectly happy for this hospital or any other hospital to know a lot about me. But it is because I know it is being used for my health care. I would not want the same data to known by other public sector bodies'	Practice
4.5	00:12:38.60 'I think there is too often an assumption that the means drives the intention whereas actually the question is what is it I am trying to achieve, what is it that I am seeking that I don't currently have, and will it actually help. And the question that was always uppermost and was with the recent Investigatory Powers Act was if you get what you think you need will it actually be applicable will it be usable will it be if you like accessible and integrated to a point to what you thought you wanted it for, will actually turn to be out of any value whatsoever.' 00:14:39.36 'I think the reason I am raising it is that you have to decide what the end product what is the end objective of all this and then try and design what you want to get rather than what you first thought of.' *In relation to Government powers and the acquisition and access to data*	Practice
4.8	00:23:41.92 'I was not thinking of the Information Commissioner although that is a dam good approach I was thinking of the oversight mechanisms that the Investigatory Powers Bill drew together mm so you know we had a surveillance commissioner and all of that and they are now one and I think that would be a really good approach which would then influence the civil servants who are having to carry through the existing legislation and it is getting to the civil servants at this early stage. The alternative is the special advisors and, on the grounds, that this is going to be helpful, it helps for instanced in the aftermath of the Westminster Bridge to actually analyse how we how we make decisions and whether we achieve what we set out to Good look with it'	Practice
7.19	00:45:01.18 'EU compliant to EU legislation, UK compliant to EU legislation uh ho!' *when discussing EU non UK responses being better than UK responses*	Practice
8.14	00:31:33.52 (4)	Practice

Ref	Comment	Sub-theme
	'a plc and a private company have pretty much the same governance requirements'	
8.15	00:37:43.03 'ICO they have been a very relaxed regulator, they are quite in the background, not really looking for conflict , not really looking to make peoples life difficult.'	Practice
8.18	00:57:02.73 'in a couple of years time or less they won't be able when the GDPR is here. They won't be able and nobody else will be able, none of us will be able, to just justify generically like that because the people in office will have to be very specific of the names of all the third parties that they will be using or might be using or may be processing the data. So, you can't get away with that any longer either when the GDPR comes in to play.' *wrt generic statements regarding the destination of data*	Practice
9.17	00:47:56.09 'I had cause to contact a private company and you know just sort of ask what sort of information they hold on me and what they are doing with it, and how long they intend to hold on to it, and all those kind of things and it is the only time I have ever put in anything purporting to be a Subject Access Request of my own. An interesting thing was that in the initial discussion with the online operator, they sort of said something like, well you know that you can follow this process and this is what you have to do but please be aware that by doing this you are not going to find out the reasons why we have made this particular decision in this case. And I am thinking under the Data Protection Act I am entitled to know you know what your intentions are towards me. But the point I, that I am coming to, is that this is a situation where it was far from clear where the information would be held, because it is a company that has a presence here, a company that has a presence I think in Gibraltar, and therefore it opens up questions about which legal regime you're, you are talking about. And therefore, which Information Commissioner you would want to go to if you had a complaint.'	Practice
1.25	01:21:19.09 'Well that strikes me as a lovely way, if I had a public company that would be a fabulous way of linking it to an advantage' *wrt error correction from SAR requests*	Processes
1.26	01:22:29.84 'they didn't see it as a conversation' *wrt error correction from SAR requests*	Processes
2.10	00:22:37.03 'Because it doesn't go the right place' *SARs do not go to the correct place within an organisation pointing to poor organisational control*	Processes
2.16	00:49:56.17 'the more likely you are to be complained about the more likely it is to be to have good positive procedures to have an easy way to get to the right person, to extract the data so that probably is part and the less likely you are the more likely you are to be at the bottom.'	Processes

Ref	Comment	Sub-theme
	an explanation of why Pubic Companies perform well and Local and Central Government badly	
2.6	00:14:06.77 'in a very, very big organisation where your analytics team could be part of a digital team, an IT team, part of the marketing team they could be absolutely anyone you simply wouldn't know where you would start internally getting an answer to the question.'	Processes
3.15	00:52:08.44 'they don't think it is very important they don't have the systems set up to access it'	Processes
8.10	00:30:23.08 (2) 'public company would have set processes in place'	Processes
8.11	00:31:33.52 (1) 'for the government, generally, central, local I am putting it down to lack of processes definitely'	Processes
9.10	00:23:22.54 'at its peak we had around 4,000 cases that were overdue back in August, September last year, and we have worked really hard through a combination of process reviews, continuous improvement methods, and just kind of influencing customer behaviour in that way as well, doing a bit of triaging, a whole range of things, to help us get completely on top of that' *wrt focus on improving SAR response times*	Processes
9.4	00:07:22.02 'And we did look at that last year to see if there was a more effective way of doing this. If you like influencing customer behaviour because for a long time it has been my belief that a one size fits all approach doesn't really work and that there must be something that we can do in that space. With that in mind that is why we brought out our fast track service which incentivises customers to narrow down the scope of their data request. We can respond to those much quicker and we pass on the time and money savings to the customers because we don't charge for those. So, we have been doing a bit to see if we really zone in on what matters most to customers so that we can provide a slightly more tailored service which then enables us to focus our resources in the best way possible and get through the requests. *Refers to a system of providing a small amount of commonly asked for data as a first response in order to cut down time of response and cost*	Processes
3.7	00:27:07.51 'I would slightly take issue with the idea that the collection of data by government is about control mm you could equally make the case that quite often it is about protection'	Protective
5.15	00:41:12.92 'It is a cultural and it is a ahh, they are very very protective about the data' *wrt HMRC* 'their main job is to fend off requests for data' *other public bodies want access to HMRC data but the confidentiality obligations they have, have created a culture of protecting data, but also, they see it as their data not the data subjects data*	Protective
5.16	00:44:58.45	Protective

Ref	Comment	Sub-theme
	'My perception is that there is an element of control, power through control' Interviewee 5 00:45:01.45 'Yes absolutely'	
5.17	00:47:08.68 'the question of whether they hold information on behalf of the population or whether the state is something that is separated from the population'	Protective
5.21	00:59:24.37 'they are custodians of the information, we are holding the information for you and everyone else you know because someone has to do this correctly '	Protective
1.1	00:01:04.77 'I don't think that we have had any requests of that nature' *when commenting that the organisation had not received any requests*	Size
1.11	00:38:55.39 'they are more poorly structured' 'they share roles more' *in relation to Public Companies and performing worse, they may share roles more as they tend to be smaller*	Size
1.13	(00:50:22.92) 'They may be too dam big that it is split into so many sub-departments'	Size
1.19	01:04:38.20 'Most app developers are either single individuals or collaborative, like hacking organisations' - *hence they are unaware of DPA responsibilities*	Size
1.3	00:10:35.86 'In my business I rarely use, if ever use, some of the latter sources'	Size
1.8	00:24:03.88 'personally, we don't get much information from other organisations'	Size
2.6	00:14:06.77 'in a very, very big organisation where your analytics team could be part of a digital team, an IT team, part of the marketing team they could be absolutely anyone you simply wouldn't know where you would start internally getting an answer to the question.'	Size
2.8	00:17:56.62 'It is difficult to find out, particularly in larger organisations, but they are expected to respond on behalf of the organisation as a whole'	Size
3.20	00:54:31.77 BLP 'But they were very happy to tell me that and I suspect that if it just 5 people in an office with one computer then they just do this this and this	Size

Ref	Comment	Sub-theme
	Interviewee 3 00:54:43.12 'I think that is right' *Wrt small organisations*	
5.10	00:27:04.77 'I think that people have good intentions but in general the data is one of capacity, a large company will have people who understand data protection' *wrt NGO responding to SARs*	Size
5.11	00:28:02.54 'for a large company it is a complete nightmare to try to find all the information.'	Size
1.11	00:38:55.39 'they are more poorly structured' 'they share roles more' *in relation to Public Companies and performing worse, they may share roles more as they tend to be smaller*	Structure
2.15	00:49:56.17 'If they are connected to a bank that takes money lends money that kind of - of course there are zillions of operations there but in a council you would be everything from picking up rubbish to running a swimming pool to teaching kids to planning applications so vast and varied but I have never worked in those environments and deliberately so because I always think that as organisations they are so fractured' *local government is very fractured and so isn't able to pull everything together*	Structure
2.6	00:14:06.77 'in a very, very big organisation where your analytics team could be part of a digital team, an IT team, part of the marketing team they could be absolutely anyone you simply wouldn't know where you would start internally getting an answer to the question.'	Structure
3.12	00:51:00.82 'simply don't hold it simply in one area' *In reference to Central Government not providing good data*	Structure
3.13	00:51:00.82 'they don't have people who are responsible for responding to it. They will have quite sophisticated Freedom of Information operations i.e. for dealing with Freedom of Information requests but clearly not in responding in the same way to, what is it data access' *In reference to Central Government not providing good data (note conflicts with UKBA interview)*	Structure
9.7	00:18:04.34 'I am not sure that there is anything more that I can add really. That is quite a tricky one for me. I suppose in terms of sharing of data that it is something which this particular area of the business isn't that sort of focused on.' *wrt data moved between departments, illustrating a compartmentalisation or a reluctance to discuss?*	Structure
1.12	00:42:58.98	Transparency

Ref	Comment	Sub-theme
	'if a public funded body has information they should share it with the public'	
1.14	00:50:22.92 'what I have generally noticed about Central or Local Government is their love of the word no' *commenting on the lack of competitive service culture within Central and Local Government*	Transparency
2.19	01:05:21.49 'when Governments become as transparent as they want you to be with your data' *Commenting on government's insistence on personal data transparency (for security reasons) but lack of governmental transparency*	Transparency
3.10	00:51:00.82 'I think that it is cock-up rather than conspiracy ' *With respect to Central Government not providing data*	Transparency
3.11	00:51:00.82 'I can understand the UK Border Agency not wanting to tell you information that may fall under the category of security '	Transparency
3.18	00:53:00.54 'they may have a set of values which is about openness' *wrt NGOs*	Transparency
3.19	00:53:00.54 'I think they have been quite hit about the thing about selling on people's data and I wonder if they are trying to be a bit open about that now in order to try and rebuild a bit of people's trust' *wrt NGOs*	Transparency
3.21	00:56:10.75 'the government organisation is more transparent.'	Transparency
4.6	00:14:39.36 'the bigger issue for me was always intrusive surveillance I always took the view that you were on firm ground if by normal traditional policing methods you would have been able to, to obtain something but with the methods you were using you could obtain it more effectively more efficiently more quickly in a modern era which is why so long as it is transparent CCTV I don't think we should be worried as long as we know, whereas if it was surveillance of your home and intrusion I was very, very jumpy about recording people in their own personal space and they didn't know'	Transparency
5.13	00:36:19.34 'HMRC and DWP are the main data holders and everyone else wants to access their data'	Transparency
5.14	00:36:39.69 'I should know who shares my data across government'	Transparency
5.18	00:48:06.67 'Firstly, I am more of a communitarian but I do believe that central and local government should be more transparent and I think if they were there would be more trust in them.'	Transparency

Ref	Comment	Sub-theme
	Interviewee 5 00:48:19.71 'Totally'	
5.7	00:21:03.23 'the question is whether they have the data and want to share it, it is seen as more sensitive' *wrt Analytics*	Transparency
7.12	00:38:02.96 'You can see what they can see, they just can't see very much' *when discussing Central Government being transparent at an ex-cabinet member commented*	Transparency
7.14	00:39:04.69 'You know if you asked MI5 what your information was I would have expected them to give you the bare minimal. If you asked the DWP it is only because they are incompetent.'	Transparency
9.12	00:37:35.88 'as a matter of course now, we will look at it as a two stage approach where we provide the IT stuff first, and then if the customer comes back to say that they are not satisfied with that, that there must be more information, and they want everything that is held on the Home Office files, we will respect that and we will do that. So, that is how we see it. We see that as a preliminary stage and then we will put something through the full process if the customer comes back and says where is the rest'	Transparency
9.6	00:14:21.51 'Although in sort of percentage terms it was a spike of requests of that nature, they were quite low in the scheme of things so actually in absolute terms it was fine.' *wrt an agency requesting data from an individual that it already has, a SAR can be used to obtain the data and to feed it back to the agency, is this a case of the government not being transparent?*	Transparency
9.7	00:18:04.34 'I am not sure that there is anything more that I can add really. That is quite a tricky one for me. I suppose in terms of sharing of data that it is something which this particular area of the business isn't that sort of focused on.' *wrt data moved between departments, illustrating a compartmentalisation or a reluctance to discuss?*	Transparency
5.18	00:48:06.67 'Firstly, I am more of a communitarian but I do believe that central and local government should be more transparent and I think if they were there would be more trust in them.' Interviewee 5 00:48:19.71 'Totally'	Trust
5.19	00:55:55.83 'they trust private companies and the continent of Europe is the opposite and people trust the government' *Anglo Saxons tend to distrust governments but trust companies, whereas for southern Europeans it is the other way round*	Trust

