Project acronym:	BYTE
Project title:	Big data roadmap and cross-disciplinarY community for addressing socieTal Externalities
Grant number:	619551
Programme:	Seventh Framework Programme for ICT
Objective:	ICT-2013.4.2 Scalable data analytics
Contract type:	Co-ordination and Support Action
Start date of project:	01 March 2014
Duration:	36 months
Website:	www.byte-project.eu

Deliverable D2.2:
Report on public perceptions and social impacts relevant to big data

Author(s):	Anna Donovan, Rachel Finn and Kush Wadhwa (Trilateral Research & Consulting)
Reviewers	Peter Stefan, National Information Infrastructure Development Institute

Dissemination level:	Public
Deliverable type:	Final
Version:	1.1
Submission date:	20 July 2015

Table of Contents

EXECUTIVE SUMMARY

This report primarily aims to provide an indication of the general public perceptions of, and aspirations for, information practices relating to big data, such as data collection, data storage, data sharing (including selling) and data analysis.[1] The collection, storage and usage of personal data have become a part of everyday life at all levels of society and as a result, users have raised concerns in relation to these processes, especially with resect to the privacy and security of personal data. Public sentiments play a role in determining which externalities are desirable and which may generate public resistance. These sentiments are important because they can indicate the extent to which the public, as a major source of data, willingly contribute to the big data process currently, and how willing they may continue to be. This work was undertaken as part of the EU FP7-funded project "Big data roadmap and cross disciplinary community for addressing societal externalities" (BYTE), within work Package 2 (WP2). This report is the product of desk research that was largely reliant on public opinion surveys, as well as academic journal articles, project reports, media materials, materials from industry and any other information relevant to public opinions of, and aspirations towards, information processes associated with the collection of data from the public.

In chapter 2, we examine public sentiments towards big data. To that end, we examine public perceptions as they relate to specific big data practices, as well as the major concerns identified in a number of public opinion surveys. These big data practices are data collection, data storage, data analysis, and data sharing (including selling of data). Here, we reveal that the main areas of concern relate to the privacy and security of data collected as well as a general distrust of those handling their data, particularly companies operating in the private sector. There does not appear to be any overwhelmingly negative sentiments about specific practices. This may be because users tend to be more informed about the collection of their personal data as they play an active role in that process by providing the information, whereas data analysis, data storage and the selling and sharing of data appear more opaque to the user because the data subject is removed from those practices. However, companies and organisations can better inform their users with more transparent policies concerning the subsequent use of the data and specifically, the benefits that can flow from information technology practices. This is particularly relevant as research suggests that the concerns expressed by members of the public can be reduced when they are aware of societal benefits, rather than resigning to the provision of their data in exchange for a service, even when they hold concerns as to the privacy and security of their data.

In chapter 3, we examine public aspirations towards big data by looking at what relevant information can tell us about how members of the public would like big data to operate in a manner that causes the least number of negative implications for them. This examination of two major public aspirations for big data, namely the identification of benefits that flow from the provision of data, and transparency of why, how, when and where data will be used following its collection, provides useful information to big data companies and organisations in the public and private sectors. The use of data by public sector organisations is more favourable to the public because of the aspirations they hold in relation to the benefits achieved through the collection and use of data by such organisations. This is particularly true

[1] We attempt to address these processes separately, although the research often considers an overlap of these processes under the umbrella terms "process", "handle" or "use". There is very little information concerning public sentiments towards the re-use of data. However, at this stage, it is plausible that the sentiment would reflect that in relation to the sharing and/ or sale of data.

when a tangible public benefit is readily identified, such as when data use produces improvements in public security or where developments in health care treatment and diagnostics are achieved. Ultimately, public aspirations for big data revolve around the collection and use of data, especially personal data, to be used by government and companies for their benefit, and in a transparent manner. Thus, members of the public are more likely to willingly disclose a greater amount of their data, if big data actors seeking to use that data to meet public or commercial objectives incorporate public aspirations into big data policies and practices.

In chapter 4, we use the information relating to public perceptions of, and aspirations towards, information practices relating to big data, to begin to develop a public perceptions good practice framework. A public perceptions framework that takes into account public perceptions and aspirations can contribute to the development and growth of the big data industry by ensuring that citizens, as a major data source, continue to comfortably and securely contribute to large data sets. To that end, it is particularly important that the following issues are examples of what can be incorporated into a relevant good practice framework:

- Consider how to address public perceptions and aspirations towards data protection and privacy;
- Consider how to address public perceptions and aspirations towards data security;
- Implement more transparent practices; and
- Implement adequate security measures.

Overall, this reports focuses on positive public sentiments and aspirations towards big data because recognising these issues is imperative to the continuation of data processing activities and the future of big data as a value adding process. Personal benefit is the strongest incentive for being in favour of the collection and use of personal data by government and companies. Conversely, if the public see little benefit from sharing their data and little confidence that they will see benefits in future, this may hinder the amounts of data available to big data actors into the future thereby, threatening the longevity of the European big data industry. Public sentiment towards issues that relate to big data is crucial to the wider examination of societal externalities of big data that the BYTE project aims to examine.

1 INTRODUCTION

1.1 OVERVIEW

Technology and big data are poised to alter many sectors of society. The collection of data from the public, and the subsequent handling of that data by public and private sector organisations enable such organisations to capture the benefits of big data. Generally, "the big data phenomenon affects many facets of contemporary life and has the potential to alter governance, the economy, and the very structures of society."[2] In many contexts, ordinary citizens are at the core of data collection and handling within these sectors, as they are the source of vast amounts of data. As such the relationship between citizens and big data requires investigation. Central to that relationship is public perceptions of, and aspirations towards, data collection, and subsequent information practices such as data storage, data sharing and selling and data analyses. These perceptions and aspirations are important because they indicate the extent to which the public, as a significant source of data, are willing to provide information to be utilised in these processes, and ultimately, contribute to the longevity of the European big data industry as a vehicle for economic growth and development in Europe.

This report primarily aims to provide an indication of the general public perceptions of, and aspirations for, information practices relating to big data, such as data collection, data storage, data sharing (including selling) and data analysis.[3] Relevantly, a 2014 Ipsos Mori survey confirms that public attitudes towards big data are not yet widely known, although this is an emerging area in science and policy.[4] The collection, storage and usage of personal data have become a part of everyday life at all levels of society and as a result, members of the public have raised concerns in relation to these processes, especially with resect to the privacy and security of personal data. According to the Eurobarometer 359 survey[5], Europeans commonly use the following types of credentials which are a source of data for businesses and governments alike: credit cards and bank cards (74%), national identity cards or residence permits (68%), government entitlement cards (65%), or driving licences (63%). In addition, 34% of respondents to that survey have an account they use on the Internet, such as email, or for social networking or commercial services. These credentials provide a wealth of data that is later combined to create large data sets that are stored, analysed, shared or sold, and/ or re-used. Data mining tools have been developed to find patterns in large collections of personal data collected from such sources, to identify individuals and to attempt to predict their interests and preferences through tracking and profiling techniques. Companies use these technologies to obtain large customer bases, and governments are increasingly analysing and exchanging information about their citizens. Should organisations collecting and handling

[2] Data and Society Research Institute, *Event Summary: The Social, Cultural and Ethical Dimensions of "Big Data",* New York University, 17 March 2014. http://www.datasociety.net/pubs/2014-0317/BigDataConferenceSummary.pdf

[3] We attempt to address these processes separately, although the research often considers an overlap of these processes under the umbrella terms "process", "handle" or "use". There is very little information concerning public sentiments towards the re-use of data. However, at this stage, it is plausible that the sentiment would reflect that in relation to the sharing and/ or sale of data.

[4] Cameron, Daniel, Sarah Pope and Michael Clemence, *Dialogue on Data: Exploring the Public's Views on Using Administrative Data for Research Purposes,* Ipsos MORI, UK, 2014, p.9 ("2014a"). http://www.ipsos-mori.com/DownloadPublication/1652_sri-dialogue-on-data-2014.pdf

[5] TNS Opinion and Social, *Special Eurobarometer 359: Attitudes on Data Protection and Electronic Identity in the European Union*, Special Eurobarometer, European Commission, 2011, p. 2. http://ec.europa.eu/public_opinion/archives/eb_special_359_340_en.htm

data wish to continue these big data practices to drive commercial development into the future, they may need to consider the role public perceptions and aspirations will play in enabling them to do so, and ultimately, to continue to enable them to capture the benefits of big data. If consideration is not given to public perceptions and aspirations relating to big data information practices, negative user sentiment may build, resulting in a general public reluctance to provide information (especially personal data) and/ or refuse to participate in the objectives of data-focussed businesses and organisations such as those in the public health sector. Thus, understanding perceptions, opinions and implementing user aspirations can assist in diminishing the negative externalities of big data such as data protection and security breaches, which in turn, builds user trust.

However, public perceptions and opinions are not fixed and will regularly alter depending upon the information available to the opinion holder at the time perceptions are formed:

> Public opinions can differ significantly depending upon how a question is asked and a topic is framed. To date, research suggests that the public are generally opposed to any form of data use and collection by government and companies, although in practice the public consider there to be no alternative, especially in relation to the sharing of their personal data with government and companies. The public also appears to expect this practice to increase in future.[6]

Despite an increase in public concern of big data practices, Singleton notes:

> people will express concerns if questioned about 'concerns', but will readily trade these 'concerns' for health or other benefits, even altruistic ones. 'Real world' choices can be very different (and constrained) from those offered in opinion surveys where costs and trade-offs may not appear.[7]

Thus, it is important to consider that the snapshot of public opinions provided in this report were provided within a context that was specific to the respondents' circumstances and concerns at the time the surveys were carried out.

Nevertheless, this examination of some of the views provided by respondents to public opinion surveys provide a general indication of recent levels of social acceptability of, and aspirations towards, particular big data technologies, practices and applications and can assist in providing a framework that begins to consider the social impacts of big data in later work.

1.2 METHODOLOGY

This report is the product of desk research that was reliant largely on public opinion surveys, as well as academic journal articles, project reports, media materials, materials from industry and any other relevant information relevant to public opinions of, and aspirations towards information processes associated with the collection of data from the public and which provide indicators of public opinions toward big data. The main sources relied upon in the writing of this report are set out in the table below:

[6] Sciencewise Expert Resource Centre ("Sciencewise"), *Big Data: Public views on the Collection, Sharing and Use of Personal Data by Government and Companies,* UK, 1 April 2014, p.1. http://www.sciencewise-erc.org.uk/cms/assets/Uploads/SocialIntelligenceBigData.pdf

[7] Cited in Sciencewise, op. cit., 2014, p.11.