Ref	Comment	Sub-theme
6.2	00:11:08.44 'The trouble is when you say that you can't be trusted because you will accidentally do it' *referring to organisations gathering data for one purpose only but using it for something else*	Trust
6.3	00:11:33.43 'So, I would suggest that the government can't be trusted at all on almost anything, because there is so much lying and cheating'	Trust
7.4	00:22:17.51 'there is something to be said for not blocking people creating digital personas because if you haven't got a digital debt persona or a digital credit persona then people are going to go why has this person not got one. Are they somebody who has been doing fraud and regularly disappearing and starting up a new persona?'	Trust
1.10	00:35:12.82 'I think we just give them the data what it is and if we didn't understand how to categorise it'	Understanding Personal Data
1.27	01:23:00.27 'yes I thought it was good' *wrt model as an aid to understanding*	Understanding Personal Data
2.18	01:00:30.89 'citizens advice bureau as the data on property basket issues with payments of benefits of certain kinds is more up to date than the governments there analytics are absolutely incredible I have never seen anything like it so I wandered over to see what they were doing I sat down and I had a two hour session with the head of analytics but I lost the thread and he really, really even he was really clear a very good communicator and I sort of turned around and he had stopped and I said that I'm dead now I can't do this anymore.' *Analytics are too hard to understand so tend not to be known about or summarised for a SAR request or customer*	Understanding Personal Data
2.2	00:14:06.77 'digital footprint and the 3rd party digital footprint are things that more people will intuitively understand'	Understanding Personal Data
2.22	01:19:22.03 'It is simple and I might use that ' *In reference to the model*	Understanding Personal Data
2.3	00:14:06.77 'going down a bit these are very, very much more hidden processes, things which are very back office, that people are not clear about'	Understanding Personal Data
2.4	00:14:06.77 'the further down the list you go the more opaque these things are so this is why are the returns so bad'	Understanding Personal Data
3.1	00:13:54.71 'my first feeling is that it would probably be useful to have some sort of categorisation of data because I think people are very confused about the uses that are made of their data in public policy and legislation and I think they are very	Understanding Personal Data

Ref	Comment	Sub-theme
	confused about the extent to which those developing public policy using their legislation compared to the private sector, using their data compared to those in the private sector.'	
3.2	00:13:54.71 'developing the ID card proposal was the approach of people who said that they didn't want government to have my data but were completely happy for Facebook and a variety of other organisations to have far more data that the government was ever going to ask for. That is one of the times when I came up against understandable concerns of people about what the nature of the data was' *It would have been useful in explaining to people what data was needed for an ID card system when compared with Facebook.*	Understanding Personal Data
3.3	00:13:54.71 'I think it would have been quite helpful to be able to categorise to have a simple categorisation of what it was that you needed to hold and where the limitations were that could be set down in legislation.'	Understanding Personal Data
3.4	00:16:16.47 'the fact that you know that somebody owns a particular mobile phone or has used a particular mobile phone at a particular moment or owns a particular laptop and has used it to access a particular web site or form of communication is distinct from knowing what in this mobile phone conversation they are talking about, in however guarded terms about carrying out a terrorist attack,' *In this case it would help to differentiate contents of communications and the meta data of those communications in order to explain to the public, in this case they would be two different footprints?*	Understanding Personal Data
3.5	00:20:44.12 'where a child has been subject to things like family breakup, use of drugs within the house, criminal, parents with a criminal record, all sorts, that sort of thing that you can imagine, where you have 4 or more of those, that you know, this is going onto who will be a criminal in the future, that is an extremely good determinant of those people who are going to become violent or criminals in the future. and therefore, how can you intervene how can you identify' *example of a digital persona*	Understanding Personal Data
3.6	00:21:26.32 'in a lot of legislation that I did the opposition rightly probing and challenging on amendments particularly in committee was what is this information going to be used for, who is collecting it, where is it being held, mm how is it going to be safeguarded, to what extent is, are individuals are going to be giving permission. So that is everything from adoption legislation, regulations about what you record in registers about kids,'	Understanding Personal Data
3.8	00:35:21.06 'I think the categorisation of data would be helpful but then alongside that the matrix if you like is nature of data / use of data So people might be willing, you know I am perfectly happy for this hospital or any other hospital to know a lot about me. But it is because I know it is being used for my health care. I would not want the same data to known by other public sector bodies'	Understanding Personal Data

Ref	Comment	Sub-theme
3.9	00:36:49.65 'the categorisation of the data is helpful but the context of its use is crucial'	Understanding Personal Data
4.1	00:01:49.63 'you can get lost the mystification of it is unbelievable' *wrt the terminology used when describing personal data*	Understanding Personal Data
4.2	00:02:54.16 'confusions about big data / metadata people just loose the plot' *There is a need in Government to differentiate between data and meta data in a clear and consistent way*	Understanding Personal Data
4.3	00:06:24.87 'if there was several people they wouldn't know which one had done it' *In relation to collecting Digital Footprints from devices such as a TV, phone, SIRI etc.*	Understanding Personal Data
4.4	00:12:38.60 'I think it would' *- in response to would the model help in legislation.*	Understanding Personal Data
4.7	'BLP 00:15:47.26 I find that interesting because there is an issue for me with CCTV cameras are they, the recordings, are they caused by the individual or are they caused by the person who put the camera there, now I personally think that it is the individual who walks through them who that creates their own digital footprint Interviewee 4 00:16:07.84 it has got to be, the other is just a mechanism for recording it BLP 00:16:13.25 but if you were to record inside somebody's house perhaps that would be a third party digital footprint? Interviewee 4 00:16:23.10 I think it would because they had no knowledge of our choice in that, if you walk through a railway station and you are in the concourse and it was being policed or there were uniform staff around and they could see what you were doing that is no different to, to the CCTV whereas in your own home you are going about personal business without any expectation that someone else will be registering it.' *Difference between digital footprints and third party digital footprints*	Understanding Personal Data
5.1	00:00:24.28 'There the one big difference would be whether the data would be about people like you or data about you' *The issue of whether data is actually about you or ascribed to you from group data for example from the use of data analytics*	Understanding Personal Data
5.12	Interviewee 5 00:33:43.86 Was it all on paper? Do you have transcripts of phone calls? BLP	Understanding Personal Data

Ref	Comment	Sub-theme
	00:34:02.30 No recordings made as no phone calls and no video was sent Interviewee 5 00:34:03.96 Ok, we got problems from 3UK and a couple of other phone companies *The question of telephone conversations and video recordings not being returned under my SARs*	
5.2	00:04:11.53 'That is one of the biggest problems' *wrt the ownership of the digitally extended self*	Understanding Personal Data
5.20	00:58:36.09 'a lot of information about you they will probably not see that as your information. They will probably have quite a strict idea of what is your information and what is their governments information.'	Understanding Personal Data
5.24	01:29:01.19 'so there is demographic as a class demographic' *the idea that you are considered a member of a class of individuals and that data about a number of that class is then attributed to the other members of the class*	Understanding Personal Data
5.3	00:07:25.51 'demographic information which is where we are finding the biggest problems in defining what is personal information' 'What is the demographic data, the detailed demographic data to the point where it can be used to effect decisions about you but it is not mm when the Department of Health shares data with the Society of Actuaries. The Society of Actuaries uses the data to re-adjust health premium for health insurance premiums. But in the end the individual, the hospital record is used to change the premium. What they did was to use it for people like you, as you are part of a class' *wrt Third Party Data and where the boundary of personal data lies*	Understanding Personal Data
5.4	00:07:25.51 'another area is anonymized data, so they say at the moment it is a computer record and in most cases it would be most appropriate to talk about pseudo anonymized because it is at a level to say that it is anonymized but one day it could be de-anonymized' *When does data become personal from impersonal (due to anonymisation)?*	Understanding Personal Data
5.5	00:12:55.70 'tyranny of the minority' *if the rest of your group provide data then there is enough to classify you - see Nissenbaum in Privacy and Big Data*	Understanding Personal Data
5.6	00:21:03.23 'even if they wanted to give you information do they understand what they have, and do they know what you mean … because the more sophisticated concepts around data constructs may be harder for them' *wrt analytics*	Understanding Personal Data

Ref	Comment	Sub-theme
6.1	00:10:42.74 'to have powers to enable you to limit what can and can't be done with certain parts of data in order to be able to do that you have to be able to define it'	Understanding Personal Data
6.5	00:17:15.80 'we were senior people in the IT industry and in discussions with them it did seem to me that very senior management did not completely understand and perhaps nobody really understands the detail of how the big data was being assembled and used'	Understanding Personal Data
6.6	00:17:15.80 'I think the trouble is you are walking into a tsunami of personal data which is getting thicker and thicker and bigger and bigger every year and it is probably, even if legislated and defined it is probably a bit unstoppable because it is very hard to police it'	Understanding Personal Data
7.3	00:05:03.19 'I think this is more of a big data thing. I think it is more a what you do with it kind of thing. It is more of an information scientist, I think they call them nowadays, they have invented new titles. So, they are kind of like data architects but not data architects. With data architects are more around how do we avoid duplication of data and how do we avoid that data getting out of sync and how do we share it between the applications whilst staying within the bounds of the law etc.' the model is applicable to information scientists, and the use of say big data, as opposed to data architects who are more interested in data duplication and how it is shared between applicants while staying within the law	Understanding Personal Data
7.4	00:22:17.51 'there is something to be said for not blocking people creating digital personas because if you haven't got a digital debt persona or a digital credit persona then people are going to go why has this person not got one. Are they somebody who has been doing fraud and regularly disappearing and starting up a new persona?'	Understanding Personal Data
7.5	00:24:12.40 'I think it is more about information science than architecture. Architecture is more concerned with creating these things'	Understanding Personal Data
7.7	00:33:26.53 'how useful was that stuff 10 years ago. I was in a different place then with a different amount of money, living in a different house.'	Understanding Personal Data
8.16	00:51:02.02 'No because they probably don't see your personal data, it is a different level isn't it. I mean they will be exposed to data but they won't recognise it as personal data because it is only personal when you are an identifiable by it. So, if your name and surname are not there automatically it is not your data.' *wrt app developers not realising that some of their data is personal data*	Understanding Personal Data
8.3	00:18:17.48	Understanding Personal Data

Ref	Comment	Sub-theme
	'In any company I would say I don't think we have any of these, number three <data descriptive of me disclosed to other parties>, because the information that we would have about you was collected for the purpose of you opening an account' *compare with comments later about credit checks and data to ?Experian? 8.5 21:47:99*	
8.5	00:21:47.99 'for the performance of your contract we are bound by law to actually run a credit search on you' *see 18:17:48* *when it is said this doesn't happen*	Understanding Personal Data
8.6	00:22:04.68 (1) 'I might know on that basis who we send it to but would I actually when you file this Subject Access Request would I actually go and say ooh but also let me mention that it has been sent'	Understanding Personal Data
8.8	00:22:04.68 (3) 'if I was to receive this in these terms obviously I would apply them as they are and I would have to go and look whether the data has actually gone and be very specific when I respond to you and say we do also send your data to da da da, in the performance of your contract da da da. Which of course then you would have to go to Experian if you wanted to know how they are managing your data etc. etc.' *Compare to 8.4 00:18:17:48*	Understanding Personal Data
9.1	00:04:19.63 (1) 'I don't think that we have too much of an issue with terminology.' *in relation to the model*	Understanding Personal Data
9.15	00:46:35.97 'Certainly, if I had any concerns about how my, about responsibly an organisation is using my data, then I think one of my questions would be I think where my information is being drawn from who else might have access to it. What exactly is going to be done with it. And I think my questions would extend to what is your data retention policy. How is it destroyed, when, and all those kind of questions.' *Note also a comment about data retention policy, matches a comment made by another Data Protection Officer*	Understanding Personal Data
9.2	00:04:19.63 (2) 'there is not too much that my department would actually do in terms of using data. It is all about gathering the information together that the department holds in order to then be able to respond to subject access requests.' *wrt not considering data structure but concentrating on retrieving known documents*	Understanding Personal Data
9.3	00:04:19.63 (3) ''I suppose in a way because it is quite an onerous process we don't get too bogged down in the terminology because although we recognise that our obligation is to provide access subject to any exemptions to the data rather than necessarily to copies of the records. The reality is that you end up providing copies of records because that is the easiest way of complying.'	Understanding Personal Data

T Interview thematic analysis summary

Interviewee	Identifier	1	6	7	2	8	9	3	4	5	Theme Total	Sub-Theme Total
	Area of Expertise	Ent	IT		DPO			Pol		TT		
Theme	**Sub-theme**											
Culture	Approach to SARs	4	7	1	4	2	7	2	0	0	27	
	Transparency	2	0	2	1	0	3	5	1	4	18	
	Customer Focus	3	0	0	1	0	0	1	0	1	6	
	Protective	0	0	0	0	0	0	1	0	4	5	
	Efficiency	0	1	2	0	0	0	0	0	0	3	59
People	Understanding Personal Data	2	3	4	5	5	4	8	5	9	45	
	Knowledge / Training	2	0	0	4	1	0	1	0	2	10	
	Trust	0	2	1	0	0	0	0	0	2	5	
	Common Requests	1	0	0	0	1	0	0	0	0	2	62
Capacity	Capability (IT or Otherwise)	0	1	9	0	1	4	1	0	2	18	
	Size	6	0	0	2	0	0	1	0	2	11	
	Processes	2	0	0	3	2	2	1	0	0	10	
	Structure	1	0	0	2	0	1	2	0	0	6	
	Competitive Situation	2	0	2	0	0	0	0	0	0	4	49
Governance	Practice	2	0	1	1	3	1	1	2	0	11	
	Mission & Vision	1	1	2	0	0	0	1	0	2	7	
	Disposition to DPA & SARs	1	0	0	1	1	0	0	0	0	3	21
Total		29	15	24	24	16	22	25	8	28	191	191

U Letter from The Office of National Statistics

Office for National Statistics

Segensworth Road
Titchfield
Fareham
Hants
PO15 5RR

Name: Martin Stringfellow
Tel: 01329 444012
Fax: 01633 652694
Email: martin.stringfellow@ons.gov.uk
www.ons.gov.uk

Mr Brian Parkinson
71 Bainton Road
OXFORD
OX2 7AG

28th March, 2014

Dear Mr Parkinson,

Thank you for your letter dated 1st March 2014 in which you repeat your request for details of the personal data held by Office for National Statistics about you.