#	Title	Organisation/ Date	Country /region	Survey/ source
1	Deliverable 7.1: Report on existing Surveys	PRISMS project, 14 March 2013	EU	Source
2	Unisys Security Index 2H10 Europe	Unisys, 22 October 2010	Europe	Source (Youtube clip)
3	Sustaining Public Trust in Health Data	Kingsley Manning, The National Health and IT Conference and Exhibition, 20 March 2014	UK	Source (Speech)
4	The Public Must be at the Heart of Any New Settlement on Data Sharing: the Data Dialogue	DEMOS	UK	Source (paper)
5	Event Summary: The Social, Cultural and Ethical Dimensions of Big Data	Data and Society Research Institute. 17 March 2014	US	Source
6	Reputation Management and Social Media: How People Monitor their Identity and Search for others Online	Pew Research Centre, Pew Internet Project, 26 May 2010	US	Source
7	What You really Agreed to in Facebook Terms and Conditions	News.com.au, Australia, 22 July 2014	Global	Source (news article)
8	Dialogue on Data: Exploring the Public's Views on Using Administrative Data for Research Purposes	Ipsos MORI, 2014	UK	Survey
9	Dialogue on Data: Exploring the Public's Views on the Changes to the Census	Ipsos MORI, 2014	UK	Survey
10	The Use of Personal Health Information in Medical Research General Public Consultation	Ipsos MORI/ Medical Research Council, June 2006	UK	Survey
11	Unisys Security Index: UK	Lieberman Research Group, 14 May 2014	UK	Survey
12	Data Nation 2013: Balancing Growth and Responsibility	Deloitte, 2013	UK/EU	Survey
13	Unisys security Index: UK Customers Switch Banks and Retailers Over Privacy Fears but Admit to Shortcomings When Protecting Themselves	Unisys, 27 October 2012	UK	Survey

14	Big data: Public Views on the Collection, Sharing and Use of Personal Data by Government and Companies	Sciencewise Expert Resource Centre, 1 April 2014	UK/ EU	Survey (review of 17 other related surveys**)**
15	Eurobarometer Flash 225 Survey	TNS Opinion & Social/ The European Commission, 2008	EU	Survey
16	Privacy 2.0: Personal and Consumer Protection in the New Media Reality	SINTEF, The Norwegian Research Council, 2 November 2009	Norway	Survey (and source – accompanying report)
17	Search Engine Use	Pew Research Centre, March 2012	US	Survey
18	Knowing More About Privacy makes Users Share Less with Facebook and Google	Siegel + Gale, March 2012	US	Survey
19	Global Opposition to U.S. Surveillance and Drones, but Limited Harm to America's Image	Pew Research, Global Attitudes Project, 14 July 2014	US	Survey (and source – accompanying report
20	Privacy: A Study of Attitudes and Behaviours in US, UK, and EU Information Security Professionals	Semantic Security, 2003	US	Survey (and source – accompanying report)

Our original aim was to review research in relation to a number of information practices separately (data collection, data storage, data sharing and selling, data reuse and data analysis). However, a large proportion of the research looks at attitudes toward the initial phase of collection of personal data, and then more generally considers the subsequent handing or use of data. As such, they do not distinguish between individual data practices that occur following the collection of the data, and which lead to organisations finding meaning in the data and utilising it to meet their own commercial or public objections. The general focus on handling data is likely due to the novelty of big data processing. The consortium has, however, considered the results of surveys relating to public attitudes towards information practices with respect to consumer behaviours and e-commerce practices, health data practices and biometrics data practices, banking and financial data practices, social networks and search engines. Further, the primary focus of this report is European public sentiment, although we have also included survey results from non-European countries, such as the US, to provide some insight into the similarities and differences across continents.

2 PUBLIC SENTIMENT TOWARDS BIG DATA

2.1 OVERVIEW

Public sentiments play a role in determining which externalities are desirable and which may generate public resistance. These sentiments are important because they can indicate the extent to which the public, as a major source of data, willingly contribute to the big data process currently, and how willing they may continue to be. This is especially important because:

> Ever increasing amounts of data are being generated, at a faster pace and in more formats than ever before. The growing power to analyse vast and complex datasets can offer great insight into complicated issues, improving the quality of decision-making, delivery of public services, scientific research and many other areas.[8]

The views of the public of data processing that leads to this "great insight" is varied, but remains a useful way of informing the public and private sectors capturing the benefits of big data. This in turn enables them to forecast the growth and development of aspects of the European big data industry. The bulk of research into public sentiment illuminates data protection and data security as issues of major public concern. However, this may be directly linked to the presence of these issues in mainstream media, which regularly reports on privacy and security breaches, thereby heightening the levels of concern held by readers. A review of public opinion surveys across Europe as part of the PRISMS project[9] found that European citizens are not just concerned about the privacy of their data, but also the security of their data.[10] Results from that analysis show that citizens do not necessarily trust private and public organisations' abilities to safeguard their data; a fundamental societal implication associated with big data.[11] A 2011 survey on data protection in the EU also revealed that individuals across Europe are concerned about disclosing their personal information, although the amount of concern differs depending upon the age and gender of the data subject.[12] These concerns are apparent across a number of sectors and nations in relation to health, online consumerism, the use of social media and search engines, banking and financial institutions. For example, a 2009 survey from Norway revealed that citizens were concerned about the subsequent use of their consumer data by private companies, which is particularly significant when considering the implications of the use of big data to predict consumer behaviour.[13] This opinion can extend to relate processing techniques such as profiling and tracking that may result in negative outcomes for the user such as discriminatory practices (addressed below). Thus, exploring public sentiment toward issues that relate to big data is crucial to the wider

[8] Sciencewise, op. cit. 2014, p.1.

[9] PRISMS project reviewed surveys that provide interesting insights into both public attitudes and public behaviour regarding their attitudes towards matters relating to (for instance): consumer behaviour on the Internet, Internet usage, measures taken to enhance privacy and security on the Internet and attitudes towards surveillance technologies. For example, it includes Eurobarometer surveys that focus on European attitudes relating to public attitudes towards trusting others with managing their personal data.

[10] Watson, Hayley and David Wright (ed.) Deliverable 7.1: Report on existing surveys, Deliverable 7.1 of the PRISMS project, 14 March 2013. http://PRISMSproject.eu/wp-content/uploads/2013/03/PRISMS-D7-1-Report-on- existing-surveys.pdf

[11] Ibid.

[12] Watson and Wright, op. cit., 2013, pp. 30-31

[13] Brandtzaeg, Petter Bae and Markia Luders, "Privacy 2.0: Personal and Consumer Protection in the New Media Reality", *SINTEF Report*, The Norwegian Consumer Council, 2 November 2009. http://sintef.academia.edu/PetterBaeBrandtz%C3%A6g/Papers

examination of societal externalities of big data that the BYTE project aims to examine. This is especially so as public sentiment may conflict with business aspirations, which underlines the importance of considering public sentiment, perceptions and opinions related to information practices that produce benefits for big data industry.

2.2 PUBLIC PERCEPTIONS OF INFORMATION PRACTICES RELEVANT TO BIG DATA

Research suggests that the key public opinions regarding information practices related to big data centre on information privacy, data protection and data security. These opinions are generally held in relation to the collection of data in the first instance, rather than the subsequent use and processing by private and public organisations. Previous research and opinion surveys do not commonly identify public opinions in relation to distinct practices, such as data storage, sharing and selling, and analysis. Instead, they tend to indicate how the public feel about the general collection of data (mainly personal data) and the subsequent use thereof across a number of sectors, including health or banking and finance sectors for example, or in relation to social media and search engine use. It is also the case that most online users are not overly concerned by big data practices, but amongst those surveyed, there are a number of concerns that are most frequently mentioned. For example, respondents to the Eurobarometer 359 Survey revealed the following main concerns: that their information was being used without their knowledge (44%); that they could become a victim of fraud (41%); and that their personal information is being shared with third parties without their knowledge (38%). Respondents were least concerned about being discriminated against (e.g., in relation to job selection) (7%), their views being misunderstood (11%) and/ or their reputation being damaged (12%). More recently, the Public Attitudes to Science project revealed that respondents are largely opposed to their personal data being used for commercial gain. Although a majority of respondents seem relatively unconcerned about the use of their records in 'big data' analysis, there is strong opposition to some of the specific ways in which private companies might operationalise this data. For example, 62% of people oppose websites using people's online browsing histories to create personalised adverts for products in which people are more likely to be interested.[14] However, irrespective of public perceptions of how their data are used, there is an overriding interest in or sense of resignation to providing personal data for the free use of services such as search engines, email accounts and social media. This sentiment was acknowledged in a 2014 Ipsos Mori survey in relation to big administrative data practices within which respondents described how data is collected from them all the time, for example when using companies' services, interacting with the government, and making applications for jobs or courses.[15] That survey reveals that, in general, respondents were either unconcerned by this, or resigned to it, seeing the modern world as one in which people collecting data from or about you on a regular basis is "just part of life".[16] In fact, many of these services have become similar to public utilities and individuals rely on their ability to trade personal data for these services in order to participate in current social and communication activities. This perception is similar with respect to the use of surveillance technologies such as CCTV, by public enforcement organisations that, despite being privacy invasive, are recognised as providing security for citizens. The issue of government surveillance generally is topical in light of recent scandals involving the US National Security Agency ("NSA") as revealed by former contractor, Edward Snowden,

[14] Cameron, Daniel, Sarah Pope and Michale Clemence, *Dialogue on Data: Exploring the Public's Views on the Changes to the Census,* Ipsos MORI Social Research Institute, 2014, p.5 ("2014b"). http://www.ipsos-mori.com/DownloadPublication/1657_sri-dialogue-on-data-2014-census.pdf
[15] Cameron, Pope and Clemence, op. cit., 2014a, p.19.
[16] Ibid.

which have been featured in global mainstream media. What this means is that public perceptions and sentiments can differ depending upon the information available to the opinion holder at the time of forming the opinion. Further, public perceptions of information processes relating to big data can also differ depending upon whether the collector is a public sector organisation or a commercial organisation. Generally, authorities and institutions – including the European Commission and the European Parliament – are trusted 55% more than commercial companies.[17] Less than one-third of respondents to opinion surveys trust phone companies, mobile phone companies and Internet service providers (32%); and just over one-fifth trust Internet companies such as search engines, social networking sites and e-mail services (22%).[18]

Ultimately, what these perceptions mean for big data actors is that they must address these issues in their manner of operation in order to reduce negative sentiment and conversely, build user trust. In the absence of such practices, the amount of data willingly provided by the public, and (lawfully) collected, may decrease, especially as the public become increasingly informed about information practices relating to big data.

2.2.1 Data collection

A great deal of research into public perceptions of information practices relating to big data, or any sized data collection, relates to the initial stage of data collection. The public appear most knowledgeable about data collection, rather than data analysis or data sharing or selling. This may be because it is the point in the data cycle that is most tangible for individual users when they manually enter personal data or other information on websites such as consumer websites.[19] Users can encounter data collection across a number of different websites and across a multitude of online transactions in any one day. Subsequent information practices such as data storage, data sharing and selling, data analysis and data re-use are more opaque due to the technical aspects of these processes and possibly the limited the references to them during ordinary online participation.