You make some other points in support of your request. You say that you have been in touch with commercial organisations who state that they receive data from ONS. If any of these organisations are claiming to be receiving data from ONS for non-statistical purposes, I would like to know. This would be illegal under the Statistics and Registration Service Act 2007 (SRSA) and it is very important to ONS that nobody gets the impression that this is done. If you could let me have details of any organisation making this claim I will contact them to ask them to stop and ensure that any other misleading and damaging claims are stopped.

ONS information is indeed used by many organisations and individuals in this country and others. The vast majority of this information is published in an anonymised form as statistics which do not allow the identification of individuals. Other releases of data are to approved researchers for purely statistical or scientific research in a safe environment to allow them to produce carefully controlled statistical outputs.

I must draw your attention to section 7 of the SRSA which limits our functions to the production of statistics that serve the public good.

You are right to identify that merely removing names from data does not prevent the identity of the individual being rediscovered. ONS is keenly aware of this and takes great care to ensure that any information it publishes does not allow an individual's personal data to be identified and is acknowledged to be a world leader in statistical methodology designed to ensure that no identity or attribute related to an identity is ever released.

The importance of confidentiality is reinforced by the legislation that controls what ONS does, the Statistics and Registration Service Act 2007. Section 39 of that Act makes it an offence, punishable by up to two years imprisonment, to disclose personal information held by ONS except in tightly defined circumstances. These exceptions allow us to disclose personal information only where Parliament has said it is allowable. ONS only discloses personal information for statistical and scientific research purposes, not where it will be used to support decisions affecting particular individuals.

I would like to emphasise that ONS does not sell personal data.

Section 33 of the Data Protection Act does not forbid ONS from answering a request for personal data but it is ONS policy to resist any attempt to force it to disclose personal data for non-statistical purposes even from data subjects.

Because of this I am repeating my previous reply that we (as well as all other UK data controllers who process data for only scientific and statistical purposes) are exempted from the section 7 subject access provisions by section 33 of the Data Protection Act and will not process your request.

I have securely shredded the identity documents and cheque that you enclosed.

Yours Sincerely,

Martin Stringfellow
Legal Services
ONS

V Letter from Royal Mail

18th June 2014

Mr Brian Parkinson
The Anvil
The Street
Kingston
Lewes
BN7 3PB

Royal Mail Group

Information Rights Team
(Data Protection Act)
2nd Floor
Royal Mail Group
Pond Street
SHEFFIELD
S98 6HR

Tel: 0114 2414217
information.rights@royalmail.com
www.royalmail.com

Dear Mr Parkinson

Re: Subject Access Request (Our Reference: SAR10135-Parkinson)

I am writing in response to your Subject Access Request initially received by Royal Mail on 11th April 2014. In your letter you stated that you were a student undertaking a doctorate and that you were researching your own electronic records. On the 17th April we wrote to you to explain that Royal Mail Group is an extremely large organisation and as such we would require further information to be able to progress your request.

We asked if you could provide an indication of what parts of our business may hold information about you as, due to the size of the business, it is not possible to complete an open ended search of all records. We stated it would be useful if you could confirm what products or services you may have used or purchased, such as online postage, redirections or keepsafe, or if you have had reason to contact any of our Customer Service teams in Royal Mail or Parcelforce. We noted that it would be particularly useful if you could provide dates of any contact, or any reference numbers you hold. We also explained that you would need to contact Post Office Limited directly to make a request to them as Royal Mail Group is a separate organisation.

You wrote to us on the 13th May 2014 and stated;

> *"You ask for clarification. As I do not know what you hold it is difficult for me to be very specific. However I have used Recorded and Signed For services plus Parcelforce. Together with the Mail Preference Service to avoid junk mail. I received parcels and bulk postings from you."*

In light of this clarification we have sought information from the most likely parts of our businesses that may hold details about you as a customer. As we have explained, it is not possible to perform an open ended search. I can confirm that Royal Mail Group does not hold any personal data relating to you at any of the addresses specified, in any of the following;

Cont...

- Our Marketing Databases.
- Our Philatelic and Stamp Buying Database.
- We do not have record of you having registered as a customer on any of our Group websites.
- Our Parcelforce telephone booking system has no records of orders to or from your addresses.
- Our Parcelforce web booking system has no records of you or any of your addresses.
- Our Parcelforce complaints system has no record of any claims, calls, or correspondence.
- Our own database shows no records of any Freedom of Information requests made whilst we were still subject to the Act.

I can confirm that the only system on which your details were located is the one relating to Royal Mail complaints. I have enclosed a copy of the information held. Customer complaints are routinely logged on a central system along with any correspondence in relation to the case. Therefore the relevant information from this system has been provided to you.

Please note that it has been necessary to redact some of the enclosed information in order to withhold the personal data of third parties which is included within it. This does includes the names and contact details of Royal Mail employees. Under Section 7(4) of the Data Protection Act, where it is not possible to comply with a request without disclosing information relating to another individual, the data controller is not obliged to comply with the request unless the third party has consented to disclosure; or it is reasonable to release the information without consent. Our employees have an expectation of privacy and we do not therefore, believe it would be reasonable to disclose the redacted information in response to a Subject Access Request. Further, there is no obligation to seek consent and the information relating to third parties has therefore been redacted.

I hope that the enclosed information is helpful. If for any reason you are unhappy with this response, you can write to the Head of Information Rights, Royal Mail Group, 2nd Floor, Royal Mail, Pond Street, SHEFFIELD, S98 6HR.

You also have the right to complain to the Information Commissioner if you believe we have failed to comply with the requirements of the Data Protection Act. Further information about the Act and your rights as an individual is available from the Information Commissioner's Office at:

Information Commissioner's Office
Wycliffe House
Water Lane
WILMSLOW
SK9 5AF

Telephone: 01625 545 745

Cont...

www.ico.org.uk

In the event you continue to experience issues in relation to the delivery of your mail please contact our Customer Services Department who would be able to investigate the issued further. They can be contacted by telephone on 03457 740 740.

Yours sincerely,

Daniel Tulp
Information Rights Team
Royal Mail Group

W Extract of letter from Equifax

Dear All

Please find below the details for the latest Subject Access Request. Please compile all the relevant information and return to the **External Subject Access/uk** in box **as soon as possible** but no later than **2nd January 2014**

Name: Brian laurence Parkinson
DOB: 06/11/1948
Address: The Anvil, The Street, Kingston, Lewes, BN7 3PB
Previous: 71 Bainton Road, Oxford, OX2 7AG
Lane Ends Barn, Elterwater, Ambleside, Cumbria, LA22 9HN

Even if you have no information on the above individual please can you respond to this e-mail.

Kind Regards,

Customer Relations

X The research findings with respect to ability and willingness from the viewpoints of the interviewees

Interviewee 1, the Owner of a Tech Company compared governmental organisations with their love of the word no (unwilling) with John Lewis and M&S who are very focused on customer service (willing) as opposed to HMRC who are not as they are not in a competitive market. In the same interview he notes the ability of Facebook to collect and curate data, and the IOS app developers who have the ability but are unaware of their Data Protection Act responsibilities.

Interviewee 2, Member of the Management Group of the National Association of Data Protection Officers , suggested that local government was unable to pull data together due to the range of services provided but also that they were unwilling to collate the data as they tend to sit in their own ghettos.

Interviewee 3, Former Member of the UK Cabinet [1], had the following observations. Local government is badly organised and so can be positioned as having a low capability. Central government do not, in general have systems set up to respond to subject access request. Whilst maintaining that central government is more transparent than the private sector the interviewee also stated that freedom of information requests were treated as a higher priority. This suggests a low capability, but a basic willingness to respond to requests albeit more muted when answering subject access requests. On the other hand, it was intimated that private sector organisations were better structured to provide the necessary data, whilst charities had a culture of openness. Interviewee 3 also suggested that people may view private sector organisations as willing as they are less suspicious of the scope of the organisation (for instance we all believe we know what the scope of John Lewis' operation is) whereas people tend to suspicious of the scope of government organisations and therefore inherently belief that they are keeping something back (which of course they are, for instance details of the Warnings Index as the UK NO Fly List is called).

Interviewee 4, Former Member of the UK Cabinet [4], nothing relevant.

Interviewee 5, Think tank - policy director, indicated from their personal experience that charities had capacity issues, in comparison a large company has people who understand data protection, although it may be difficult to pull it all together from a distributed organisation. In government HMRC and DWP are the main data holders but HMRC foundation charter focuses on confidentiality and therefore a resistance to sharing data is cultural for that organisation. In addition, central government will often not see a lot of data which is descriptive of an individual as personal data and so will be unwilling to share it.

Interview 6, Recently retired IT Director of FTSE 100 company, states his belief that government is unable as they have worse quality staff that public companies and that they are unwilling due to lack of competition.

Interviewee 7, Senior IT Professional Magic Circle Law Firm, states from his experience in central government that they do not understand their data, and don't know how to use it (they are incompetent) as opposed to being unwilling, whereas public and private companies have a better understanding of their data and have the ability to use it. Utilities however are more akin to the government in that they don't know their data but are willing to provide it once it was signposted to them. It was summarised by saying that those who want to get money off you are the hottest at analysing data (able) and those that are not e.g. credit reference agencies, would tend to perform a little worse. Suggesting a continuum of ability relative to the need to know the customer in order to make money.

Interviewee 8, Data Protection Officer from a utility, believes that companies and NGOs would respond quickly as a compliance issue (i.e. they are willing due to a corporate culture of compliance) and that companies would have processes in place thus making them able. On the other hand, local and central government lack staff and processes thus making them unable and by inference unwilling. This is compounded by silos within organisations not talking to each other for instance the NHS. Compliance is seen as an enabler to a perceived open culture so for instance the FCA have strong sanctions within the Financial Services industry and so compliance is the norm, including for subject access requests, as a result these organisations will be able and appear willing. In other sectors is was considered that on-line shopping and charities both need data to exist and so would be expert in its storage and extraction although this would not necessarily explain their

willingness to provide data. App developers would also fall into this category in that they are expert in the collection and perhaps sale of data but it was suggested that they do not understand what personal data is and so do not have the ability to recognise it and so respond to subject access requests.

Interviewee 9, Data Protection Officer from central government, states that many departments have resource issues (e.g. HMRC, MOJ, UKBA) which affect their ability to respond to customers. (Note the customer service vocabulary) but also that HMRC have records going back to the 1960s on microfiche similarly affecting their ability to respond, whilst UKBA systems rely on manually printing screens one by one. Despite this there was surprise that central government were for at responding suggesting an internal viewpoint of willingness and success (despite the restrictions mentioned above). In comparison to other interviewees this person had occasion to raise their own subject access request but found the private organisation unhelpful the more so because it was located across international boundaries.

Y Mailchimp's initial response to data access request

From: **MailChimp Legal** legal@mailchimp.com **Subject:** Re: Personal Information Request

Date: 13 January 2014 17:59 **To:** Brian Parkinson mail@brian.parkinson.name

---

Hello Brian,

Thank you for reaching out to us.

Because of our commitment to Member privacy, we can't comment on, divulge information about, or block access to a Customer's account unless you submit to us a valid court order or subpoena from the State of Georgia. Please direct any documents to our legal team, which can be reached at legal@mailchimp.com.

If you wish to unsubscribe from any newsletters, or would like to report the receipt of unsolicited mail, please provide the full campaign headers to any/all campaigns received and we will be happy to take immediate action. Below you will find a link that outlines how to obtain that information. We may also globally block your email in our system, if you wish. This would prevent any MailChimp user from contacting you in the future.

http://mailchimp.com/contact/campaign-id/

Thank you,

~Legal Chimps

Z Letter from John Lewis - Follow-up Response

partnership
John Lewis Waitrose

PO Box 5137
Coventry, CV3 9EP

Telephone Number 0845 300 3833

Mr Brian Parkinson
The Anvil
The Street
Kingston
Lewes
East Sussex
BN7 3PB

03 June 2014 Our Ref: GN

Dear Mr Parkinson

Re: Subject Access Request – 5420 1138 4398 4578

Thank you for your recent communication.

We can confirm Data Analysis is not personal data; therefore, information around this would not be provided in a Data Subject Access Request.

The BA score is the Behavioural score and BK score is the Bankruptcy score. This is an in house scoring system used by John Lewis Financial Services and is not divulged outside this group. Your usage and payment history is reflected in the above scores.

Your account information is not transferred to John Lewis Stores; however, this is shared between departments within John Lewis Financial Services departments. We update Experian and Equifax with your account information, to obtain a copy of this, a request will need to be sent to the Experian or Equifax as this is not personal information held on our systems.

Your personal information is not given or sold to third parties for marketing without your consent, please be advised your agreement to our terms and conditions gives us consent to process your information.

Please accept my apologies as it seems that the full transaction history was not provide in the first instance, I have enclosed a list of all transactions applied to your account from December 2004.

Should you have any further queries, please contact *partnership card* services on 0845 300 3833.

Yours sincerely

Michelle Dean
Senior Manager

Enc

AA Letter from HM Revenue and Customs

HM Revenue & Customs

Data Protection Unit
PAYE & Self Assessment
Foyle House
Duncreggan Road
LONDONDERRY
BT48 0AH

Phone +44 (0) 3000 537277

Fax +44 (0) 3000 537407

www.hmrc.gov.uk

DX

Mr Brian L Parkinson
71 Bainton Road
OXFORD
OX2 7AG

Date 15 January 2014
Our ref DPU 23623/12/13
Your ref YL889475B

Dear Mr Parkinson

Thank you for your letter dated 09 December 2013 requesting personal information under the terms of the Data Protection Act 1998 (DPA).

Please find enclosed the following electronic records as held on HMRC's systems –

- PAYE personal details summary.
- PAYE case history notes, expanded where necessary.
- PAYE P14 pay, tax and NIC's details from 2007-08 to 2012-13.
- PAYE tax code calculations from 2007-08 to 2013-14.

- Self Assessment (SA) personal details summary.
- SA case history notes.
- Screen prints of your SA tax returns from 1996-97 to 2012-13 together with associated tax calculations.
- SA payments/credits history.
- SA repayments history.

Please be advised that due to constraints with HMRC's Data Security Policy, to remain compliant with DPA and in line with HMRC internal data retention regulations I am unable to respond to any of your questions pertaining to a detailed breakdown or analysis of data contained or used on HMRC systems.

You should be aware that the Data Protection Act provides for a number of exemptions to the disclosure of information. HMRC like all registered Data Controllers under the Act, are entitled to apply these exemptions where permitted.

I have enclosed our fact sheet which tells you more about data protection and Subject Access Requests.

BB Validation tables: Simple digital mosaic, full digital mosaic, digital persona and the digitally extended self

Validation Tables	Term from Literature	Usage	Example of Usage
2 Simple Digital Mosaic	2.1 digital dossier	2.1.1 dossiers compiled from a person's uploads	Gelman (2009)
	2.2 digital footprint	2.3.1 referring to the collection of digital footprints	Chretien et al. (2009) Ess (2009)
	2.3 digital mosaic	2.2.1 'He was a digital mosaic … storing his data in starfish satellites' p112	DeLillo (1991)
		2.2.2 Google search terms used by an individual and their associated data	Floridi (2006b)
		2.2.3 'our transactions, our media consumption, our locations and travel, our communications, and our relationships' p2	Wittes (2011)

Table BB.1. Validation 2: Simple Digital Mosaic - mapping of literature to the model.

Term in Model	Term from Literature	Usage	Example of Usage
3 Full Digital Mosaic	3.1 data shadow	3.1.1 combination of digital artefacts	Westin (Westin, 1967) Garfinkel (2000) Floridi (2005)
		3.1.2 'records and data about the self' p167	Smithson (1985)
	3.2 digital biography	3.2.1 'an electronic collage' p1394, 'a life captured in records' p1394, 'bits and pieces of stored information about one's life' p70	Solove (2001) Ploeg (2003)
	3.3 digital doppelganger	3.3.1 a collection of digital artefacts which provide a picture of a life	Cherry (2005)
	3.4 digital dossier	3.4.1 collections of footprints e.g. from Facebook	Gross & Acquisti (2005) Garfinkel (2000) Solove (2004)
	3.5 digital footprint	3.5.1 digital artefacts some known to us others not	Sellen et al. (2009)
	3.6 digital identity	3.6.1 aggregated data about an individual, but only that which are publically available	Palfrey & Gasser (2008)

Term in Model	**Term from Literature**	**Usage**	**Example of Usage**
	3.7 digital mosaic	3.7.1 individual searches by law enforcement agencies may not intrude an individual's privacy but multiple ones produce a mosaic of information which can be a breach of privacy	Dennis (2012)
		3.7.2 a collection of artefacts which can present an image of an artist to a fan e.g. YouTube, Twitter	Hanna et al. (2011)
		3.7.3 mosaic of information and analyses that create a picture of a company	Schwartau (1994)
3 Full Digital Mosaic (continued)	3.8 digital person	3.8.1 'a life captured in records' p1	Solove (2004)
	3.9 digital persona	3.9.1 persona created by postings onto the internet - does not consider analyses of these postings	Clark (2010)

Term in Model	Term from Literature	Usage	Example of Usage
		3.9.2 'each digital persona is defined by the combination of profile data captured by the person and others into one or more SNS [social networking systems]' p. 11	Clarke (2008)
	3.10 dossier	3.10.1 aggregated data about an individual	Solove (2006)
		3.10.2 aggregated data about individual, includes data not publically available	Palfrey & Gasser (2008)

Table BB.2. Validation 3: Full Digital Mosaic - mapping of literature to the model.

Term in Model	Term from Literature	Usage	Example of Usage
4 Digital Persona	4.1 digital biography	4.1.1 'bits and pieces of stored information about my life and behavior, an embodied identity' p70	Ploeg (2003)
		4.1.2 data and profiles	Solove (2004)
	4.2 digital - doppelganger, digital self , second self	4.2.1 focuses on data from social networks and data aggregation	Andrews (2013)
	4.3 digital persona	4.3.1 describes projected and imposed personae but does not explicitly include profile data, but does consider context	Ardagna et al. (2010)
		4.3.2 analysis of data especially transaction generated data	Blanchette & Johnson (2002)
		4.3.3 personas derived from profiling and data mining	Hildebrandt & Gutwirth (2008)
	4.4 digital persona, data shadow, digital individual	4.4.1 'the digital persona is a model of the individual established through the collection, storage and analysis of data about that person' p1	Clarke (1993)

Term in Model	Term from Literature	Usage	Example of Usage
	4.5 digital personality profile	4.5.1 'aggregating, analyzing, or 'mining' personal information, when it is or can be used to uniquely identify, locate, or contact that person' p142	Ludington (2006)
	4.6 ersatz double	4.6.1 Facebook profiles and postings	Sanchez (2009)
	4.7 online identity, digital self	4.7.1 does not explicitly allow for the inclusion of profile data in further profiles	Briggs (2013)

Table BB.3. Validation 4: Digital Persona - mapping of literature to the model.

Term in Model	Term from Literature	Usage	Example of Usage
5 Digitally Extended Self	5.1 digital doppelganger	5.1.1 a similar concept but constrained to social networking data	Andrews (2013)
	5.2 digital dossier	5.2.1 'Taken together, all the digital information held, in many different hands, about a given person' p39	Palfrey & Gasser (2008)
	5.3 digital persona	5.3.1 analysis of transactional data combined with other records e.g. demographics	Blanchette & Johnson (2002)
	5.4 virtual self	5.4.1 analyses computed by marketing companies and government departments augmented by further transactions	Lyon (2000)

Term in Model	Term from Literature	Usage	Example of Usage

Table BB.4. Validation 5: Digitally extended self - mapping of literature to the model.

List of References

Abu-Laban, Y. and Nath, N. (2007). From deportation to apology: The case of Maher Arar and the Canadian state. *Canadian Ethnic Studies*, 39(3), pp. 71-98.

Acquisti, A. (2004). Privacy in electronic commerce and the economics of immediate gratification. In: *Proceedings of the 5th ACM Conference on Electronic Commerce,* New York. Available at: http://www.academia.edu/download/30777892/Acquisti04.pdf [Accessed 2nd March 2018].

Acquisti, A., Brandimarte, L. and Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221), pp. 509-514.

Agger, B. (2008). *The Virtual Self: A Contemporary Sociology*. Blackwell Publishing Ltd, Oxford.

Akalu, R. (2018). Privacy, consent and vehicular ad hoc networks (VANETs). *Computer Law & Security Review*, 34(1), pp. 37-46.

Ali, M.A. and Mann, S. (2013). The inevitability of the transition from a surveillance-society to a veillance-society: Moral and economic grounding for sousveillance. In: *2013 IEEE International Symposium on Technology and Society (ISTAS),* Toronto. Available at: http://ieeexplore.ieee.org/document/6613126/ [Accessed 20th March 2018].

Allen, A.L. (1988). *Uneasy Access: Privacy for Women in a Free Society*. Rowman & Littlefield Publishers Inc, Lanham, MD.

Altheide, D.L. (1987). Reflections: Ethnographic content analysis. *Qualitative sociology*, 10(1), pp. 65-77.

Anderson, J., Stender, M., Myers West, S. and York, J.C. (2016). *Unfriending Censorship: Insights from Four Months of Crowdsourced Data on Social Media Censorship*. Available at: Onlinecensorship.org_Report_-_31_March_2016.pdf [Accessed 24th February 2018].

Anderson, R., Brown, I., Dowty, T., Inglesant, P., Heath, W. and Sasse, A. (2009). *Database State: A Report Commissioned by the Joseph Rowntree Reform Trust*. The Joseph Rowntree Reform Trust, London.

Andrejevic, M. (2014). The big data divide. *International Journal of Communication*, 8, pp. 1673-1689.

Andrews, L. (2013). *I Know Who You Are and I Saw What You Did: Social Networks and the Death of Privacy*. Free Press, New York, NY.

Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016). *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks.* Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Accessed 18th February 2018].

Apple (2018). *Differential Privacy*. Available at: https://images.apple.com/privacy/docs/Differential_Privacy_Overview.pdf [Accessed 6th March 2018].

Apthorpe, N., Reisman, D., Sundaresan, S., Narayanan, A. and Feamster, N. (2017). Spying on the smart home: privacy attacks and defenses on encrypted IoT traffic. *arXiv:1708.05044v1 [cs.CR]*,

Archer, S. (2018). *Google Could be the First $1 Trillion Company if It Told Us More About What It Does (GOOG, GOOGL)*. Available at: http://markets.businessinsider.com/news/stocks/google-stock-price-could-be-the-first-1-trillion-company-if-it-told-us-what-it-does-2018-2-1014971527 [Accessed 1st March 2018].

Ardagna, C.A. et al. (2010). Exploiting cryptography for privacy-enhanced access control: a result of the prime project. *Journal of Computer Security*, 18(1), pp. 123-160.

Armstrong, D., Gosling, A., Weinman, J. and Marteau, T. (1997). The place of inter-rater reliability in qualitative research: an empirical study. *Sociology*, 31(3), pp. 597-606.

Ausloos, J. (2012). The 'Right to be Forgotten' – worth remembering. *Computer Law & Security Review*, 28(2), pp. 143-152.

Australian Government (1988). *Privacy Act 1988*. Available at: https://www.legislation.gov.au/Details/./ccd82b75-6a6a-4ece-87f0-9d74669c38fe [Accessed 24th February 2018].

Ayer, A.J. (2001). *Language, Truth, and Logic*. Penguin, London.

Bachmann, R. (2001). Trust, power and control in trans-organizational relations. *Organization Studies*, 22(2), pp. 337-365.

Baker, S. (2008). *The Numerati: How They'll Get My Number and Yours*. Jonathan Cape, London.

Bakshi, H. (2016). How can we measure the modern digital economy? *Significance*, 13(3), pp. 6-7.

Balka, E. and Star, S.L. (2015). Mapping the body across diverse information systems: Shadow bodies and how they make us human. In: *Boundary Objects and Beyond: Working with Leigh Star,* (Eds, Bowker, G.C., Timmermans, S., Clarke, A.E. and Balka, E.). MIT Press, Cambridge, MA, pp. 417-434.

Barber, B. (1983). *The Logic and Limits of Trust*. Rutgers University Press, New Brunswick, NJ.

Barbour, R.S. and Schostak, J. (2005). Interviewing and focus groups. In: *Research Methods in the Social Sciences,* (Eds, Somekh, B. and Lewin, C.). SAGE Publications Ltd, London, pp. 41-48.

Barnes, S.B. (2006). A privacy paradox: Social networking in the United States. *First Monday,* 11(9). Available at: http://journals.uic.edu/ojs/index.php/fm/article/view/1394/1312 [Accessed 21st March 2018].

Barth, S. and de Jong, M.D.T. (2017). The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telematics and Informatics*, 34(7), pp. 1038-1058.

Batchelor, R., Bobrowicz, A., Mackenzie, R. and Milne, A. (2012). Challenges of ethical and legal responsibilities when technologies' uses and users change: Social networking sites, decision-making capacity and dementia. *Ethics and Information Technology*, 14(2), pp. 99-108.

Batra, S. (2014). Big data analytics and its reflections on DIKW hierarchy. *Review of Management*, 4(1/2), pp. 5-17.

Bauman, Z. and Lyon, D. (2013). *Liquid Surveillance: A Conversation*. Polity Press, Cambridge.

Bell, C.G., Gemmell, J. and Haag, J. (2009). *Total Recall: How The E-Memory Revolution Will Change Everything*. Dutton, New York, NY.

Belyaev, K., Sun, W., Ray, I. and Ray, I. (2018). On the design and analysis of protocols for personal health record storage on personal data server devices. *Future Generation Computer Systems*, 80, pp. 467-482.

Benhabib, S. (1999). Sexual difference and collective identities: The new global constellation. *Signs*, 24(2), pp. 335-361.

Bentham, J. (1995). *The Panopticon Writings*. Verso, London.

Beresford, A.R., Kübler, D. and Preibusch, S. (2012). Unwillingness to pay for privacy: A field experiment. *Economics Letters*, 117(1), pp. 25-27.

Bessière, K., Seay, A.F. and Kiesler, S. (2007). The ideal elf: Identity exploration in World of Warcraft. *Cyberpsychology and Behavior*, 10(4), pp. 530-535.

Bhaskar, R. (1998). *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences*. Routledge, London.

Binns, R., Zhao, J., Van Kleek, M., Shadbolt, N., Liccardi, I. and Weitzner, D. (2017). My bank already gets this data. In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems,* Denver, CO. Available at: http://dx.doi.org/10.1145/3027063.3053255 [Accessed 27th February 2018].

Blaikie, N. (2009). *Approaches to Social Enquiry*. Polity Press, Cambridge.