Data collection occurs regularly and through a multitude of online transactions - from personal banking – to health - to social media – to search engine use. One survey revealed that a majority of Europeans were concerned about their behaviour being recorded via payment cards, their mobile phone or on the Internet. That concern might be related to the limited trust in commercial organisations that collect these data.[20] Whilst the collection of information by users is seen as necessary or worthwhile give the returns, the public is concerned more generally with whether subsequent use of the data collected has any relevance to the reason for the initial collection. For example, 34% of respondents to the Eurobarometer 359 survey said they were concerned that their information is being used without their knowledge and 23% were concerned about their information being used in different contexts from the ones that were disclosed to them.[21] Similarly, Demos found people's principle concerns to be companies using their data without their permission (80%).[22] This research indicates that

[17] TNS Opinion and Social, op. cit., 2011, p.2.
[18] Ibid.
[19] Although this is not to say that data collection does not occur through other means such as cookie use.
[20] TNS Opinion and Social, op. cit., 2011, p. 138. However note that residents of different countries display varying degrees of trust.
[21] cited in Sciencewise, op. cit., 2014, p.9.
[22] Ibid.

respondents, who may be reflective of the wider community, wish to be better informed of the purpose of data collection, as well as the contexts in which that data may be later used. This sentiment is reiterated by the results of the Eurobarometer 359 survey, which revealed that 65% of respondents believe it to be a "bad" thing if a search engine collects information about searches and uses that information to rank future search results, because it may limit the information a user is presented with.[23] Further, 70% of people surveyed within the European Union are concerned that personal data held by companies may be used for a purpose other than that for which it was collected.[24] A similar sentiment is shared amongst American online users. The Pew Internet & American Life survey in February 2012[25] included several questions probing respondents for how they feel about search engines and other websites collecting information about them to either shape their search results or target advertising to them. That survey revealed most search users disapprove of personal information being collected for search results or for targeted advertising.[26] Relevantly, search engine use is one of the most popular online activities: in February 2012, 73% of all Americans regularly used search engines.[27] However, whilst the public are not necessarily comfortable with the collection of data, they are aware that their online presence comes with the risk of tracking and tailoring future search results and advertising according to tracked online behaviours. This is despite indications that users see the sharing of personal information as a necessary part of being an online consumer and may be willing to provide personal information as a trade for a particular digital service. In that regard, a 2011 Eurobarometer survey found that 65 per cent of the public agreed with the following statement: "There is no alternative to disclose personal information if one wants to obtain products or services".[28]

Data collection remains topical with members of the public becoming less trusting of a wide range of institutions, whether it be the police, the banks or big government. This may be a result of scandals featured in mainstream mass media such as the recent scandal involving the NSA and Edward Snowden, as well as retailers involved in data and security breaches following the collection of personal data. This is especially so when scandals focus on privacy invasive activities that collect data as a means of carrying out surveillance on citizens. Companies and businesses conduct surveillance with cameras, whilst mobile phones sending location information to the network providers enables contextual advertising and mapping for commercial purposes. In particular, the covert collection of data by government organisations has been topical following the Snowden revelations of NSA surveillance practices. Whilst that was primarily relevant to Americans, a survey conducted by Pew Research Center, in 44 countries among 48,643 respondents from March 17 to June 5, 2014, confirms the rising distrust of US surveillance activities: "In nearly all countries polled, majorities oppose monitoring by the U.S. government of emails and phone calls of foreign leaders or their citizens."[29] Although a number of those surveys are sceptical of the U.S Government's

[23] However, this was not case for all survey participants. 29% said it is a "good" thing if a search engine collected information about searches and used it to rank future search results, because it displays more relevant results.
[24] TNS Opinion and Social, op. cit., 2011, p.2.
[25] Conducted during January and February 2012 among 2,253 adults age 18 and over.
[26] Purcell, Kristen, Joanna Brenner and Lee Rainie, *Search Engine Use 2012,* Pew Research Centre, Washington D.C., March 2012, p.2. http://www.pewinternet.org/files/old-media//Files/Reports/2012/PIP_Search_Engine_Use_2012.pdf
[27] Ibid, p.3.
[28] cited in Sciencewise, op. cit., 2014, p.6.
[29] Pew Research, "Global Opposition to U.S. Surveillance and Drones, but Limited Harm to America's Image", *Pew Research Global Attitudes Project,* 14 July 2014. http://www.pewglobal.org/2014/07/14/global-opposition-to-u-s-surveillance-and-drones-but-limited-harm-to-americas-image/

respects for civil liberties, a median of 66% across the seven European Union nations surveyed express a favourable opinion of the U.S.[30] These results indicate that Europeans may favour digital surveillance if they perceive there to be a benefit of such practices by governments, such as personal security and to combat to criminal activity.

Therefore, as the public seem to be better informed about the process of data collection generally, other than less transparent and obvious information practices (addressed below), it follows that citizens tend to hold stronger opinions about data collection. This is especially so when individuals are largely concerned about whether or not the data collected is, in reality, used for the initial purpose of collection.

2.2.2 Data storage

Different authorities (government departments, local authorities, agencies) and private companies collect and store personal data, which may also be of concern to data subjects. Such concerns are also a result of data storage being related to the issue of data security, which for individuals, translates into whether an adequate level of security is afforded to sensitive personal information once it has been collected. However, data storage is a matter left to organisations' internal processes and procedures, many of which the public know very little about. This lack of information may be why European users feel strongly about retaining some control over their data after it has been collected and once it is stored. For example, three-quarters of European Internet users surveyed say that they would like personal data that has been collected and stored through a website to be deleted at the users' discretion.[31] Respondents to the 2010 State of the Nation Survey[32] revealed a general distrust of government proposals for handling their personal information, including storage and sharing. More specifically, when asked whether personal information should be stored on a large computer system and shared across government departments, 34% felt this was a "very bad idea", while only 6% indicated that it was a "very good idea". ICM identified similar findings with regard to holding all medical records on a centralised computer system, with 29% indicated that practice to be a "very bad idea" and only 13% of respondents indicating that that it is a "very good idea". Government access to phone, e-mail and Internet browsing records saw even greater opposition with 55% of respondents thinking it was a "very bad idea" and 3% of respondents thinking it was a "very good idea".[33]

These public opinions underline trust (or distrust rather) as possibly the bigger issue relating to information practices employed by big data actors, and also a variance in the trust respondents have for public organisations versus commercial companies. For example, individuals surveyed revealed they are more likely to trust institutions storing their data such as health and medical care (78%) and national public authorities (70%), rather than shops (39%), communication companies (32%), and Internet companies (22%).[34] Perceptions of

[30] Pew Research, "Global Opposition to U.S. Surveillance and Drones, but Limited Harm to America's Image", *Pew Research Global Attitudes Project,* 14 July 2014. http://www.pewglobal.org/2014/07/14/global-opposition-to-u-s-surveillance-and-drones-but-limited-harm-to-americas-image/

[31] TNS Opinion and Social, *op. cit.,* 2011, p.172.

[32] Published in February 2010The survey covers a range of topics including public opinion regarding British government and the actions taken by the government, public opinion on government policies, questions concerning a proposed Bill of Rights, perceptions of surveillance technologies and individuals' political identities. This survey provides an understanding of British public opinion towards, among other topics, privacy, surveillance, and security.

[33] Cited in Watson and Wright, op. cit., 2013, P.91.

[34] TNS Opinion and Social, op. cit., 2011, p. 138.

data storage by public and private sector organisations indicate that more transparent storage processes are required in order to gain users' trust, so that they may continue to collect and store the information users provide.

2.2.3 Data analysis

Data analysis often occurs through automated means that monitor large volumes of data.[35] The ever-increasing volume, velocity and variety of data have reached the point that standard software tools and statistical skills are no longer sufficient for managing the size and complexity of data.[36] The growing power to analyse vast and complex datasets can offer significant insight into complex issues, improving the quality of decision-making, public services, scientific research and many industries.[37] In terms of public perceptions, data analysis falls into the general category of what organisations collecting data do with it after they have collected data, but there is little research that points directly to opinions, perceptions and sentiments of data analysis per se. As with the example of data storage, data analysis may be more opaque and less understood by users generally due to the technical nature of the process. Nevertheless, whilst data analysis is rarely discussed as an individual and distinct information technology practice, the sentiments towards it can be understood in light of a general distrust of certain organisations collecting data and what they subsequently do with that data.

2.2.4 Data sharing (including selling data)

Data sharing by organisations presents opportunities for those sharing the data, despite representing a concern for some members of the public, especially when personal data is shared or sold without the data subjects knowledge.. However, generally:

> people are generally happy for data to stay with one organisation, but are concerned when it's shared. They expressed a fear of the "master file" and the "data lock nightmare", where errors are perpetuated and you're trapped in a cycle of data you've given. There is a desire to have an element of control over their own personal information.[38]

Further, 33% of respondents to the Eurobarometer 359 survey said that they are concerned about their information being shared with third parties without their agreement.[39] Similarly, Demos found that 76% of people surveyed said a principle concern is data being shared with third parties.[40] In 2008, IIPS also found that people were more concerned about data sharing than the collection of data: "People are happy to give personal information: it's seen as a necessary part of modern life, and you get something in exchange for it, such as money off, or a quicker service. However, the worry comes later about how it's used, or could be used."[41]

Further, data sharing can occur through the sale of data as a source of generating income for businesses operating in the digital sphere, such as social media networks that sell user data to

[35] Lieberman Research Group, *Unisys Security Index: UK 2014*, 14 May 2014, p.11.
[36] Sciencewise, op. cit., 2014, p. 2.
[37] Ibid.
[38] Cited in Sciencewise, op. cit., 2014, p. 9.
[39] Ibid.
[40] Cited in Sciencewise, op. cit., 2014, p.9.
[41] Cited in Sciencewise, op. cit., 2014, p.6.

advertising companies. This is important because users may be unaware that the personal information they provide, when completing an online purchase or as part of an application or online subscription process, can be shared or sold for advertising purposes. Whilst this is ordinarily covered in the fine print amidst a number of other terms and conditions imbedded in a company privacy policy, the chances are that the extent to which a users' personal information is shared or sold is unknown to them. This may be why "the public is particularly concerned about losing control of their personal data, with fear that they will become a victim of fraud or identity theft, and that their data will be shared with others without their knowledge or agreement."[42] This is despite the fact that they may have technically agreed to it, without any real understanding as to what they were agreeing to.

However, there is significant nuance in the public's views depending upon the type of personal data in question:

> Even those who would typically accept their data being shared between public services, in order that they did not have to give it more than once, are extremely cautious about some types of data being shared. Bank account details and information about savings and pensions were found by IIPS (2008) to be particularly sensitive, with an even majority (74% and 63% respectively) of those who would typically be happy for their data to be shared, saying they were not happy for these types of data to be shared. There are also some important nuances within data types; for example, certain types of health information, such as mental health data, are considered by some to be particularly sensitive (Wellcome Trust, 2013*).*[43]

However, the sharing of personal data by government organisations and agencies, as distinct from big data companies, appears generally to be expected: "Participants commonly assumed that governmental administrative data is already linked and shared across departments, and supported this for operational uses."[44] In that regard, data sharing is common in the public health sector as it is considered as critical for the monitoring of public performance and for the powering the transformation of clinical services and development. Kingsley Manning, Chair of the Health and Social Care Information Centre confirms this:

> A common and quite reasonable assumption made by service users is that the NHS is a joined up organisation. That their information, their records, will be shared across care locations so that the care they receive is informed, safe and effective. […] Enabling the sharing of information between GPs and A&E departments, between patients and the GPs and between social care and healthcare professionals. Yet we know the formidable technical barriers to achieving this….interoperability, progress has been painfully slow.[45]

However, sentiments can change when health care data leaves the public sector and is sold to commercial organisations. This practice has been the subject of recent controversy with some observers expressing discontent with the provision of such data to intermediaries and indeed to pharmaceutical companies. In that regard, Kingsley opines:

> Quite rightly however, the public are suspicious that these arrangements are in some way unfairly tipped in favour of the profit makers. This suspicion has been fuelled by

[42] Sciencewise, op. cit., 2014, p.1.
[43] Cited in Sciencewise, op. cit., 2014, p.13.
[44] Ibid., p.6.
[45] Manning, Kingsley, "Sustaining Public Trust in Health Data", *The National Health and IT Conference and Exhibition (Speech),* 20 March 2014. http://www.hscic.gov.uk/media/13726/Kingsley-Manning-speech-HC2014/pdf/201314_HSC2014_FINAL.pdf

> our innocent lack of transparency. If we are to sustain public trust we not only need to demonstrate that their data is secure and that it is used effectively, but we need to be transparent in everything we do. The current arrangements governing the release of data are undoubtedly confusing and there is inadequate representation of the public voice in our decision-making.[46]

Such sentiments do not take into account instances where the sale and sharing of health data between public and private organisations aid the advancement of national and social care services. [47] Also, the sale of health care data to the pharmaceutical industry can be critical in developing new treatments and contribute to breakthroughs in treating deadly diseases Thus, transparency in big data processing activities is important, especially in relation to processing activities that implicate sensitive personal data. Transparency is useful in highlighting the positive societal externalities, such as health benefits, and can lead to a change in negative perceptions where they exist. In 2013, research undertaken by Deloitte found that some individuals say they would be happy for organisations to share their data with other organisations, where otherwise they wouldn't have been, if they were informed of how their data would be used for their or the public benefit.[48] Moreover, there is already some suggestion that that data subjects are more willing to support the sharing and sale of the data when a perceived benefit exists: "Opinion surveys suggest some willingness on the part of the public to trade-off their concerns against the potential benefits to themselves or the wider public."[49]

Therefore, despite some indication of negative public sentiments towards the sharing and sale of data, the health care sector provides good examples of how big data information technology practices can produce positive externalities for society. In turn, this can alter negative perceptions relating to technology practices when they are opaque, and data subjects are not adequately informed of the potential benefits they can produce.

2.2.5 Summary

Overall, some of the public perceptions referred to above in relation to big data information technology practices, including collection, data storage, analysis and sharing and selling of data reveal that the main area of concern relates to the privacy and security of data collected, as well as a general distrust of those handling their data, particularly companies operating in the private sector. Thus, there does not appear to be any overwhelmingly negative sentiments about specific practices. This may be because users tend to be more informed about the collection of their personal data as they play an active role in that process by providing the information, whereas data analysis, data storage and the selling and sharing of data appear more opaque to the user because the data subject is removed from those practices. However, companies and organisations can better inform their users with more transparent policies concerning the subsequent use of the data and specifically, the benefits that can flow from information technology practices. This is particularly relevant as research suggests that user concerns can be reduced when they are aware of societal benefits, rather than users resigning to the provision of their data in exchange for a service, even when they hold concerns as to the privacy and security of their data.

[46] Manning, op. cit., 2014.
[47] Ibid.
[48] cited in Sciencewise, op. cit., 2014, p.12.
[49] Ibid.

2.3 PUBLIC CONCERNS ABOUT BIG DATA USAGE

Though the opportunity offered by big data is great, there is also significant potential for big data to be misused and/or have unintended negative consequences for individuals and society. In particular, the explosion of data and the increasingly sophisticated way it is processed has raised concerns about how, when and why data is collected, stored, analysed, shared and otherwise used or processed To date, the major concerns about data raised by the public relate to data protection and the privacy of personal information, as well as data security. Some of the surveys reviewed for this report reveal that Internet users are not just concerned about the privacy of their personal information, but at times they are also concerned about the safety of their personal information on the Internet.[50] Research reveals that the public have also expressed concern in relation to tracking and profiling techniques that may negatively affect their lives or online participation, although research focussed on the extent of public concern for these issues is limited. However, the level of concern expressed for the aforementioned issues tend to differ subject to the industry, sector or the stakeholder specified, especially when it comes to sensitive personal data pertaining to health or financial circumstances (bank account details or pension details etc.). This is perhaps reflective of increased consumer awareness of privacy and data security related issues that are covered by mainstream media, as well as the European Commission's regulatory focus on safeguarding consumer data protection rights. For example, in 2012, Deloitte found that 82 per cent of the public report having some degree of awareness that private sector companies and public sector bodies collect data on people and their activities.[51] Public concerns are likely to have an increased as a result of increased awareness and generally speaking, "the public can be segmented into a number of groups sitting along a continuum between pro- and anti-sharing of data."[52] However, public concerns can be reduced when users are offered a specific personal or public benefit by the organisation or company collecting and using their data.

2.3.1 Data security

The issue of data security, particularly who has access to data, is a major concern for online users, especially due to the threat of fraud or identity theft. Public concern can lead to changes in online consumer behavior. The Unisys Security Index 2014[53] reveals that cyber security is the UK public's chief concern, with 85% of the population surveyed worried about bankcard fraud and 55% concerned about falling victim to identity theft. As a result, nearly one in 10 Britons have switched banks or retailers because of unhappiness with the way they protected their identity/privacy. Similarly, Demos found that among the principle concerns of those surveyed are companies losing their personal data (76%) and ID theft (70%).[54] However, the threat of cyber security attacks is ever present when "today's traditional security is proving ineffective against advanced persistent attacks on corporate networks via the Internet—ranging from data theft to phishing attacks, from breach of data to carefully-planned 'denial of service' attacks". [55] Thus, maintaining superior security monitoring, awareness and reporting capabilities in holistic cyber security frameworks to protect stored data from internal and external threats presents a challenge for organisations and companies, and the urgency of

[50] Watson and Wright, op. cit., 2013, p.127.
[51] cited in Sciencewise, op. cit., 2014, p.4.
[52] Sciencewise, op. cit., 2014, p.1.
[53] Lieberman Research Group, op. cit., 2014.
[54] cited in Sciencewise, op. cit., 2014, p.9.
[55] Lieberman Research Group, op. cit., 2014, p.11.

meeting this challenge is further prompted by levels of public concern for security. Data security is both an issue of public concern and a pressing organisational challenge. This is because:

> organisations and governments today confront potential security threats that didn't exist a generation ago. The community's sense of security is a critical determinant of public confidence in how governments and private organisations respond. Security threats are global and can impact any individual.[56]

In particular, respondents to a 2014 survey by conducted by Ipsos MORI regarding big administrative data revealed that keeping their personal data secure was very important to them and they worried about their data being leaked, lost, shared or sold by organisations that hold it.[57] Participants in that survey also felt that they had little control over their personal data. These general concerns about data security more widely drove particular security fears with relation to administrative data linking. For members of the public, data security translates into how adequately their personal information is kept safe and confidential.

Security is applied to a range of different contexts, including technologically secure systems.[58] The 2010 Unisys Security Index[59] identified moderate concerns over Internet security in eight of the 10 countries it examined, with the exception of the Netherlands and Germany. Of the eight countries where moderate public concern was identified, it was noted that concern is greater in relation to the threat of viruses than e-commerce, with the exception of Australia and the UK. Further differences in level of public concern were highlighted between Germany and Spain for example, where in Germany, public concerns are greater in relation to the threat of viruses and e-commerce security breaches than the level of concern evident in Spain. Spain and Australia were the only countries of the 8 countries examined where public concern over Internet security has increased between 2009 and 2010.[60] Further, this report previously mentions the general distrust members of the public surveyed in various surveys reviewed for this report have in public and private organisations to keep and use their data securely and appropriately.[61] In particular, with respect to the security of personal data, members of the UK public surveyed rank the following organisations from greatest to least trusted with personal data: financial organisations (56% trust), the NHS (50%), Employers (44%), central government (31% trust), private companies (23% trust), and service providers (only 16% trust).[62] However, levels of public distrust are decreasing in certain areas of data security. For example, just 30% of Britons recently surveyed by Unisys are seriously concerned about computer security in relation to viruses or spam, and this figure is observed to have dropped considerably since 2013.[63] Despite this decrease, data security remains a public concern, particularly personal data security. As Ipsos Mori observed "personal data security was very important to participants, and this framed much of the discussion".[64] These

[56] Lieberman Research Group, op. cit., 2014, p.3.
[57] Cameron, Pope and Clemence, op. cit., 2014, p.9.
[58] Watson and Wright, op cit., 2013, p.16.
[59] The *Unisys Security Index* is a regular survey conducted twice every year, the specific survey discussed here was published in April 2010. Every six months, the survey provides insight into the attitudes of consumers in ten countries in relation to four security issues: national security, financial security, Internet security and personal security. Although this survey has revealed that financial threats are the greatest concern, this analysis will focus on the other three topics; national, Internet and personal security.
[60] Review of survey results in Watson and Wright, op. cit., 2013, p.100.
[61] Sciencewise, op. cit., 2014, p.14.
[62] Lieberman Research Group, op. cit., 2014, p.9.
[63] Ibid., p.7.
[64] cited in Sciencewise, op. cit., 2014, p.9.

levels of public concern is vital information for big data actors to assist them in understanding the necessity of implementing appropriate and adequate measures to safeguard users' data to protect consumers from cyber threats such as identity theft and fraud and by doing so, address public concern in this area.

In the meantime, public concern over data security can result in citizens taking their own steps to protect their data by limiting the amount of personal information they disclose. This can mean a reduction in the volumes of personal data available to companies and organisations that rely on the collection vast amounts of data to meet business and organisational objectives. The Eurobarometer 359 survey provides insight into the various measures European citizens may be taking to enhance their security on the web, even if those measures are simply the measures that are most readily accessible to them, such as simply avoiding sharing information.[65] Thus, public concern over data security can result in members of the public wishing to gain back control of their data.

Therefore, public concern over data security makes it a vital consideration for big data companies and organisations seeking to capture the benefits of big data. Looking to the future, big data strategies can incorporate data security measures that address the major aspects of this public concern, namely that inadequate data security can lead to identity theft and fraud. Transparent and effective data security can build trust amongst users who will be more likely to disclose their personal data and less likely to implement their own personal measures that can have an adverse effect on the growth of the European big data industry.

2.3.2 Privacy (including surveillance)

Technology continues to modify the ways in which big data is collected, stored, analysed and shared, and in turn, raises concerns related to the privacy of personal data. In fact, data privacy is a major concern for members of the public, alongside data security, particularly in relation to online transactions. Data privacy is also of growing concern when personal data is collected through more overt methods such as surveillance. This section will examine levels of public concern for data privacy primarily in relation to online transactions, as well as touching upon the issue of public concern regarding surveillance activities. The potential for data protection breach and privacy invasion has in part resulted from technology that enables practices that can be privacy invasive, and also as a result of increasing awareness of surveillance activities that are privacy invasive. The technology that supports the development of big data as an industry through the collection of vast amounts of personal data, also presents opportunities to abuse the privacy of that personal data:

> The advent of large databases maintained by companies that specialize in collecting huge numbers of public records allows for the trivial monitoring and investigation of an individual. Data mining makes the process of inference cheap and easy, and the move from cash to credit cards, phones to cellular phones and paper mail to email make the task of investigating a particular citizen easier.[66]

The concept of "data privacy" or "information privacy" is ever evolving and has many varied definitions. As such, there is no one consistent "public view" on what constitutes personal data, although, "Data about who you are (i.e. personal information) is generally considered by

[65] Watson and Wright, op. cit., 2013, p.134.