Blanchette, J.F. and Johnson, D.G. (2002). Data retention and the panoptic society: The social benefits of forgetfulness. *The Information Society*, 18(1), pp. 33-45.

Bloustein, E.J. (1964). Privacy as an aspect of human dignity: An answer to Dean Prosser. *New York University Law Review*, 39, pp. 962-1007.

Blume, P. (2014). EU adequacy decisions: The proposed new possibilities. *International Data Privacy Law*, 5(1), pp. 34-39.

Brady, M.K. and Cronin, J.J. (2001). Customer orientation: Effects on customer service perceptions and outcome behaviors. *Journal of Service Research*, 3(3), pp. 241-251.

Brandeis, L.D. (1928). Olmstead V. U.S., 277 U.S.438, 478 (Mr. Justice Brandeis, Dissenting).

Brandeis, L.D. (1913). What publicity can do. *Harper's Weekly - 20th December*, pp. 10-13.

Breaux, T.D., Smullen, D. and Hibshi, H. (2015). Detecting repurposing and over-collection in multi-party privacy requirements specifications. In: *2015 IEEE 23rd International Requirements Engineering Conference,* Ottawa, ON, Canada. Available at: https://pdfs.semanticscholar.org/6755/b3926ad5ff03c04eed5e06c78f9d6a12f187.pdf [Accessed 19th March 2018].

Briggs, P. (2013). *Future Identities: Changing Identities in the UK – the Next 10 Years. DR 4: Will an Increasing Element of Our Identity be 'Devolved' to Machines?* The Government Office for Science, London.

Brin, D. (1999). *The Transparent Society: Will Technology Force Us to Choose Between Privacy and Freedom*. Perseus Books Group, Cambridge, MA.

Bryman, A. (2008). *Social Research Methods*. Oxford University Press, Oxford.

Butler, J. (1990). Gender trouble, feminist theory, and psychoanalytic discourse. In: *Feminism/Postmodernism,* (Ed, Nicholson, L.J.). Routledge, London, pp. 324-340.

Byron, T. (2008). *Safer Children in a Digital World: The Report of the Byron Review: Be Safe, Be Aware, Have Fun*. Department for Children, Schools and Families, London.

Caiza, J.C., Martín, Y.-S., Del Alamo, J.M. and Guamán, D.S. (2017). Organizing design patterns for privacy. In: *Proceedings of the 22nd European Conference on Pattern Languages of Programs,* Irsee, Germany. Available at: http://dx.doi.org/10.1145/3147704.3147739 [Accessed 19th March 2018].

Callero, P.L. (2003). The sociology of the self. *Annual Review of Sociology*, 29, pp. 115-134.

Castells, M. (1996). *The Information Age: Economy, Society and Culture: The Rise of the Network Society (Vol. 1)*. Blackwell Publishers Ltd, Oxford.

Chang, A.Y.C., Parrales, M.E.P., Jimenez, J., Sobieszczyk, M.E., Hammer, S.M., Copenhaver, D.J. and Kulkarni, R.P. (2009). Combining Google Earth and GIS mapping technologies in a dengue surveillance system for developing countries. *International Journal of Health Geographics*, 8, pp. 49-59.

Chaum, D. (1985). Security without identification: transaction systems to make big brother obsolete. *Communications of the ACM*, 28(10), pp. 1030-1044.

Cheong, M. and Lee, V. (2009). Integrating web-based intelligence retrieval and decision-making from the twitter trends knowledge base. In: *Proceedings of the 2nd ACM Workshop on Social Web Search and Mining,* Hong Kong, China. Available at: https://dl.acm.org/citation.cfm?id=1651439 [Accessed 21st March 2018].

Cherry, S. (2005). Total Recall (life recording software). *Spectrum, IEEE*, 42(11), pp. 24-30.

Chodorow, N.J. (1995). Gender as a personal and cultural construction. *Signs*, 20(3), pp. 516-544.

Choi, H., Park, J. and Jung, Y. (2018). The role of privacy fatigue in online privacy behavior. *Computers in Human Behavior*, 81, pp. 42-51.

Chretien, K.C., Greysen, S.R., Chretien, J.P. and Kind, T. (2009). Online posting of unprofessional content by medical students. *The Journal of the American Medical Association*, 302(12), pp. 1309-1315.

Cichy, P. and Salge, T.-O. (2017). Creating value from personal data – on the legitimacy of business practices in the field of internet-enabled services. In: *International Conference on Information Systems 2017,* Seoul. Available at: http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1102&context=icis2017 [Accessed 19th March 2018].

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press, New York.

Clark, J.E. (2010). The digital imperative: making the case for 21st century pedagogy. *Computers and Composition*, 27(1), pp. 27-35.

Clarke, R. (1993). Computer matching and digital identity. In: *Proceedings of the Third ACM Computers, Freedom & Privacy Conference,* Burlingame, CA. Available at: http://cpsr.org/prevsite/conferences/cfp93/clarke.html/ [Accessed 19th March 2018].

Clarke, R. (2008). Web 2.0 as syndication. *Journal of Theoretical and Applied Electronic Commerce Research*, 3(2), pp. 30-43.

Clemons, E.K., Barnett, S. and Appadurai, A. (2007). The future of advertising and the value of social network websites: Some preliminary examinations. In: *Proceedings of the Ninth International Conference on Electronic Commerce,* Minneapolis, MN. Available at: https://dl.acm.org/citation.cfm?id=1282153 [Accessed 19th March 2018].

Commission on Funding of Care and Support (2011). *Fairer Care Funding*. The Stationery Office, London.

Companies House (2018). *Incorporated Companies in the UK: October to December 2017*. Available at: https://www.gov.uk/government/publications/incorporated-companies-in-the-uk-october-to-december-2017/incorporated-companies-in-the-uk-october-to-december-2017 [Accessed 16 Feb 2018].

Cook, K. (2001). Trust in society. In: *Trust in Society: Volume II,* (Ed, Cook, K.). Russell Sage Foundation, pp. xi-xxviii.

Cooley, T.M. (1879). *A Treatise on the Law of Torts or the Wrongs which arise Independent of Contract*. Callaghan, Chicago, IL.

Corbin, J. and Strauss, A. (2014). *Basics of Qualitative Research*. SAGE Publications Ltd., London.

Council of Europe (1950). *European Convention on Human Rights*. Available at: https://www.echr.coe.int/Documents/Convention_ENG.pdf [Accessed 23rd March 2018].

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, and repealing Directive 95/46. *Official Journal of the European Union (OJ)*, 59, pp. 1-88.

Crain, M. (2018). The limits of transparency: Data brokers and commodification. *New Media & Society*, 20(1), pp. 88-104.

Craswell, G. and Poore, M. (2012). *Writing for Academic Success*. SAGE Publications Ltd, London.

Crenshaw, K.W. (1993). Beyond racism and misogyny: Black feminism and 2 Live Crew. In: *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment,* (Eds, Matsuda, M.J., Lawrence, C.R., Delgado, R. and Crenshaw, K.W.). Westview Press, Boulder, CO, pp. 246-263.

Crompton, M. (2010). User-centric identity management: An oxymoron or the key to getting identity management right? *Information Polity: The International Journal of Government & Democracy in the Information Age*, 15(4), pp. 291-297.

Crooks, E. (2017). *GE Loses Crown as Biggest US Manufacturer by Market Cap*. Available at: https://www.ft.com/stream/42dab372-28d1-364c-b4f2-fc7c4a07e906 [Accessed 1st March 2018].

Custers, B. and Uršič, H. (2016). Big data and data reuse: A taxonomy of data reuse for balancing big data benefits and personal data protection. *International Data Privacy Law*, 6(1), pp. 4-15.

Daskal, J.C. (2018). Borders and Bits. *Vanderbilt Law Review*, 71(1), pp. 179-239.

de Durand, E.T. (2008). *Trust: The Bedrock Principle of Good Governance*. Available at: http://ut.suagm.edu/sites/default/files/uploads/Centro-Gobernanza/Articulos/Trust-The-Bedrock-Principle-of-Good-Governance.pdf [Accessed 4th March 2018].

du Gay, P., Evans, J. and Redman, P. (2000). General introduction. In: *Identity: A Reader,* (Eds, du, G., Paul, Evans, J. and Redman, P.). Sage Publications Ltd, London

De Hert, P. and Gutwirth, S. (2006). Privacy, data protection and law enforcement. opacity of the individual and transparency of power. In: *Privacy and The Criminal Law,* (Eds, Claes, E., Anthony, D. and Gutwirth, S.). Intersentia, Oxford, pp. 61-104.

Degli Esposti, S. (2014). When big data meets dataveillance: The hidden side of analytics. *Surveillance & Society*, 12(2), pp. 209.

DeLillo, D. (1991). *Mao II*. Viking, New York, NY.

Dennis, E.S. (2012). Mosaic shield: Maynard, the Fourth Amendment, and privacy rights in the digital age. *Cardozo Law Revue*, 33, pp. 738 - 771.

Denshire, S. (2014). On auto-ethnography. *Current Sociology*, 62(6), pp. 831-850.

Department of Education (2018). *Longitudinal Education Outcomes Study: How We Use and Share Personal Data*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/664729/LEO_privacy_notice_Dec17_v1.0.pdf [Accessed 18th March 2018].

Desouza, K.C. and Jacob, B. (2017). Big data in the public sector: Lessons for practitioners and scholars. *Administration & Society*, 49(7), pp. 1043-1064.

Deutsch, A.L. (2014). *WhatsApp: The Best Facebook Purchase Ever?* Available at: https://www.investopedia.com/articles/investing/032515/whatsapp-best-facebook-purchase-ever.asp [Accessed 24th February 2018].

Dong, X., Guo, B., Duan, X., Shen, Y., Zhang, H. and Shen, Y. (2016). DSPM: A platform for personal data share and privacy protect based on metadata. In: *2016 13th International Conference on Embedded Software and Systems,* Chengdu. Available at: http://dx.doi.org/10.1109/icess.2016.10 [Accessed 19th March 2018].

Dötzer, F. (2005). Privacy issues in vehicular ad hoc networks. In: *Privacy Enhancing Techologies, 5th International Workshop,* Cavtat, Croatia. Available at: https://link.springer.com/chapter/10.1007/11767831_13 [Accessed 20th March 2018].

Dunne, M., Pryor, J. and Yates, P. (2005). *Becoming a Researcher: A Research Companion for the Social Sciences*. Open University Press, Maidenhead.

Elshtain, J.B. (1993). *Public Man, Private Woman: Women in Social and Political Thought*. Princeton University Press, Princeton, NJ.

Elvy, S.-A. (2017). Paying for privacy and the personal data economy. *Columbia Law Review*, 117(6), pp. 1369-1459.

EPIC (2014). *Comments of the Electronic Privacy Information Center to the Office of Science and Technology Policy*. Available at: https://www.epic.org/privacy/big-data/EPIC-OSTP-Big-Data.pdf [Accessed 28th February 2018].

Ess, C. (2009). *Digital Media Ethics*. Polity Press, Cambridge.

Etzioni, A. (1999). *The Limits of Privacy*. Basic Books, New York, NY.

European Commission (2016). *The EU-U.S. Privacy Shield*. Available at: https://ec.europa.eu/info/law/law-topic/data-protection/data-transfers-outside-eu/eu-us-privacy-shield_en [Accessed 21st March 2018].

European Commission (2018). *What is Personal Data*. Available at: https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en [Accessed 24th August 2018].

European Union (2016). General Data Protection Regulation (EU) 2016/679. *Official Journal of the European Union*, L 119, pp. 1-88.

European Union (2018). Directive (EU) 2016/680 of the European Parliament and of the Council. *Official Journal of the European Union*, L 119, pp. 89-131.

Fanucci, F. (2008). Access to information country study - United Kingdom. In: *Arab-European Dialogue Working Group 'Access to Information and Human Rights',* Cairo, Egypt. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2057609 [Accessed 20th March 2018].

Fife, R. (2018). *Up to 100,000 Canadians could be affected by no-fly list, research suggests*. Available at: https://www.theglobeandmail.com/news/politics/up-to-100000-canadians-could-wrongly-be-on-no-fly-list-research-suggests/article37299604/ [Accessed 24th August 2018].

Floridi, L. (1999). Information ethics: On the philosophical foundation of computer ethics. *Ethics and Information Technology*, 1(1), pp. 33-52.

Floridi, L. (2005). The ontological interpretation of informational privacy. *Ethics and Information Technology*, 7(4), pp. 185-200.

Floridi, L. (2006a). Four challenges for a theory of informational privacy. *Ethics and Information Technology*, 8(3), pp. 109-119.

Floridi, L. (2006b). Word of mouse. *The Philosophers' Magazine*, 33, pp. 17.

Floridi, L. (2008). Information ethics: A reappraisal. *Ethics and Information Technology*, 10(2), pp. 189-204.

Foucault, M. (1991). *Discipline and Punish: The Birth of the Prison*. Penguin, London.

Foucault, M. (1998a). *Technologies of the Self: A Seminar with Michel Foucault*. University of Massachusetts Press, Amherst, MA.

Foucault, M. (1998b). *The History of Sexuality: The Will to Knowledge*. Penguin, London.

Foucault, M. and Gordon, C. (2002). *Michael Foucault: Power Essential Works of Foucault 1954-1984*. Penguin, London.

Francis, P., Eide, S.P. and Munz, R. (2017). Diffix: High-utility database anonymization. In: *Annual Privacy Forum 2017,* Vienna. Available at: http://dx.doi.org/10.1007/978-3-319-67280-9_8 [Accessed 20th March 2018].

Fried, C. (1970). *An Anatomy of Values*. Harvard University Press, Cambridge, MA.