[66] Gordon, Sarah, "Privacy: A Study of Attitudes and Behaviors in US, UK, and EU information Security Professionals", *Symantec Security Response White Paper,* Semantic Security, California, 2003, p. 6. https://www.symantec.com/avcenter/reference/privacy.attitudes.behaviors.pdf

the public to be more personal than data about what you do (i.e. behavioural data)."[67]. Respondents to the 2008 Eurobarometer Flash 225 survey[68] were asked which information and data they consider to be personal. Around three-quarters of the European respondents believe the following data to be personal data: financial information, such as salary, bank details and credit record (75%), medical information such as patient records, health information (74%), and their national identity number and / or card number or passport number (73%). Additionally, a majority of respondents agreed that fingerprints (64%), home address (57%) and mobile phone number (53%) are personal. Almost half of the Europeans surveyed consider photos of them (48%) and their name (46%) to constitute personal data. Close to a third of respondents believe heir work history (30%) and who their friends are (30%) are also personal data. Around a quarter of respondents also think that information about their tastes and opinions (27%), their nationality (26%), things they do, such as hobbies, sports and places they go (25%), and the websites they visit (25%) are personal.[69] Further, Demos and the Wellcome Trust suggest that public views on the use and collection of personal data sit along a spectrum.[70] Effectively, this means that public opinion is stretched across a range of positions, with the balance of opinion falling in different places depending upon the particular circumstances.[71] For example, an individual's age and social class both appear to have some bearing on their views on data, with younger generations typically sharing more but being less aware and older generations sharing less but being more aware, and higher social classes being more comfortable with sharing their personal data than lower social classes. More specifically, in 2013, the Wellcome Trust found that their focus group participants distinguished between types of personal data according to:

- · the degree of seriousness/risk if the data were misused or stolen;
- · the perceived level of security of the data;
- · Anonymous vs. personally identifiable data;
- · Recognition of the value of data collection (to self vs. to others) vs. unclear benefit;
- · free choice to create data vs. enforced/necessary existence of the data;
- · Government and non-government data.[72]

The aforementioned survey results are indicative of a considered approach to the disclosure of personal information and an awareness of factors that could exacerbate or minimise the risk of invasions of personal data privacy or other lead to other consequences for the data subject. Thus, individual perceptions of the use of personal data are influenced by a general awareness of data collection by government agencies and companies although there is a low level of what this means in practice.[73] At the time of writing this report, further research suggests that generally, public views of the collection and use of personal data by governments and companies can be summarised as:

- The public consider the collection and use of personal data to be a big issue;
- When asked, the public are ostensibly opposed to any form of data use and collection by government and companies;
- In practice, the public consider there to be no alternative to sharing personal information with government and companies in the modern world and expect it to increase in future;

[67] Sciencewise, op. cit., 2014, p.1.
[68] published in 2008 and examined public perceptions relating to data protection.
[69] See summary in Watson and Wright, op. cit., 2013, p.69.
[70] Cited in Sciencewise, op. cit., 2014, p.14.
[71] Ibid.
[72] Sciencewise, op. cit., 2014, p.14.
[73] Ibid., p.3.

- A significant proportion of the public expects to feel less comfortable about sharing personal data in future. [74]

Therefore, irrespective of the level of concern raised in relation to data privacy, and other personal characteristics that impact upon the level of concern held by a data subject, a majority of users see disclosing personal information as an ever increasing part of modern life, which is not so dissimilar to the 74% of Europeans surveyed in 2011.[75] Day to day disclosure of personal data can be open and deliberate in some cases, such as on social networking sites or in exchange for services. Alternative means of disclosure/ collection can be unintentional or hidden. An example of the latter is when behaviour is being tracked through websites, mobile phones or credit cards.[76] However, as with the issue of data security, the level of trust that survey respondents hold in organisations that collect personal data varies subject to the type of organisation or business collecting it and with that, the perceived level of privacy that will be afforded to personal information by that organisation or business. A 2010 Pew Internet & American Life Project survey[77] reveals important insights with regard to the extent to which American adults surveyed trust Internet companies, and that their distrust is linked to the collection and handling of their personal information as well. Thus, distrust is not just an issue of concern for Europeans, but a concern for citizens of other countries too. That survey revealed that Internet users were more likely to distrust social networking sites (65%), and younger adults (those aged between 18 and 29) were most distrustful of social networking sites. The category that attracted the most trust from users was news websites. Those between the age of 18 and 29 were more likely to "always trust" news websites (11%). Those between the age of 30 and 49 were more likely to trust newspapers and television news (6%) and those aged over 50 were more likely to always trust websites that provided them with health information (6%).[78] Thus, the level of trust users have in a company or organisations is linked to how well those users believe their personal data will be safeguarded.

The level of public concern for personal data privacy can have practical repercussions for companies and organisations that are least trusted by consumers. According Neil Fisher, Unisys', Vice President of Global Security Solutions, as at 2010, 40 million Europeans switched banks or retailers due to their concerns over protection of their personal data.[79] These findings are significant in a sector traditionally associated with low attrition rates.[80] Further, a Unisys 2014 survey reveals that 75% of British people will not shop or bank with people they cannot trust to safeguard their personal information.[81] This can potentially be the case across a number of retail sectors, although as previously highlighted, survey participants have suggested that they are less likely to take active measures such as switching service providers or banks if they perceive the privacy of their personal data is a trade-off for something of benefit to them. In that regard, Britons take a selective approach to online

[74] Sciencewise, op. cit., 2014, p.1.
[75] TNS Opinion and Social, op. cit., 2011, pp. 1 &11.
[76] Ibid., p.12.
[77] Madden, Mary and Aaron Smith, "Reputation Management and Social Media: How People Monitor Their Identity and Search for Others Online", *Pew Research Internet Project,* 26 May 2010. http://www.pewinternet.org/2010/05/26/reputation-management-and-social-media/
[78] Cited in Watson and Wright, op. cit., 2013, p.103.
[79] Unisys, "Unisys Security Index 2H10 Europe", *YouTube,* 22 October 2010. http://www.youtube.com/watch?v=lHqw--B8iOI
[80] Unisys, "Unisys Security Index: UK Customers Switch Banks and Retailers Over Privacy Fears but Admit to Shortcomings When Protecting Themselves", *Unisys Online*, 27 October 2012. http://www.unisys.com/unisys/countrysite/news/index.jsp;jsessionid=7BCEA3828EEBC9E5546591C586E9D613?cid=300008&id=1400004
[81] Lieberman Research Group, op. cit., 2014.

safety. While a majority limit access to personal information on social media sites, 39% admit to rarely considering privacy protection when shopping or banking online.[82] This reiterates a common the common acceptance amongst those surveyed that users must provide a certain degree of personal information in order to complete online transactions:

> Personal benefit is the strongest incentive for being in favour of the collection and use of personal data by government and companies, but the public report currently seeing little benefit from sharing their data and little confidence that they will see benefits in future. The public also identifies public goods (e.g. health research, prevention and detection of crime, and unearthing of dishonesty or fraudulent behaviour) as potential benefits of personal data use.[83]

Thus, unless members of the public see some benefit for them in providing personal information, collectors and users of big data would be at risk of a decreased amount of data at their disposal. In fact, citizens are already taking measures to limit the amount of personal data they provide as a result of concerns over potential misuse of their personal data, citizens are taking measures in order to protect their personal information and enhance their privacy. This has the potential to limit the amount of data available to big data actors looking to capture the benefits flowing from the big data industry, and at times, purposely providing false and inaccurate information as a result of their concerns. The Globalization of Personal Data Project and the Special Eurobarometer 359 identify the following measures that are being implemented by users to avoid or limit the amount of personal data they provide:

- Refusing to provide personal information to companies and government,
- Asking a company not to sell information,
- Asking a company to be removed from their marketing list,
- Reading online privacy policies,
- Avoiding the sharing of their user name and password, and
- Avoiding the disclosure of payment details online,
- Using anti-spyware; and
- Deleting cookies.

Less favourable measures to enhance privacy reportedly include:

- Purposefully giving false information,
- Asking to see what information is held on record,
- Asking for personal information to be removed,
- Using a dummy e-mail account and
- Shredding information.[84]

Thus, whilst users to protect their own personal information implement a number of the aforementioned measures, a number of the measures are aimed at distorting the information collected by companies and agencies. If big data companies and organisations implemented more transparent policies or indeed, could guarantee that they would better safeguard personal data, both the consumer and the data collector would benefit. The Eurobarometer survey 359 offers the following additional insight into what measures European citizens take to help protect their identity, which reiterates the level of public concern for their personal data. However, these additional measures are implemented by users primarily to protect himself or herself, rather than to distort the information available to a data collector that users hold little

[82] Ibid.
[83] Sciencewise, op. cit., 2014, p.1.
[84] Watson and Wright, op. cit., 2013, p.128.

or not trust in. Aside from generally avoiding sharing personal information with people or organisations individuals do not trust (47%), these additional measures include:

- Providing only the minimum required information (62%);
- Avoiding disclosing bank details or their pin number (56%);
- Avoiding sharing their user name and password (45%);
- Not disclosing payment details online (29%);
- Shredding private information, such as bills (29%); and
- Providing inaccurate information (7%).[85]

When comparing results by country, the Netherlands and the Scandinavian countries were more likely to have taken certain measures to protect their identities. They were also the countries that were less concerned about their behaviour being recorded. Measures were less likely to be taken in Southern European, Baltic and central countries, and Poland, Hungary and Romania.[86] Relevantly, public concern over the privacy of personal data was also found outside of Europe, examples include: Canada, the United States, Australia, Japan and China. Thus, concern over privacy is not simply a European phenomenon, but one that manifests globally.[87] Overall, public concerns relating to the collection and use of personal information seem to be linked to the lack of transparency surrounding the information practices relating to big data, which leave users concerned about the privacy of their personal data. Better practices by big data actors can marry organisations' information practices to address the levels of concern in members of the public have for data privacy. It may also enable a consideration of the aspirations that are held by their users, as this too can be a valuable source of how public concerns can be addressed. By abating public concern, big data companies and organisations foster public perceptions that build trust. This may lead to a reduced number of people implementing their own data limiting privacy measures, and eradicate the need they feel to purposely provide inaccurate or false data. However, data privacy and surveillance activities are not so straightforward given the covert nature of the activities.

Data privacy relating to surveillance activities is also gaining momentum as a public concern where personal information is obtained through surveillance technologies.[88] Whilst surveillance activities may primarily be undertaken with the objective of collecting personal data, surveillance methods, such as CCTV in public places, undoubtedly collect information that amounts to personal data. Surveillance activities are usually more convert and thus, perceived to be potentially more invasive than it is as such activities lack transparency. Public concerns are also heightened in relation to the surveillance activities, such as those undertaken by public organisations and agencies, with the intention of collecting personal data by more covert means. An example of the latter is the recent revelations of the NSA's participation of blanket surveillance of citizens. When news of the NSA's activities first broke, it also raised questions about government surveillance of other national governments and the surveillance of citizens in other nations. However, mainstream surveillance technologies, such as the use of CCTV, appear to have public support to the extent that they safeguard citizens' security. The PRISMS project examined 20 international public survey opinions regarding, amongst other things, public opinions of surveillance technologies. That examination revealed: eight of

[85] Cited in Watson and Wright, op. cit., 2013, p.111.
[86] Ibid.
[87] Watson and Wright, op. cit., 2013, p.127.
[88] For an examination of how technologies aimed at enhancing security are subjecting citizens to an increasing amount of surveillance and, in many cases, causing infringements of privacy and fundamental rights see the PRISMS project.

the 12 surveys provide evidence that some individuals respond positively to the use of surveillance measures to help enhance their security.[89] However, it was also observed that support for surveillance technologies waivered subject to the threats posed to individual security and safety: "respondents were not entirely supportive of surveillance technologies invading their privacy; rather, individuals' opinions were more likely to be supportive when there was a greater threat to security."[90] Overall, the findings of relevant public opinion surveys support the contention that "citizens are willing to give up some aspects of their privacy in the face of some surveillance technologies deployed to enhance their security. However, this does not necessarily mean that citizens trust the institutions implementing these measures."[91] This notion of trading off privacy rights for a societal or individual benefit, such as safety, applies many of the scenarios raised by this report where there exists potential for data protection breaches and invasions of privacy. As previously mentioned, a significant majority of the public appear willing to trade-off their concerns about the collection of many – if not all – types of personal data (including sensitive data such as medical records) if there is a strong enough personal benefit in return.[92]

Overall, data privacy is a major concern for members of the public, particularly in relation to online transactions. Whilst privacy is a major concern when considering the potential surveillance activities that are undertaken and directly or indirectly collect personal data, citizens may not be in a position to directly limit the personal data they provide as they may not always be aware that surveillance is being conducted. This, however, does not minimise the concern felt for those activities, it simply means that data subjects have less control over the provision of personal information. It is important to recognise that data privacy is a major public concern. There are two reasons for this: first, the big data industry is reliant to a certain extent on the consensual provision of personal data in order to capture benefits that flow from vast amounts of that data; and second, the issues raised in relation to this public concern provide big data actors with a valuable insight into what consumers value and how those values can be met before the trend to limit the provision of personal data grows to such an extent that it limits the growth of big data. Addressing public concern for data privacy in practice can ultimately benefit the actors and participants of the big data industry.

2.3.3 Profiling

Consumer profiling occurs when Internet companies that provide free online services, such as search engines, e-mail accounts or social media network accounts, collect information about users to create profiles of consumers' online preferences, behaviours and other characteristics that can be shared with or sold to advertisers. Advertisers can then use profiles to reach users through personalised or targeted advertising, and websites use information about consumer online activity to tailor advertisements or content to their hobbies and interests.[93] One of the most common methods of profiling is through the use of "cookies" that can be used to create a profile a particular user or computer across multiple websites. This is not a new method but a common means by which websites collect information about users' online behaviours.

Profiling represents a concern for European Internet users. A recent survey into attitudes towards administrative data revealed that 62% of people oppose websites using people's

[89] Watson and Wright, op. cit., 2013, p.135.
[90] Ibid.
[91] Watson and Wright, op. cit., 2013, p.9.
[92] Sciencewise, op. cit., 2014, p.1.
[93] TNS Opinion and Social, *op. cit.,* 2011, p.74.

online browsing histories to create personalised adverts for products that people are more likely to be interested in.[94] The Eurobarometer 359 survey considered public attitudes towards profiling resulting from the collection of information about them pertaining to preferences and online behaviour.[95] Generally, more than half of the Europeans interviewed, 54% of those surveyed, feel uncomfortable with Internet profiling, although four in ten of survey participants are comfortable with it (39%).[96] When distinguishing between respondents to the survey from different European countries, the survey revealed countries where two-thirds or more of respondents feel uncomfortable are the Czech Republic (72%), Germany (69%), Greece (68%), and Latvia (67%). The lowest percentages are found in Bulgaria (30%), Poland (34%), Romania and Ireland (both 40%).[97] Overall, in all but seven of the Member States, the number of respondents feeling uncomfortable is larger than the number of respondents feeling comfortable with Internet profiling.[98] Users of social networking sites and sharing sites are much more comfortable about profiling on the Internet. A relative majority of social networkers feel comfortable (48% vs. 47%) whereas 60% of Internet users who do not use social networks feel uncomfortable. Further, 47% of sharing site users feel comfortable (vs. 48% who do not) whereas 58% of Internet users who do not use these sites feel uncomfortable about it. Relevantly, purchasing online has no impact on the results with 60% of survey participants who shop online feel uncomfortable with Internet profiling.[99] Thus, there appears to be a correlation between the level of online activity and the degree of comfort with Internet profiling. This may be because regular users and sharers are more familiar with the concept of profiling because they can be better informed about it by way of privacy policies or other similar terms or conditions published on websites. For example, Google Inc. blatantly refers to user profiling in its privacy policy. That policy provides that when users are signed into Google's programs, their online behaviour can be collected and combined with other information collected about that user to form a cohesive user profile. This includes material from Google's search engine, the Google+ social networking site, YouTube video-sharing site, and Gmail. Since, 83% of all PEW survey respondents said that Google was their primary search engine[100], it follows that vast amounts of data are being collected for the purposes of creating consumer profiles. However, Google's profiling policy is buried amidst a myriad of other terms and conditions. In consequence, users may be unaware of this policy, and may have unknowingly created profiles that are being used to meet advertising and other commercial objectives.

Therefore, consumer profiling is useful to companies partaking in targeting advertising. However, research suggests that nearly half of people say they are unhappy about the use of profiling to achieve this form of marketing. This level of discomfort amongst users can lead to user modifying or restricting their online behaviour, which can translate into business models employing profiling techniques for commercial gain being less sustainable in the long term.[101]

---

[94] Cameron, Pope and Clemence, op. cit., 2014, p.5.
[95] TNS Opinion and Social, op. cit., 2011, pp.74 - 75.
[96] Ibid., p. 74.
[97] TNS Opinion and Social, op. cit., 2011, p.74.
[98] These exceptions are Bulgaria, Poland, Romania, Ireland, Italy, Finland, Portugal and the United Kingdom: TNS Opinion and Social, op. cit., 2011, page 74.
[99] TNS Opinion and Social, op. cit., 2011, p.75.
[100] Purcell, Kristen, Joanna Brenner and Lee Rainie, *Search Engine Use 2012,* Pew Research Centre, Washington D.C., March 2012, p.9.
http://www.pewinternet.org/files/old-media//Files/Reports/2012/PIP_Search_Engine_Use_2012.pdf
[101] Lewis, Harvey, Cecilia Liao and Neha Pandey, *Data Nation 2013: Balancing Growth and Responsibility,* Deloitte, UK, 2013, p. 4.

2.3.4 Tracking

Tracking represents another concern for European users of online services, especially as there is a large number of technical means by which tracking can be performed. Tracking is the analysis of visitor behaviour on a website, and can continue once the user has continued beyond that initial website visit. Four in ten Europeans surveyed in the Eurobarometer 359 are concerned about their behaviour being tracked on the Internet.[102] According to the results of that survey, a majority of Europeans are concerned about the recording of their behaviour via payment cards (54% vs. 38%), mobile phones (49% vs. 43%) or mobile Internet (40% vs. 35%).[103] Although the analysis of an individual visitor's behaviour in relation to the aforementioned activities may be used to provide that visitor with options or content that relates to their implied preferences, the tracking and profiling that enabled this is often still performed without the knowledge of the user.[104] This is of concern particularly as tracking can extend to other websites and mobile applications beyond that which launched the tracking activity. For example, Facebook partners with data broker firms to monitor both online and offline behaviour and specifically, to measure the relationship between the ads users see on Facebook and the purchases made offline.[105] Thus, online activities can be closely monitored, and even where users have not provided personal data when accessing services on the Internet, those users can be identified through the Internet Protocol (IP) addresses of their computer. In addition, users can be identified through digital cookies or electronic identifiers left on their browser by web sites. Internet communication and browsing tends to leave logs of web pages visited, e-mail and instant message senders and recipients, voice over IP callers, goods examined and purchased, advertisements viewed and searches.[106] In particular, tracking is an issue closely related to search engine use. The Pew Research survey revealed that 73% of American survey respondents say they would "not be okay" with a search engine keeping track of searches and using that information to personalise future search results because they feel it is an invasion of privacy, compared with 23% of those surveyed who said they would "be okay" with a search engine keeping track of their searches and using that information to personalise future search results. Another common way that users' behaviour is tracked online is via Adware (software that displays ads on the user's machine randomly, or that target ads based upon user profile) that is "piggybacked" with other, useful applications. One controversial piece of adware – and certainly one of the most well known – is the Gator Advertising Information Network (GAIN). This software provides several useful functions – and also can gather information about surfing habits etc. Gator is given as an example, however, because the End User License Agreement and privacy policy clearly describes the functionality of the software.[107] However, similar but less legitimate applications such as Spyware performs the same function as adware but without requesting permission from the

---

http://www.deloitte.com/assets/Dcom-UnitedKingdom/Local%20Assets/Documents/Market%20insights/Deloitte%20Analytics/uk-da-data-nation-2013.pdf

[102] TNS Opinion and Social, op. cit., 2011, p.74.

[103] Ibid., p.1.

[104] Although, tracking practices carried out within the EU are regulated by the E-Commerce Directive.

[105] "What You Really Agreed to in Facebook Terms and Conditions", *news.com.au,* 22 July 2014. http://www.news.com.au/finance/business/what-youve-really-agreed-to-in-facebooks-terms-and-conditions/story-fn5lic6c-1226997762948

[106] TNS Opinion and Social, op. cit., 2011, p.6.

[107] Gordon, Sarah, "Privacy: A Study of Attitudes and Behaviors in US, UK, and EU information Security Professionals", *Symantec Security Response White Paper,* Semantic Security, California, 2003, pp. 8. https://www.symantec.com/avcenter/reference/privacy.attitudes.behaviors.pdf

user prior to installation which enables the application operator to then silently track personally-identifiable information, and use this to modify content. It goes without saying that such applications pose concerns for users due to their covert nature. User concerns about tracking can also mean that their distrust or discomfort of associated techniques such as personalised advertising grows. For example, 68% of users surveyed by Pew Research replied, "I'm not okay" when asked how they felt about targeted advertising, based on the fact that respondents said they didn't like having their online behaviour tracked and analysed.[108] This level of discomfort has resulted in users are taking more responsibility for their personal and behavioural information online. For example, the Pew survey reveals that 65% of respondents had changed browser settings; 81% of respondents had deleted web history; and 75% had used the privacy settings of websites.[109] Whilst this can alleviate levels of concern amongst users, it ultimately limits he potential growth of industry for actors whose growth relies on the ability to track users and capture specific information about user behaviour.

As the big data industry develops, the value of, and demand for, gathering data on groups of users and individual behaviours for commercial purposes is likely to increase through tracking. Whilst there is a legitimate desire for online marketers and web site creators to tailor content and offers to users for commercial purposes, especially where it enables websites to offer services without charging a subscription fee, the level of consumer concern about information technology practices such as tracking remain an important consideration for data companies operating into the future.

2.3.5 Summary

Data security, data privacy, profiling and tracking represent areas of public concern. The level of public concern is related to the level of trust consumers have in the big data actors collecting their information. Furthermore, data security and data privacy are of particular concern in connection with personal data, particularly sensitive personal data such as medical information and financial information. As awareness of big data practices and the implications they have on users increases, big data actors can consider public aspirations towards big data to assist them in managing user concerns.

[108] Purcell, Brenner and Rainie, op. cit., 2012, p.2.
[109] Ibid., p.40.