Friedman, M. (2007). The social responsibility of business is to increase its profits. In: *Corporate Ethics and Corporate Governance,* (Eds, Zimmerli, W.C., Holzinger, M. and Richter, K.). Springer, Berlin, pp. 173-178.

Friman, M., Gärling, T., Millett, B., Mattsson, J. and Johnston, R. (2002). An analysis of international business-to-business relationships based on the commitment-trust theory. *Industrial Marketing Management*, 31(5), pp. 403-409.

Fromm, E. (1994). *Escape From Freedom*. Henry Holt & Company Inc, New York, NY.

Fromm, E. (2002). *The Sane Society*. Routledge, Abingdon, Oxfordshire.

Gabel, D. and Hickman, T. (2016). *Unlocking the EU General Data Protection Regulation*. Available at: https://www.whitecase.com/publications/article/gdpr-handbook-unlocking-eu-general-data-protection-regulation [Accessed 6th March 2018].

Gamble, A. and Kelly, G. (2001). Shareholder value and the stakeholder debate in the UK. *Corporate Governance: An International Review*, 9(2), pp. 110-117.

Gantz, J. and Reinsel, D. (2010). The digital universe decade -are you ready. *http://idcdocserv. com/925,* Available at: https://ci.nii.ac.jp/naid/10031099270 [Accessed 2nd March 2018].

Gantz, J. and Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC analyze the future,* Available at: https://www.emc-technology.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf [Accessed 2nd March 2018].

Garfinkel, S. (2000). *Database Nation: The Death of Privacy in the 21st Century*. O'Reilly Media, Inc., Sebastopol, CA.

Gavison, R. (1980). Privacy and the limits of law. *The Yale Law Journal*, 89(3), pp. 421-471.

Gawer, A. (2017). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, 43(7), pp. 1239-1249.

Gefen, D., Karahanna, E. and Straub, D.W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, 27(1), pp. 51-90.

Gelman, L. (2009). Privacy, free speech, and blurry-edged social networks. *Boston College Law Review*, 50(5), pp. 1315-1344.

Giddens, A. (1991). *The Consequences of Modernity*. Stanford University Press, Stanford, CA.

Gieseke, J. (2014). *The History of The Stasi : East Germany's Secret Police, 1945-1990*. Berghahn Books, Oxford.

Gill, M. and Spriggs, A. (2005). *Assessing the Impact of CCTV*. Home Office Research, Development and Statistics Directorate, London.

Gill, O. (2018). *Apple, Amazon and Alphabet in Race to $1 trillion Market Cap as Tech Titans Post Massive Sales Numbers*. Available at: http://www.cityam.com/279934/apple-amazon-and-alphabet-race-1-trillion-market-cap-tech [Accessed 1st March 2018].

Glikman, P. and Glady, N. (2015). *What's the Value of Your Data?* Available at: https://techcrunch.com/2015/10/13/whats-the-value-of-your-data/ [Accessed 25th February 2018].

Glushko, R.J., Annechino, R., Hemerly, J. and Wang, L. (2013). Categorization: Describing resource classes and type. In: *The Discipline of Organising,* (Ed, Glushko, R.J.). O'Reilly, Sebastopol, CA, pp. 235-272.

Gobetti, D. (1992). *Private and Public: Individuals, Households, and Body Politic in Locke and Hutcheson*. Routledge, London.

Godfrey, P.C., Merrill, C.B. and Hansen, J.M. (2009). The relationship between corporate social responsibility and shareholder value: An empirical test of the risk management hypothesis. *Strategic Management Journal*, 30(4), pp. 425-445.

Goffman, E. (1971). *The Presentation of Self in Everyday Life*. Pelican, Harmondsworth.

Goffman, E. (1990). Role distance. In: *Life as Theater: A Dramaturgical Sourcebook,* (Eds, Brissett, D. and Edgley, C.). Aldine de Gruyter, New York, NY, pp. 101-111.

GoGlobe (2018). *Things That Happen on Internet Every 60 Seconds (Updated 2017)*. Available at: https://www.go-globe.com/blog/things-that-happen-every-60-seconds/ [Accessed 8th March 2018].

Google (2018a). *Requests for User Information*. Available at: https://transparencyreport.google.com/user-data/overview [Accessed 3rd March 2018].

Google (2018b). *Search Removals Under European Privacy Law*. Available at: https://transparencyreport.google.com/eu-privacy/overview [Accessed 6th March 2018].

Gopal, R., Hidaji, H., Patterson, R.A., Rolland, E. and Zhdanov, D. (2018). How much to share with third parties? A website's dilemma and users' privacy concerns. *MIS Quarterly*, 42(1), pp. 143-164.

Grayling, A.C. (2009). *Liberty in the Age of Terror: A Defence of Civil Liberties and Enlightenment Values*. Bloomsbury, London.

Greenhalgh, H. (2015). *Personal Data - Your Digital Life in Their Hands*. Available at: https://www.ft.com/content/9046ec52-8168-11e5-8095-ed1a37d1e096 [Accessed 6th March 2018].

Greenleaf, G. (2012). The influence of European data privacy standards outside Europe: implications for globalization of Convention 108. *International Data Privacy Law*, 2(2), pp. 68-92.

Greysen, S.R., Kind, T. and Chretien, K.C. (2010). Online professionalism and the mirror of social media. *Journal of General Internal Medicine*, 25(11), pp. 1227-1229.

Gross, R. and Acquisti, A. (2005). Information revelation and privacy in online social networks. In: *ACM Workshop on Privacy in the Electronic Society,* Available at: http://inference-reseaux-sociaux.googlecode.com/files/Information Revelation and Privacy in Online Social Networks - Document.PDF

Guerin, B., McCrae, J. and Shepheard, M. (2018). *Accountability in modern government: What are the issues?* Institute for Government, London.

Guild, E., Bigo, D. and Carrera, S. (2017). Trump's travel bans: Harvesting personal data and requiem for the EU-US Privacy Shield. *CEPS Policy Insights*, 2017(13), pp. 1-9.

Gutwirth, S. (2002). *Privacy and the Information Age*. Rowman & Littlefield, Lanham, MD.

Hadar, I., Hasson, T., Ayalon, O., Toch, E., Birnhack, M., Sherman, S. and Balissa, A. (2018). Privacy by designers: Software developers' privacy mindset. *Empirical Software Engineering*, 23(1), pp. 259-289.

Haddaway, N.R., Collins, A.M., Coughlin, D. and Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PLOS one*, 10(9), pp. 1-17.

Haggerty, K.D. and Ericson, R.V. (2003). The surveillant assemblage. *The British Journal of Sociology*, 51(4), pp. 605-622.

Halbert, D.J. (2009). Public lives and private communities: The terms of service agreement and life in virtual worlds. *First Monday,* 14(12). Available at: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2601/2405 [Accessed 20th March 2018].

Hankin, C et al. (2013). *Foresight Future Identities (2013): Final Project Report.* The Government Office for Science, London.

Hanna, R., Rohm, A. and Crittenden, V.L. (2011). We're all connected: The power of the social media ecosystem. *Business Horizons*, 54(3), pp. 265-273.

Hardin, R. (2002). *Trust and Trustworthiness*. Russell Sage Foundation Publications, New York, NY.

Hardy, K. and Maurushat, A. (2017). Opening up government data for big data analysis and public benefit. *Computer Law & Security Review*, 33(1), pp. 30-37.

Hartzog, W. and Rubinstein, I. (2017). The anonymization debate should be about risk, not perfection. *Communications of the ACM*, 60(5), pp. 22-24.

Heeney, C. (2012). Breaching the contract? Privacy and the UK census. *The Information Society*, 28(5), pp. 316-328.

Heinderyckx, F. (2014). Reclaiming the high ground in the age of onlinement: ICA Presidential Address, 2014. *Journal of Communication*, 64(6), pp. 999-1014.

Hengstler, J. (2011). Managing your digital footprint: Ostriches v. eagles. *Education For A Digital World*, 1, pp. 89-139.

Henttonen, P. (2017). Privacy as an archival problem and a solution. *Archival Science*, 17(3), pp. 285-303.

Hern, A. (2016). *Muslim Professor Blocked from Game because his Name was on US Blacklist*. Available at: https://www.theguardian.com/technology/2016/jan/13/muslim-professor-blocked-paragon-game-name-us-blacklist [Accessed 13th January 2016].

Hersey, R. and Blanchard, T. (1988). *Management of Organizational Behaviour: Utilizing Human Resources*. Prentice-Hall Inc, Eaglewood, NJ.

Hildebrandt, M. (2009). Who is profiling who? Invisible visibility. In: *Reinventing Data Protection,* (Eds, Gutwirth, S., Poullet, Y., de Hert, P., de Terwangne, C. and Nouwt, S.). Springer, Dordrecht, Netherlands, pp. 239-252.

Hildebrandt, M. and Gutwirth, S. (2008). *Profiling the European Citizen: Cross-disciplinary Perspectives*. Springer Science, Berlin.

Hildebrandt, M., O'Hara, K. and Waidner, M. (2013). Introduction. In: *Digital Enlightenment Yearbook 2013: The Value Of Personal Data,* (Eds, Hildebrandt, M., O'Hara, K. and Waidner, M.). IOS Press BV, Amsterdam, pp. 1-25.

Hoffman, D.L., Novak, T.P. and Peralta, M. (1999). Building consumer trust online. *Communications of the ACM*, 42(4), pp. 80-85.

Hsieh, H.F. and Shannon, S.E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), pp. 1277-1288.

Hume, D. (2003). *A Treatise of Human Nature*. Dover Publications, New York.

Hunt, T. (2018). *';--Have I Been pwned?* Available at: https://haveibeenpwned.com [Accessed 2nd March 2018].

Hypponen, M. (2014). *Why Should You Be Worried About NSA Surveillance?* Available at: https://www.npr.org/templates/transcript/transcript.php?storyId=265386281 [Accessed 11th March 2018].

Information Commissioner's Office (2012). *Determining What Is Personal Data*. Available at: https://ico.org.uk/media/for-organisations/documents/1554/determining-what-is-personal-data.pdf [Accessed 15 March 2017].

Information Commissioner's Office (2014). *Findings from ICO Audits of 16 Local Authorities*. Available at: https://ico.org.uk/media/action-weve-taken/audits-and-advisory-visits/2081/outcomes-report-local-authorities.pdf [Accessed 6th July 2015].

Information Commissioner's Office (2018). *Information Commissioners Office: Income and Expenditure*. Available at: https://ico.org.uk/about-the-ico/our-information/income-and-expenditure/ [Accessed 18th February 2018].

Inness, J.C. (1996). *Privacy, Intimacy, and Isolation*. Oxford University Press, Oxford.

Jasserand, C. (2018). Law enforcement access to personal data originally collected by private parties: Missing data subjects' safeguards in directive 2016/680. *Computer Law & Security Review*, 34(1), pp. 154-165.

Jones, M. (1990). Super conducting super collider: Evolution of facility layout requirement and CAD system development. *Unique Underground Structures Symposium,* Available at: http://www.osti.gov/scitech/servlets/purl/6148417-PFSjrN/ [Accessed December 3 2014].

Jørgensen, R.F. (2017). What platforms mean when they talk about human rights. *Policy & Internet*, 9(3), pp. 280-296.

Kapadia, A., Henderson, T., Fielding, J. and Kotz, D. (2007). Virtual walls: Protecting digital privacy in pervasive environments. In: *5th International Conference on Pervasive Computing,* Toronto, Canada. Available at: http://www.cs.indiana.edu/~kapadia/papers/walls_pervasive07.pdf [Accessed 19th March 2018].

Kaplan, B. (2016). How should health data be used. *Cambridge Quarterly of Healthcare Ethics*, 25(2), pp. 312-329.

Kathryn R (2015). *dunnhumby: how Tesco destroyed £1.3bn of value in 9 months*. Available at: https://digit.hbs.org/submission/dunnhumby-how-tesco-destroyed-1-3bn-of-value-in-9-months/ [Accessed 14th March 2018].

Kehr, F., Kowatsch, T., Wentzel, D. and Fleisch, E. (2015). Blissfully ignorant: The effects of general privacy concerns, general institutional trust, and affect in the privacy calculus. *Information Systems Journal*, 25(6), pp. 607-635.

Kember, S. (2002). *Cyberfeminism and Artificial Life*. Routledge, London.

King, D.K. (1988). Multiple jeopardy, multiple consciousness: The context of a black feminist ideology. *Signs*, 14(1), pp. 42-72.

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 64, pp. 122-134.

Kristeva, J. (1982). *Desire in Language: A Semiotic Approach to Literature and Art*. Columbia University Press, New York, NY.

Kvale, S. (1996). *Interviews: An Introduction to Qualitative Research Interviewing*. SAGE Publications Inc, Thousand Oaks, CA.

L'Hoiry, X.D. and Norris, C. (2015). The honest data protection officer's guide to enable citizens to exercise their subject access rights: Lessons from a ten-country European study. *International Data Privacy Law*, 5(3), pp. 190-204.

Lane, C. and Bachmann, R. (1998). *Trust Within and Between Organizations: Conceptual Issues and Empirical Applications*. Oxford University Press, Oxford.

Lastowka, G. and Hunter, D. (2004). Artificial intelligence's new frontier: Artificial companions and the fourth revolution. *California Law Review*, 92(1), pp. 3-73.

Laudon, K.C. (1996). Markets and privacy. *Communications of the ACM*, 39(9), pp. 92-104.