3 PUBLIC ASPIRATIONS TOWARDS BIG DATA

3.1 OVERVIEW

Big data faces public suspicion because it is an emerging industry. Despite suspicion and concerns, consumers also largely consent to data collection, storage and analysis in exchange for convenient and subscription free access to online and other services as well as perceived benefits in areas such as health, crime fighting and public safety. Further, the public may be less suspicious of data collection and subsequent use of data if they are better informed as to the use of their data and the real implications of such use. Thus, public aspirations towards big data hint at the need for more transparent policies and practices by big data companies. Whilst this represents an immediate challenge for business and organisations handling big data, meeting public aspirations can in turn build user trust resulting in a more a secure and long-term relationship between data subjects and data collectors and handlers. Public aspirations are a means by which big data companies and organisations can understand how to adapt their processes to ensure that data continues to be available for their commercial purposes, as well as to produce public benefit and ensure the development of the big data industry.

3.2 PUBLIC ASPIRATIONS FOR INFORMATION PRACTICES/ TECHNOLOGIES RELEVANT TO BIG DATA

Despite negative public sentiment towards some of the information practices examined in this report, (such as data collection, data storage, and data sharing and selling), the public nevertheless holds aspirations for better services and more transparent and consumer-friendly practices. These aspirations are a valuable indication of how big data actors can innovate to marry their objectives with user aspirations to minimise public concern. The Unisys 2014 survey discussed the importance of minimising public concerns:

> In today's hyper-connected world, people are wary of losing valuable data and assets to cybercrimes and want assurances that their personal information is safe and will not be misused. It would be prudent for organisations to deploy solutions that not only enhance enterprise security with visibility across all aspects of security operations, but also inspire customer confidence to prevent loss of business and reputation.[110]

This is important because the users may limit the amount of data they willingly contribute if their frustrations are ignored. From the surveys reviewed, it is clear that users are more likely to impart their personal information, and act as willing participants in the big data industry, to the extent they derive a benefit from that participation. Alternatively, users can implement measures to limit the information they disclose, especially personal information. Further, greater transparency by big data companies and organisations is a common aspiration of users, some of whom feel uninformed about the extent of the user of the their information, particularly, personal data. For example, public opinion surveys reviewed indicate that users take issue with information about big data practices being buried amidst a sea of terms and conditions contained within privacy policies. Users also want to understand how they benefit from being the source of data that is ultimately generating wealth for commercial companies: "It just feels like I don't gain anything from letting companies use my data, none of them have ever told me how I benefit."[111]

[110] Lieberman Research Group, op. cit., 2014, p.10.
[111] Bartlett, Jamie, *The Public Must Be at the Heart of Any New Settlement on Data Sharing: the Data Dialogue* UK, 2012, p.36. http://www.demos.co.uk/files/The_Data_Dialogue.pdf?1347544233

3.2.1 Benefits for users and society

Nevertheless, there are significant potential benefits that big data can bring to society. BYTE Deliverable 1.3, *Big data initiatives*, outlines a number of potential societal benefits that big data collection and processing are expected to foster.[112] Specifically, policy-makers in Europe, the US and Australia, and on an international level, are supporting policy initiatives to enable gains in job creation, efficiency, information gathering and evidence-based policy-making, including saving people's lives through the use of data for improved crisis management and development as described in the UN Global Pulse initiative[113]. Big data initiatives also aim to generate new insights and discoveries in health and the sciences (see for example the European Bioinformatics Institute[114]). Finally, corporate entities seek to use big data to provide better and more personalised services, to assist industrial competitiveness and increase profits.[115] Members of the public support the use of big data for the realisation of many of these benefits, despite the concerns around privacy, re-use and profiling discussed above. Thus, big data practitioners and policy-makers need to find ways to support both big data practices and members of the public.

Those surveyed for public opinion surveys related to big data indicate that users are more likely to willingly participate in the big data industry by freely disclosing their data if big data companies and organisations are transparent about the how, where, when and why their data is processed, as well as the benefit/s available to users and to broader society that are produced by providing data. For example, when users are offered a specific public good, acceptance of data sharing can increase significantly. In addition to the public goods outlined in BYTE Deliverable 1.3, public survey findings identify health research, prevention and detection of crime, and unearthing of dishonesty or fraudulent behaviour, as benefits of personal data use. For example, UK National Health System focus group participants said data sharing within the NHS is considered to be positive, and that more data sharing could be done within the NHS[116] However, respondents to a 2012 Deloitte opinion survey[117] were split on whether public bodies should share more data about people between themselves, with 32% agreeing and 38% disagreeing that public sector organisations should share more data about people to improve the services they provide. Prior to that, in 2008, results of an IIPS survey revealed that 9 per cent and 42 per cent of respondents said they would be happy for "all information" and "some information" to be shared by public bodies respectively. However, 47 per cent of respondents to that survey said they would not be happy for any information to be shared.[118] Nevertheless, when specific public benefits are identified, respondents appear more willing for their personal data to be shared and used. For example, the public often agree with the use of personal data in contexts where there is a tangible public benefit, such as in medicine, transport and policing. Most people (56%) support combining the data held by multiple government departments and using them to better tailor public services to individuals.[119] Thus, survey respondents are more supportive of their data being used when there are tangible public service benefits. In 2013, the Wellcome Trust found the main

[112] Finn, Rachel, Anna Donovan, Kush Wadhwa, Lorenzo Bigagli, José María García, Big data initiatives, BYTE Deliverable 1.3, 31 October 2014. http://byte-project.eu/research/#wp1-setting-the-stage-on-big-data
[113] UN Global Pulse, *White Paper*: *Big Data for Development: Opportunities & Challenges*, May 2012. http://www.unglobalpulse.org/projects/BigDataforDevelopment
[114] The European Bioinformatics Institute, "Home", 2014. http://www.ebi.ac.uk/
[115] Finn, et al., op. cit., 2014.
[116] Cameron Pope and Clemence, op. cit., 2014, p.7.
[117] Deloitte, op. cit., 2012, p.19.
[118] Sciencewise, op. cit., 2014 p.7.
[119] Cameron, Pope and Clemence, op. cit., 2014, p.7.

benefits identified by members of the public, which would encourage their willing participation in big data, to include:

- The Government identifying needs, planning resources and services, and allocating funds;
- 'Prevention and detection of crime and, including terrorism;
- 'Tailored marketing;
- 'Identifying social/population trends and statistics;
- 'Convenience and time-saving when shopping and doing other transactions, if personal data were already held;
- 'Unearthing dishonesty (e.g. fraudulent benefit claimants and tradesmen); and
- 'Availability of vital medical information in a medical emergency. [120]

Specifically, medical records are a good example where users readily accept tangible benefits, even though generally, the public is sceptical about providing such personal data. What this means for big data companies and organisations is that they can better position themselves in the market by being transparent about their practices and the benefits they produce. It follows that the likely increase in data when tangible benefits are identified can result in innovation and development. Other such developments could include larger interoperable databases for more seamless sharing of data, which may be supported by users. In that regard, 91 per cent of respondents to a 2008 IIPS survey[121] agreed with the proposition that medical staff across the country should have access to their GP medical records, meaning that their medical history would be available to services if they needed medical care outside of their area. Almost as many (89%) agreed that medical records should also be accessible to the police and emergency services in order that they could be accessed if they were involved in an accident.

In relation to the provision of data to commercial companies, particularly personal data, a 2012 Deloitte survey revealed 29 per cent of users were in favour of doing so only because they received more tailored and personalised services or recommendations, with just 15 per cent thinking it would benefit society as a whole.[122] A mere 4 per cent wanted to provide their data to assist companies to do better, make more profit or be more efficient. In fact, Ipsos Mori found significantly more users support the use of personal data for public benefits than for commercial benefits, concluding that, "[p]eople on balance oppose personal data being used for commercial gain. Support for the collection and use of personal data therefore relies on individuals believing that they or wider society, not just companies, will derive some benefit from it."[123] This latter result provides invaluable insight for big data companies who can better focus on creating a more meaningful relationship with users.

Therefore, community aspirations for big data, and the related information practices, involve the provision of a mixture of both personal and public goods in exchange for data. Users are more willing to participate as data subjects to the extent that it benefits them. Thus, big data companies and organisations seeking to maximise their collection of data can place themselves in a better position to do so by identifying the benefits that can flow from the collection and use of data. This can also form part of a broader objective to be transparent about the use of data they collect.

[120] Ibid.
[121] cited in Sciencewise, op. cit., 2014, p.13.
[122] Deloitte, op. cit., 2012, p.9.
[123] Sciencewise, op. cit., 2014, p.8.

3.2.2 Transparency

Transparency is key for users to feel secure about the role their data plays in the big data industry. This is especially so when it involves providing personal data. An IIPS survey from 2008 found that people who reject data sharing between public services often do so because of a lack of information about the purpose behind it: "53% of rejections, as compared to 35% of acceptors, 'Don't know' why the government is keen for public services to share information about citizens."[124] Furthermore, research suggests that citizens are keen to have more control over the use of their personal data and want stronger safeguards in place through processes such as anonymisation.[125] This desire for more control is directly related to an unease felt by users about what happens to their data after it has been collected, which is product of a lack of transparency on the part of big data companies and organiations.

It is evident that aspirations are more concrete in relation to some sectors' use of data than others, such as data collection and use in the public health sector. An example of how public aspirations are being met by health care organisations is the "care.data" initiative to share personal data within the NHS and with some third parties. This is a result of the connection users feel to the public benefits derived from information practices utilised in this sector to achieve breakthroughs in medical treatments. Nevertheless, users still wish to be informed about the use of their data, and particularly, they wish to be in a position to consent to that use. Respondents to these various surveys appear to take quite a principled approach to the question of consent, with the vast majority of respondents to a Public Attitudes to Science survey say that was concern about the lack of explicit consent for the different uses that organisations might make of the personal data they collect.[126] In fact, a few respondents thought that the argument against seeking consent for the linking of administrative data on the basis that it would be extremely difficult and expensive to do so was not strong enough to warrant the linking of data without consent.[127] Additional aspirations relating to greater transparency include:

- Assurance beforehand that the information they provide would probably be kept confidential prompts just over six in ten (62%) to say they would be certain or more likely to provide their information;
- Leaflets giving information about the project in advance would inspire half the general public to consider allowing their personal health information to be used, whilst websites would have a lesser effect (36% would be more likely);
- Just over half (56%) say that information about the risks and benefits of a research project would make them more likely or indeed certain to allow their information to be used; and
- Six in ten would be more predisposed to allowing their personal health information to be used if they knew that the research it was intended for has the approval of an independent ethics committee.[128]

The Public Attitudes to Science survey also revealed that respondents still consider transparency to be vital in relation to the linking of administrative data by government departments[129].

[124] cited in Sciencewise, op. cit., 2014, p.12.
[125] Sciencewise, op. cit., 2014, p.16.
[126] Cameron, Pope and Clemence, op. cit., 2014, p.7.
[127] Ibid., p.33.
[128] Ipsos MORI, *The Use of Personal Health Information in Medical Research General Public Consultation,* Medical Research council, UK June 2006, p.9. http://www.ipsos-mori.com/Assets/Docs/Archive/Polls/mrc.pdf
[129] Cameron, Pope and Clemence,*op. cit.,* 2014, p.51.