Lee, K.M. (2006). Presence, explicated. *Communication Theory*, 14(1), pp. 27-50.

legislation.gov.uk (2018a). *Data Protection Act 1998*. Available at: https://www.legislation.gov.uk/ukpga/1998/29/pdfs/ukpga_19980029_en.pdf [Accessed 21st March 2018].

legislation.gov.uk (2018b). *Public Records Act 1958*. Available at: http://www.legislation.gov.uk/ukpga/1958/51/pdfs/ukpga_19580051_en.pdf [Accessed 21st March 2018].

Lehtiniemi, T. (2017). Personal data spaces: An intervention in surveillance capitalism? *Surveillance & Society*, 15(5), pp. 626-639.

Lessig, L. (2006). *Code: Version 2.0*. Basic Books, Cambridge, MA.

Li, H., Sarathy, R. and Xu, H. (2010). Understanding situational online information disclosure as a privacy calculus. *Journal of Computer Information Systems*, 51(1), pp. 62-71.

Lim, S., Woo, J., Lee, J. and Huh, S.-Y. (2018). Consumer valuation of personal information in the age of big data. *Journal of the Association for Information Science and Technology*, 69(1), pp. 60-71.

Lovelock, P. and Farhoomand, A.F. (2000). EIU's ViewsWire: new wine in a new bottle. In: *Proceedings of the Twenty First International Conference on Information Systems,* Brisbane. Available at: http://aisel.aisnet.org/icis2000/90/ [Accessed 19th March 2018].

Lucas, L. (2018). *China's Weibo Eclipses Rival Twitter's Market Capitalisation*. Available at: https://www.ft.com/content/c46c55b4-f1b8-11e6-8758-6876151821a6 [Accessed 1st March 2018].

Ludington, S. (2006). Reining in the data traders: A tort for the misuse of personal information. *Maryland Law Review*, 66(1), pp. 140-193.

Ludwig, L.F., Lauwers, J.C., Lantz, K.A. and Burnett, G.J. (1997). *Multimedia collaboration system with separate data network and a/v network controlled by information transmitting on the data network*. Vicor, Inc., 5617539, Palo Alto, CA.

Luhmann, N., Poggi, G. and Burns, T. (1979). *Trust and Power*. John Wiley & Sons, Chichester.

Luo, S., Morone, F., Sarraute, C., Travizano, M. and Makse, H.A. (2017). Inferring personal economic status from social network location. *Nature Communications*, 8:15227, pp. 1-7.

Lupton, D. and Michael, M. (2017). 'Depends on who's got the data': Public understandings of personal digital dataveillance. *Surveillance & Society*, 15(2), pp. 254-268.

Lyon, D. (1994). *The Electronic Eye: The Rise of Surveillance Society*. University of Minnesota Press, Minneapolis, MN.

Lyon, D. (2000). *Jesus in Disneyland: Religion in Postmodern Times*. Polity Press, Cambridge.

Lyon, D. (2009). *Identifying Citizens: ID Cards as Surveillance*. Polity Press, Cambridge.

Maciejewski, M. (2017). To do more, better, faster and more cheaply: Using big data in public administration. *International Review of Administrative Sciences*, 83(1_suppl), pp. 120-135.

MacKinnon, C.A. (1991). *Toward a Feminist Theory of the State*. Harvard University Press, Cambridge, MA.

Madaan, N., Ahad, M.A. and Sastry, S.M. (2018). Data integration in IoT ecosystem: Information linkage as a privacy threat. *Computer Law & Security Review*, 34(1), pp. 125-133.

Madden, M., Fox, S., Smith, A. and Vitak, J. (2007). *Digital Footprints: Online Identity Management and Search in the Age of Transparency*. Pew/Internet & American Life Project, Washington, DC.

Madden, M. and Rainie, L. (2015). *Americans' Attitudes About Privacy, Security and Surveillance*. Available at: http://www.pewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/ [Accessed 16th March 2018].

Maltseva, K. and Lutz, C. (2018). A quantum of self: A study of self-quantification and self-disclosure. *Computers in Human Behavior*, 81, pp. 102-114.

Mansell, R. and Collins, B.S. (2007). *Trust and Crime in Information Societies*. Edward Elgar Publishing Ltd, Cheltenham.

Mansell, R. and Steinmueller, W.E. (2002). *Mobilizing the Information Society: Strategies for Growth and Opportunity*. Oxford University Press, Oxford.

Manthorpe, R. (2018). *Sam Amrami Tracks You in Pret. And at Starbucks. And Down The Pub*. Available at: http://www.wired.co.uk/article/tamoco-sam-amrami-proximity-tracking-mwc [Accessed 24th February 2018].

Marcuse, H. (1968). *One Dimensional Man*. Sphere Books, London.

Marr, B. (2010). *The Intelligent Company: Five Steps to Success with Evidence-based Management*. John Wiley & Sons, Chichester.

Marsden, P. (2010). Social commerce: Monetizing social media. *Syzygy UK Ltd.,* Available at: https://digitalintelligencetoday.com/downloads/White_Paper_Social_Commerce_EN.pdf [Accessed 21st March 2018].

Mayer-Schönberger, V. (2009). *Delete: The Virtue of Forgetting in the Digital Age*. Princeton University Press, Princeton, NJ.

Mayer-Schönberger, V. and Cukier, K. (2013). *Big Data: A Revolution that Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt, Boston MA.

McDonagh, E.L. (1997). *Breaking the Abortion Deadlock: From Choice to Consent*. Oxford University Press, USA, New York, NY.

McInnerney, J.M. and Roberts, T.S. (2004). Online learning: Social interaction and the creation of a sense of community. *Educational Technology & Society*, 7(3), pp. 73-81.

McKechnie, L.E.F. (2008). Reactivity. In: *The SAGE Encyclopedia of Qualitative Research Methods,* (Eds, Given, L.M. and Saumure, K.). SAGE Publications Inc, Thousand Oaks, CA, pp. 729.

McWilliams, A. and Siegel, D. (2001). Corporate social responsibility: A theory of the firm perspective. *The Academy of Management Review*, 26(1), pp. 117-127.

Mead, G.H. (1934). *Mind, Self and Society*. University of Chicago Press, Chicago, IL.

Merlo, O., Eisingerich, A., Auh, S. and Levstek, J. (2018). The benefits and implementation of performance transparency: The why and how of letting your customers 'see through' your business. *Business Horizons*, 61(1), pp. 73-84.

Merritt, A.C., Effron, D.A. and Monin, B. (2010). Moral self-licensing: When being good frees us to be bad: Moral self-licensing. *Social and Personality Psychology Compass*, 4(5), pp. 344-357.

Metzinger, T. (2010). *The Ego Tunnel: The Science of the Mind and the Myth of the Self*. Basic Books, New York, NY.

Millard, C. and Hon, W.K. (2012). Defining 'personal data' in e-social science. *Information, Communication & Society*, 15(1), pp. 66-84.

Miller, J. (2014). *Microsoft Pays $2.5bn for Minecraft Maker Mojang*. Available at: http://www.bbc.co.uk/news/technology-29204518 [Accessed 2nd March 2018].

Mingers, J. (2004). Real-izing information systems: Critical realism as an underpinning philosophy for information systems. *Information and Organization*, 14(2), pp. 87-103.

Misztal, B.A. (1996). *Trust in Modern Societies: The Search for the Bases of Social Order*. Polity Press, Cambridge.

Molloy, M. (2016). *Lucky Man Becomes $123 Million 'Richer' After Banking Error*. Available at: https://www.telegraph.co.uk/news/2016/05/26/lucky-man-becomes-a123-million-richer-after-banking-error/ [Accessed 24th August 2018].

Moore, K. and McElroy, J.C. (2012). The influence of personality on Facebook usage, wall postings, and regret. *Computers in Human Behavior*, 26, pp. 267-274.

Moustakas, C. (1994). *Phenomenological Research Methods*. SAGE Publications Ltd., London.

Murray, J.A.H. (1971). *The Compact Edition of the Oxford English Dictionary.* Oxford University Press, Oxford.

Neville, A. (2017). Is it a human right to be forgotten: Conceptualizing the world view. *Santa Clara Journal of International Law*, 15, pp. 157-172.

Ngo, F. and Jaishankar, K. (2017). Commemorating a decade in existence of the International Journal of Cyber Criminology: A research agenda to advance the scholarship on cyber crime. *International Journal of Cyber Criminology*, 11(1), pp. 1-9.

Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford Law Books, Stanford, CA.

Nissenbaum, H. (2016). Must Privacy Give Way to Use Regulation. *Lecture at the Watson Institute, Brown University, March 15th 2016,* Available at: https://www.sas.upenn.edu/andrea-mitchell-center/sites/www.sas.upenn.edu.dcc/files/Nissenbaum-UPenn-Democracy.pdf [Accessed 26th February 2016].

Nissenbaum, H. (2017). *Deregulating Collection: Must Privacy Give Way to Use Regulation?* Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3092282 [Accessed 24th February 2018].

O'Hara, K. (2010). Intimacy 2.0: Privacy rights and privacy responsibilities on the World Wide Web. In: *Web Science Conference,* Raleigh, NC, USA. Available at: http://journal.webscience.org/294/2/websci10_submission_3.pdf [Accessed 19th March 2018].

O'Hara, K. and Stevens, D. (2006). *inequality.com: Power, Poverty and the Digital Divide*. Oneworld, Oxford.

O'Keeffe, G.S. and Clarke-Pearson (2011). The impact of social media on children, adolescents, and families. *Pediatrics*, 127(4), pp. 800-804.

Offe, C. (2009). Governance: An "empty signifier". *Constellations*, 16(4), pp. 550-562.

Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, 57, pp. 1701.

Okazaki, Y. (2017). Implications of big data for customs - how it can support risk management capabilities. *World Customs Organisation,* Available at: http://www.wcoomd.org/~/media/wco/public/global/pdf/topics/research/research-paper-series/39_okazaki_big-data.pdf [Accessed 1st March 2018].

Olson, E.T. (2007). *What Are We?: A Study in Personal Ontology*. Oxford University Press, Oxford.

Overton, D. (2016). *Interim report 2- Flagship Next Generation Internet (19th December 2016)*. Available at: https://ec.europa.eu/futurium/sites/futurium/files/20161219_ec_20161215_interim_report_2_v5.0.pdf [Accessed 26th February 2018].

Palfrey, J. and Gasser, U. (2008). *Born Digital: Understanding the First Generation of Digital Natives*. Basic Books, New York, NY.

Parent, W.A. (1983). A new definition of privacy for the law. *Law and Philosophy*, 2(3), pp. 305-338.

Parkinson, B., Millard, D.E., O'Hara, K. and Giordano, R. (2017). The Digitally Extended Self: A lexicological analysis of personal data. *Journal of Information Science*, pp. 016555151770623.

Pascalev, M. (2017). Privacy exchanges: restoring consent in privacy self-management. *Ethics and Information Technology*, 19(1), pp. 39-48.

Pathak, B., Garfinkel, R., Gopal, R.D., Venkatesan, R. and Yin, F. (2010). Empirical analysis of the impact of recommender systems on sales. *Journal of Management Information Systems*, 27(2), pp. 159-188.

Pearson, E. (2009). All the World Wide Web's a stage: The performance of identity in online social networks. *First Monday,* 14(3). Available at: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2162/2127 [Accessed 20th March 2018].

Peppard, J. and Ward, J. (2004). Beyond strategic information systems: Towards an IS capability. *The Journal of Strategic Information Systems*, 13(2), pp. 167-194.

Pieters, R. and Zeelenberg, M. (2005). On bad decisions and deciding badly: When intention–behavior inconsistency is regrettable. *Organizational Behavior and Human Decision Processes*, 97(1), pp. 18-30.

Pieters, W. (2017). Beyond individual-centric privacy: Information technology in social systems. *The Information Society*, 33(5), pp. 271-281.

Pinnegar, S. and Daynes, J.G., (2006). Locating narrative inquiry historically: Thematics in the turn to narrative. In: *Handbook of Narrative Inquiry,* (Ed, Clandinin, D.J.). SAGE Publications Ltd., London, pp. 3-34.

Ploeg, I.V.D. (2003). Biometrics and the body as information. In: *Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination,* (Ed, Lyon, D.). Routledge, London, pp. 57-73.

Pollach, I. (2007). What's wrong with online privacy policies. *Communications of the ACM*, 50(9), pp. 103-108.

Polonetsky, J., Tene, O. and Finch, K. (2016). Shades of gray: Seeing the full spectrum of practical data de-identification. *Santa Clara Law Review*, 56, pp. 593-629.

Posner, R.A. (1978). The right of privacy. *The Georgia Law Review*, 12(3), pp. 393-422.

Post, R.C. (1963). Three concepts of privacy. *Georgetown Law Journal*, 89(6), pp. 2087-2098.

Powell, R.A. and Single, H.M. (1996). Methodology matters - V. *International Journal for Quality in Health Care*, 8(5), pp. 499-504.

President's Council of Advisors on Science and Technology (2014). *Big Data and Privacy: A Technological Perspective*. Executive Office of the President, White House, Washington DC.

Pullinger, J. (1997). The creation of the Office for National Statistics. *International Statistical Review*, 65(3), pp. 291-308.

Rachels, J. (1975). Why privacy is important. *Philosophy and Public Affairs*, 4(4), pp. 323-333.

Rauhofer, J. (2014). 'Look to yourselves, that we lose not those things which we have wrought.' What do the proposed changes to the purpose limitation principle mean for public bodies' rights to access third-party data. *International Review of Law, Computers & Technology*, 28(2), pp. 144-158.

Raz, J. (1986). T*he Morality of Freedom*. Clarendon Press, Oxford.

Redman, T.C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM*, 41(2), pp. 79-82.

Regan, P.M. (2004). Old issues, new context: Privacy, information collection, and Homeland Security. *Government Information Quarterly*, 21(4), pp. 481-497.

Reinsel, D. (2007). *The Expanding Digital Universe*. Available at: http://www.emc.com/collateral/demos/microsites/idc-digital-universe/iview.htm [Accessed 15th March 2018].

Reinsel, D., Gantz, J. and Rydning, J. (2017). *Data Age 2025: The Evolution of Data to Life-Critical Don't Focus on Big Data; Focus on Data That's Big*. Available at: https://www.seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf [Accessed 26th February 2018].

Richardson, L. (2007). *What Terrorists Want: Understanding the Enemy, Containing the Threat*. Random House Publishing Group, New York, NY.

Robert, E.S. (1995). *The Art of Case Study Research*. SAGE Publications Inc, Thousand Oaks, CA.

Robertson, A. et al. (2010). Implementation and adoption of Nationwide Electronic Health Records in secondary care in England: Qualitative analysis of interim results from a prospective national evaluation. *British Medical Journal*, 341 :c4564,

Robinson, P. (2004). Anti-money laundering regulation - next generation. In: *City & Financial Conference,* Available at: http://www.fsa.gov.uk/Pages/Library/Communication/Speeches/2004/SP174.shtml [Accessed 3rd March 2018].

Robinson, S.C. (2017a). Disclosure of personal data in ecommerce: A cross-national comparison of Estonia and the United States. *Telematics and Informatics*, 34(2), pp. 569-582.

Robinson, S.C. (2017b). What's your anonymity worth? Establishing a marketplace for the valuation and control of individuals' anonymity and personal data. *Digital Policy, Regulation and Governance*, 19(5), pp. 353-366.

Romanou, A. (2018). The necessity of the implementation of privacy by design in sectors where data protection concerns arise. *Computer Law & Security Review*, 34(1), pp. 99-110.

Rosenberg, J.M. (1969). *The Death of Privacy*. Random House, New York, NY.

Rössler, B. (2005). T*he Value of Privacy*. Polity Press, Cambridge.

Rouvroy, A. and Poullet, Y. (2009). The right to informational self-determination and the value of self-development: Reassessing the importance of privacy for democracy. In: *Reinventing Data Protection,* (Eds, Gutwirth, S., Poullet, Y., de Hert, P., de Terwangne, C. and Nouwt, S.). Springer, Dordrecht, Netherlands, pp. 45-76.

Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), pp. 163-180.

Safran, C., Bloomrosen, M., Hammond, W.E., Labkoff, S., Markel-Fox, S., Tang, P.C., Detmer, D.E. and Expert, P. (2007). Toward a national framework for the secondary use of health data: An American Medical Informatics Association White Paper. *Journal of the American Medical Informatics Association*, 14(1), pp. 1-9.

Sanchez, A. (2009). Facebook feeding frenzy: Resistance-through-distance and resistance-through-persistence in the societied network. *Surveillance & Society*, 6(3), pp. 275-293.

Schlosser, A. (2018). *You may have heard data is the new oil. It's not*. Available at: https://www.weforum.org/agenda/2018/01/data-is-not-the-new-oil/ [Accessed 8th March 2018].

Schmidt, E. (2010). *9 - Atmosphere: Fireside chat with Eric Schmidt*. Available at: http://www.youtube.com/eventsatgoogle#p/u/5/qBaVyCcw47M [Accessed 24th February 2018].

Schneier, B. (2008). The myth of the "transparent society". *Wired News,* 6. Available at: https://www.wired.com/2008/03/securitymatters-0306/ [Accessed 20th March 2018].

Schoeman, F.D. (1984). Privacy: Philosophical dimensions of the literature. In: *Philosophical Dimensions of Privacy: An Anthology,* (Ed, Schoeman, F.D.). Cambridge University Press, Cambridge,

Schwartau, W. (1994). *Information Warfare: Chaos on the Electronic Superhighway*. Thunder's Mouth Press, New York, NY.

Scott, J. (1990). *A Matter Of Record*. Polity Press, Cambridge.

Sellen, A., Rogers, Y., Harper, R. and Rodden, T. (2009). Reflecting human values in the digital age. *Communications of the ACM*, 52(3), pp. 58-66.

Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., Hall, W. and Schraefel, M.C. (2012). Linked open government data: Lessons from data.gov.uk. *IEEE Intelligent Systems*, 27(3), pp. 16-24.

Shields, R. (2003). *The Virtual*. Routledge, London.

Siemens, G. and Long, P. (2011). Penetrating the fog: Analytics In learning and education. *Educause Review*, 46(5), pp. 31-40.

Siibak, A. (2009). Constructing the self through the photo selection - visual impression management on social networking websites. *Journal of Psychosocial Research on Cyberspace,* 3(1). Available at: http://www.cyberpsychology.eu/view.php?cisloclanku=2009061501&article=1 [Accessed December 1 2014].

Simon, P. and Graham, S. (2017). Potential privacy ramifications of modern vehicle software and firmware. In: *16th European Conference on Cyber Warfare and Security, Dublin,* Academic Conferences and Publishing International Limited, Reading, pp. 452-457.

Sirdeshmukh, D., Singh, J. and Sabol, B. (2002). Consumer trust, value, and loyalty in relational exchanges. *The Journal of Marketing*, 66(1), pp. 15-37.

Slaughter, A.-M. (2017). *3 Responsibilities Every Government has Towards its Citizens*. Available at: https://www.weforum.org/agenda/authors/anne-marieslaughter [Accessed 2nd March 2018].

Smithson, M. (1985). Toward a social theory of ignorance. *Journal for the Theory of Social Behaviour*, 15(2), pp. 151-172.

Solove, D.J. (2001). Privacy and power: Computer databases and metaphors for information privacy. *Stanford Law Review*, 53(6), pp. 1393-1462.

Solove, D.J. (2004). *The Digital Person: Technology and Privacy in the Information Age*. New York University Press, New York, NY.

Solove, D.J. (2006). A taxonomy of privacy. *University of Pennsylvania Law Review*, 154(3), pp. 477-564.

Solove, D.J. (2008). *Understanding Privacy*. Harvard University Press, London.

Solove, D.J. (2013). Introduction: Privacy self-management and the consent dilemma. *Harvard Law Review*, 126(7), pp. 1880-1903.

Spiekermann, S., Acquisti, A., Böhme, R. and Hui, K.-L. (2015). The challenges of personal data markets and privacy. *Electronic Markets*, 25(2), pp. 161-167.

Sprenger, P. (1999). Sun on privacy: 'Get over it'. *Wired,* Available at: http://www.wired.com/politics/law/news/1999/01/17538 [Accessed 22nd September 20127].

Statista (2018a). *Most Concerning Issues About Online Usage According to Internet Users In the United States as of May 2017*. Available at: https://www.statista.com/statistics/248488/frequency-with-which-us-internet-users-worry-about-online-privacy/ [Accessed 4th March 2018].

Statista (2018b). *Price Paid Per User of the Acquired Company in Selected Tech Acquisitions (in U.S. Dollars)*. Available at: https://www.statista.com/statistics/222363/price-per-user-at-selected-tech-acquisitions/ [Accessed 2nd March 2018].

Steiger, S. (2017). The unshaken role of GCHQ. In: *Privacy, Data Protection and Cybersecurity in Europe,* (Eds, J. Schünemann, W. and Baumann, M.-O.). Springer, Cham, Switzerland, pp. 79-95.

Stockport, G.J. (2010). Google: Organising the worlds' information. *International Journal of Technology Marketing*, 5(1), pp. 27-43.

Suter, L. (2017). Taxman unleashes its 'snooper computer': what information does its have on you? *The Telegraph,* Available at: taxman-unleashes-snooper-computer-information-does-have [Accessed 20th March 2018].

Svendsen, L. (2008). *A Philosophy of Fear*. Reaktion Books Ltd, London.

Svenonius, E. (2009). *The Intellectual Foundation of Information Organization*. MIT Press, Cambridge, MA.

Swan, M. (2015). Connected car: Quantified self becomes quantified car. *Journal of Sensor and Actuator Networks*, 4, pp. 2-29.

Tapp, A. (2008). *Principles of Direct and Database Marketing*. Pearson Education Limited, Harlow.

Taylor, C. (1992). *Sources of the Self: The Making of the Modern Identity*. Cambridge University Press, Cambridge.

The Economist (2017). *The world's most valuable resource is no longer oil, but data.* Available at: https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource [Accessed 8th March 2018].

Thomas, J. (1993). *Doing Critical Ethnography*. SAGE Publications Ltd., London.

Tolbert, P. and Hall, R. (2009). *Organizations: Structures, Processes and Outcomes*. Routledge, Abingdon.

Tomlinson, H. and Evans, R. (2005). *Tesco Stocks Up on Inside Knowledge of Shoppers' Lives*. Available at: https://www.theguardian.com/business/2005/sep/20/freedomofinformation.supermarkets [Accessed 16th March 2018].

Torpey, J.C. (2000). *The Invention of the Passport: Surveillance, Citizenship, and the State*. Cambridge University Press, Cambridge.

Turkle, S. (1994). Constructions and reconstructions of self in virtual reality: Playing in the MUDs. *Mind, Culture, and Activity*, 1(3), pp. 158-167.

Turkle, S. (1999). Cyberspace and identity. *Contemporary Sociology*, 28(6), pp. 643-648.

Turner, J.R. (1998). *The Handbook of Project-based Management*. McGraw-Hill, Maidenhead.

United Nations (1949). *United Nations Universal Declaration of Human Rights 1948*. Available at: http://www.jus.uio.no/lm/un.universal.declaration.of.human.rights.1948/portrait.a4.pdf [Accessed 18 March 2018].

Vander Valk, F. (2008). Identity, power, and representation in virtual environments. *MERLOT Journal of Online Learning and Teaching*, 4(2), pp. 205-211.

Vedder, A. (2000). Medical data, new information technologies and the need for normative principles other than privacy rules. In: *Law and Medicine: Current Legal Issues Volume 3,* (Eds, Freeman, M. and Lewis, A.). Oxford University Press, Oxford, pp. 441-459.

Vervier, L., Zeissig, E.-M., Lidynia, C. and Ziefle, M. (2017). Perceptions of digital footprints and the value of privacy. In: *Proceedings of the 2nd International Conference on Internet of Things, Big Data and Security,* Porto, Portugal. Available at: http://dx.doi.org/10.5220/0006301000800091 [Accessed 19th March 2018].

Vogelstein, F. (2009). Great wall of Facebook: The social network's plan to dominate the internet - and keep Google out. *Wired Magazine*, 17, pp. 96-121.

von Solmes, S. and van Heerden, R. (2015). The consequences of Edward Snowden NSA related information disclosures. In: *10th International Conference on Cyber Warfare and Security, Sukuza, South Africa,* Academic Conferences and Publishing International Limited, Reading, pp. 358-368.

Warren, S.D. and Brandeis, L.D. (1890). The right to privacy. *Harvard Law Review*, IV(5), pp. 193-220.

Weber, R.P. (1990). *Basic Content Analysis*. Sage Publications Inc, Newbury Park, CA.

Weber, R.H. (2011). The right to be forgotten: More than a pandora's box? *Journal of Intellectual Property, Information Technology and E-Commerce*, 2(2), pp. 120-130.

Westin, A.F. (1967). *Privacy and Freedom*. Atheneum, New York, NY.

Wiese, J., Das, S., Hong, J.I. and Zimmerman, J. (2017). Evolving the ecosystem of personal behavioral data. *Human–Computer Interaction*, 32(5-6), pp. 447-510.

Wilbanks, J. (2014). Portable approaches to informed consent and open data. In: *Privacy, Big Data, and the Public Good: Frameworks for Engagement,* (Eds, Lane, J., Stodden, V., Bender, S. and Nissenbaum, H.). Cambridge University Press., Cambridge, pp. 234-252.

Will, M.A., Garae, J., Tan, Y.S., Scoon, C. and Ko, R.K.L. (2017). Returning control of data to users with a personal information crunch - a position paper. In: *2017 International Conference on Cloud Computing Research and Innovation,* Available at: http://dx.doi.org/10.1109/icccri.2017.12 [Accessed 19th March 2018].

Wittes, B. (2011). Databuse: Digital privacy and the mosaic. *Brookings Institute,* Available at: https://www.technologylawdispatch.com/wp-content/uploads/sites/26/2011/05/GRE-Blog-May-17-2011-3.pdf [Accessed July 2 2013].

Wolcott, H., F. (2008). *Ethnography: A Way of Seeing*. AltaMira Press, Plymouth.

Woodhouse, A. (2018). *Tencent's market cap surpasses Facebook*. Available at: https://www.ft.com/content/01a7308c-b947-35ae-ade3-3abfbc200560 [Accessed 1st March 2018].

Wottrich, V.M., van Reijmersdal, E.A. and Smit, E.G. (2018). The privacy trade-off for mobile app downloads: The roles of app value, intrusiveness, and privacy concerns. *Decision Support Systems*, 106, pp. 44-52.

Yin, R.K. (2003). *Case Study Research Design and Methods*. SAGE Publications Inc, Thousand Oaks, CA.

Yin, R.K. (2014). *Case Study Research Design and Methods*. SAGE Publications Inc, Los Angeles, CA.

Ying-chun, C.A.O. (2009). Discussion on the influence of internet to undergraduates' self-identity. *Economic Research Guide*, 20, pp. 104-111.

Zakhary, S. and Benslimane, A. (2018). On location-privacy in opportunistic mobile networks, a survey. *Journal of Network and Computer Applications*, 103, pp. 157-170.

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), pp. 75-89.