Whilst transparency can be achieved through the use of privacy policies, they do not always achieve this result as they generally include a great deal of information in a way that is less user friendly than other practices. This can be because they are an avenue through which companies display their compliance with relevant data protection laws, rather than seeking to meet user aspirations. Privacy policies generally describe what information is collected, how it is collected, stored and shared, and how a person might manage such activities (by opting in or out, if possible).[130] Results of one survey revealed that because Google and Facebook privacy policies (for example) are too difficult for users to comprehend:

> Users surveyed don't understand which information is public; users surveyed don't understand how Facebook and Google track and store their information and activity; users don't understand how their information is shared and with whom. In fact, it is reported that users understand banks and government agencies better than they understand Google and Facebook, which is disconcerting given the high number of users each site has. Carnegie Mellon researchers determined that it would take the average person 76 workdays to read all the privacy policies they agreed to each year. So you're not avoiding the reading out of laziness; it's literally an act of job preservation.[131]

Perhaps for the aforementioned reasons, which may be relevant to many privacy policies and not just those used by larger companies such as Google Inc., few users read them. The Eurobarometer survey 359[132] revealed that just 58% of respondents read privacy statements; a further 25% stated that they "usually do not" read them; 5% stated that they did not know where to find them; and 8% stated that they ignored them. Researchers did not identify any significant differences for ignoring privacy statements across different socio- demographic variables. The most common reason cited for not reading privacy statements included:

- 41 per cent said it was sufficient to see that websites have a privacy policy;
- 27 per cent believe that the law protects them;
- 24 per cent think the website would not honour their privacy policies anyway;
- 15 per cent didn't know..[133]

Relevantly, 70 per cent of those who stated that they read privacy statements indicated that they had changed their behaviour as a result.[134] However, the survey did not include any additional questions that could have examined why or how they had changed their behavior.[135] Moreover, the issue of transparency is pertinent for commercial big data companies, not only because privacy policies are not always an effective means of achieving transparency, but also because the benefits of providing data to commercial companies is not perceived to produce any greater public good.

---

[130] Siegel + Gale, *Knowing more about privacy makes users share less with Facebook and Google,* Simplicity Lab Consumer Research Survey, March 2012, p.1.
http://www.siegelgale.com/download/8165c4e29090b0a827c34a26ab8f04f2/Privacy-Policy-Report-2012April_FINAL-online.pdf

[131] cited in "What You've really Agreed to in Facebook Terms and Conditions", *news.com.au,* 22 July 2014. http://www.news.com.au/finance/business/what-youve-really-agreed-to-in-facebooks-terms-and-conditions/story-fn5lic6c-1226997762948

[132] Cited in Watson and Wright, op. cit., 2013, 113.

[133] ibid., p.118.

[134] Cited in Watson and Wright, op. cit., p.114.

[135] Ibid.

Therefore, relevant survey results reveal that members of the public would like to be better informed of what happens to their data after it has been collected. If transparency is a key public aspiration of big data, then recognition of this by big data actors and the implementation of objectives to better inform users are vital to the development of the big data industry. If big data actors fail to recognise the importance of transparency, then they are less likely to build user trust. This can translate into users providing only the bare minimum data to proceed with transactions or utilise digital services because they simply do not know what will happen to their data.

3.3 SUMMARY

An examination of two major public aspirations for big data, namely the identification of benefits that flow from the provision of data, and transparency of why, how, when and where data will be used following its collection provides useful information to big data companies and organisations in the public and private sectors. The use of data by public sector organisations is more favourable to the public because of the aspirations they hold in relation to the benefits achieved through the collection and use of data by such organisations. This I particularly true when a tangible public benefit is readily identified, such as when data use produces improvements in public security or where developments in health care treatment and diagnostics is achieved. Ultimately, public aspirations for big data revolve around the collection and use of data, especially personal data, to be used by government and companies for their benefit, and in a transparent manner. Thus, users are more likely to willingly disclose a greater amount of their data if big data actors seeking to use that data to meet public or commercial objectives incorporate public aspirations into big data policies and practices.

4 PUPLIC PERCEPTIONS FRAMEWORK

Public perceptions of, and aspirations towards, information practices relating to big data are useful in developing a public perceptions good practice framework. A public perceptions framework that takes into account public perceptions and aspirations can contribute to the development and growth of the big data industry by ensuring that citizens, as a major data source, continue to comfortably and securely contribute to large data sets. To that end, it is particularly important that perceptions and aspirations towards data protection and privacy, data security and transparent practices be considered and implemented in a move towards responsible innovation, particularly in terms of implementing adequate security and being transparent about the information practices.

A good practice framework requires an all-encompassing security strategy. This strategy could involve data analytics to protect sensitive information. Such a strategy assists in building user trust. For example, building a practice framework around security is a goal of the UK Health and Social Care Information Centre because, "Because a key element in our strategy has been to sustain public trust on the collection, analysis and use of health care data. Indeed as I have said the HSCIC was created specifically to be a trusted, safe haven for the nation's most precious and personal data."[136] This approach also identifies the importance of giving the public a guarantee in relation to security. As Kingsley suggests: "Whilst we can never guarantee the absolute security of data we must give the public a guarantee that we as a system have taken all reasonable steps to protect and keep safe their data. We have no defence if we are found not to be compliant."[137] Good practice frameworks can also focus on the public benefit or good to be derived from the collection and use of data:

> Offering a specific personal or public benefit can significantly increase the general public's acceptance of the collection, sharing and use of their data by government and companies, but even when a specific benefit is offered, the public remain concerned about the collection, sharing and use of particular types of personal data (e.g. bank account, savings and pension details).[138]

In terms of a strategy for the increased protection of personal data (particularly sensitive personal data), transparency of how data is used and to whom it is disclosed ought be a focus, in addition to technical measures such as privacy by design. A strategy therefore involves more user-friendly and transparent measures to abate public concern that their personal information is being misused. This strategy could introduce the idea of more overt user control over their information. For example, in terms of browsing the Internet, there are many controls and configuration settings with web browsers that help facilitate privacy. In this regard, a focus on privacy enhancing technologies as part of a privacy-by-design framework is encouraged. As privacy is one of the biggest concerns of online users, Siegel + Gale make useful suggestions that may be incorporated into a relevant framework:

- Use simpler policies that inform and educate by conveying three main types of information (what information is collected and how; how the information is stored and hared; and how a user can manage their privacy). This is because the simple way to alleviate privacy concerns is through transparency;
- Standardize policies to save time and money, and possibly regulated;

[136] Manning, op. cit., 2014.
[137] Ibid.
[138] Sciencewise, op. cit., 2014, p.1.

- Let users opt-in to sharing and publicizing information. For example, the Obama administration recently announced a plan to make it easier for users to control online tracking of their personal information; and
- Design a feedback loop into digital interfaces. Designers of websites and applications should integrate feedback into their interfaces to raise awareness and inform users of potential privacy issues as they occur during use. For example, when a user is about to post a photo on Facebook, the site could tell the user how many people they're potentially sharing the photo with based on their privacy settings.[139]

Ultimately, these strategies also assist organisations and companies collecting and handling big data to reduce the risk and actuality of security and privacy breaches occurring. As this translates into to less negative publicity, the views of the public, that are susceptible to media influence, will likely change over time with users feeling more secure abut the role they play in the big data industry.

Overall, as users become increasingly aware of the risks associated with big data information practices, it is vital that big data industry actors focus on implementing good practice frameworks. Whilst the above examples are not an exhaustive list, the consideration of security, transparency and privacy by design, especially in relation to the collection and use of personal data, is a solid starting point. This is particular relevant in light of the negative public perceptions of big data information practices such as data collection, data storage, and data sharing and selling, and even more so in light of the fact that the publics aspirations towards big data are being expressed. To ignore these aspirations and not focus on a good practice framework will likely be of detriment to those wishing to capture the positive externalities of big data.

[139] Siegel + Gale, op. cit., 2012, p.5.

5 CONCLUSION

When exploring the social impact of big data, personal information privacy and data security appear to be the main concern for users, as revealed by a number of survey results relating to big data. The potential benefits of big data can be better realised if big data actors in both the public and private sectors take heed of the information pertaining to public perceptions of their practices, as well as aspirations for big data. In doing so, big data companies can build user trust. If the majority of the public trusts the collectors, sharers, users and analysers of the data, they will be more willing participants in the big data industry than they currently appear to be. The implementation of a good practice framework is how big data actors can show users that they are a valued source of data. If this does not occur and users "opt-out" of sharing their data, the future of big data industry may be limited. This is particularly relevant to private organisations as the public is less trusting of commercial organisations than of public organisations that produce what the public perceive to be tangible benefits such as better health care. Central to this is ensuring that security of data and in particular the safeguarding of personal data (both sensitive and non-sensitive) is at the forefront of the technological processes.

The public perception identified relate to the information technology practices such as data collection, storage, and data sharing and selling. Users seem better informed about the process of data collection generally, other than less transparent and obvious information practices (addressed below), it follows that citizens tend to hold stronger opinions about data collection. This is especially so when individuals are largely concerned about whether or not the data collected is, in reality, used for the initial purpose of collection. Otherwise, there is not a great deal of information about user perception relating to other specific practices such as data storage, analysis and sharing or selling. However, surveys have revealed a stronger negative sentiment towards the sale of data, particularly from the public health sector to private data companies, as well as concern for the security of data during its storage. Understanding the potential social impacts of big data, and implementing measures to change poor public perceptions and negative sentiments relating to information practices for big data are important for a number of reasons. Data information and knowledge are critical for the sustainability of the European big data sector of society. Further, significant developments across a number of sectors can be driven by access to, and use, of accurate data. The public is more likely to contribute accurate data if they feel secure doing so. This is particularly relevant in the public sector where taxpayers fund big data initiatives.

In addition to public perceptions of information technology practices relating to big data, research undertaken for this report identified a number of major concerns for users. These concerns include data security, data privacy, profiling and tracking. The level of public concern is related to the level of trust consumers have in the big data actors collecting their information. Furthermore, data security and data privacy are of particular concern in connection with personal data, particularly sensitive personal data such as medical information and financial information. As awareness of big data practices and the implications they have on users increases, big data actors can consider public aspirations towards big data to assist them in managing user concerns.

However, despite some negative sentiment towards information practices related to big data, and the identification of a number of areas of concern for users, users are generally willing to provide their data in exchange for individual or public benefits. However, big data industry can benefit by fostering a more positive interaction between big data actors and users. One of

the ways in which this can be achieved is through the recognising public aspirations, particularly the delivery of benefits and transparent practices, by incorporating them big data practices and policies. This is also reflected in the suggestions made in relation to a good practice framework. However, big data practitioners should not seek to convince members of the public to "trade" privacy or other considerations to achieve these potential benefits, instead, initiatives should be developed that incorporate privacy protection into these systems to enable the gains associated with big data and address the significant concerns expressed by members of the public in these surveys.

Therefore, positive public sentiments towards big data are imperative to the continuation of data processing activities processes, and the future of big data as a value adding institution/ process. Personal benefit is the strongest incentive for being in favour of the collection and use of personal data by government and companies. Conversely, if the public see little benefit from sharing their data and little confidence that they will see benefits in future, this may hinder the amounts of data available to big data actors into the future thereby, threatening the longevity of the European big data industry. Public sentiment towards issues that relate to big data is crucial to the wider examination of societal externalities of big data that the BYTE project aims to examine.