Research on Identity Ecosystem

Decentralised Citizens ENgagement Technologies

Specific Targeted Research Project Collective Awareness Platforms

d-cent

FP7 – CAPS
Project no. 610349
D-CENT
Decentralised Citizens
ENgagement Technologies

Lead beneficiary: Nesta

D3.3
Research on Identity Ecosystem

June 2015
Version number: 2

Authors:
Francesca Bria (Nesta)
Javier Ruiz (ORG)
Gemma Galdon Clavell (Eticas)
José María Zavala (Eticas)
Laura Fitchner (Eticas)
Harry Halpin (W3C)

Additional research:
Grégoire Delette (ORG)

The work leading to this publication has received funding from the European Union's Seventh Framework Programme (FP7/2007 – 2013) under grant agreement n° 610349.

Content

Executive Summary

This report presents an in-depth analysis of the latest evolution of the Identity Ecosystem in the Big Data context, focusing on the economic value of data and identity within the current digital economy.

This study aims at understanding the philosophical, economic and cultural implications of machine-mediated identity systems, focusing on the role of identity in the current economic system, and the way individual and collective identity in the form of personal and social data is mined and analysed through machine learning algorithms to predict future economic and societal trends, in this way redefining financial evaluations. The aggregated data extracted from the analysis of the identity and behavioural patterns of the user, is analysed in depth with the objective of maximising value extraction (e.g. for marketing, social control, and surveillance). A broader investigation and the understanding of the implication of such mechanisms are crucial for the understanding of future knowledge-based economic models and for the design of alternative effective instruments of social interaction.

This research offers an exhaustive multidisciplinary framework, tackling key conceptual issues on the evolution of the concept of identity and its role in the current digital ecosystems. At the same time however, it offers practical and regulative integrated examples of models of self-governance of identity, in the context of the knowledge-based economy. In the current internet digital ecosystem we are observing a battleground between a small number of closed, proprietary, and vertically integrated platforms mainly based in the US. Digital networks represent the space of widespread social cooperation and new forms of democratic organisation and at the same time the new attempt to capture the power of collective intelligence by a capitalism based on the biopolitical production of the common. A few private actors manage the identity, the forms of communication and the social relations of the connected multitude of users. This study investigates how to escape this and claim a free collective production for a wealth that is equally distributed (data commons and privacy-aware identity infrastructures). The internet must remain a social grassroots space for collective intelligence to thrive, and therefore must be re-appropriated to build a new kind of democracy, and to organise a new common.

In order to emphasise the benefit of these alternative models based on the commons it is necessary to move from a transactional paradigm that sees personal data as a new "asset class" to a relational and ecological paradigm that considers social data as a common that can valorise the social cooperation of communities and re-appropriate the collective value generated by citizens and invest it for social good. This requires transforming personal data to social data with the appropriate open technical standards for access control.

This study presents an initial review of the concept of identity alongside a concrete analysis of the economic, policy, and technical alternatives to develop an identity ecosystem and management of data for the common good that respects citizens' rights, privacy and data protection. This research also presents a map of the key players in the identity industry (such as data brokers and data aggregators), including empirical case studies in key sectors, showing how identity is managed in practice. The socio-economic analysis is tightly integrated with the reflections at a legal and technical level. Technical solutions do not work by themselves, therefore legal and business solutions must be based in technology and integrated with the appropriate policy framework

This research has a direct impact on the design and implementation of the D-CENT platform (WP5, in particular D5.4), since it will give central importance to users' ownership of data and to communities ability to own, share, contribute to data and build large scale collaboration and collective awareness, while at the same time keeping control over common resources to achieve empowerment. D-CENT will also experiment within the platform architecture, developing novel ways to preserve trust, privacy and data ownership in big data environments.

The report will be organised in six sections. In the first, entitled "The identity Ecosystem in the Big data and mass surveillance paradigm" we will see that identity is a complex collective issue that includes questions of autonomy, freedom and surveillance that need to be fully analysed and understood.

In the second section we will go in-depth into the Socio-Economic Framework, and analyse Reality vs Myth concerning the role of data in the Digital Economy.

The third part will deal with an empirical analysis of the emergence of an "**identity market**" where personal data emerges as a valuable commodity, and where new actors such as "**data brokers**" have a major role to play. This part takes a broader perspective on identity taking into account externalities, social construction of value, etc.

The fourth part presents some specific empirical case studies in the field of consumer financial data, sharing economy, digital identities in public service provision, political profiling and personal data market in e-education.

The fifth section provides a concise overview of the regulatory frameworks and standards existing in Europe in the context of new technological development and the European Digital Single Market (e.g. EU Privacy and Data protection Directive, E-privacy Directive; Competition Law and other relevant European regulations affecting Digital Identities). The report focuses on some key issues within EU Data Protection, such as Personal Data, anonymity and pseudonymous data, Privacy Policies, Data portability, Data Protection by design and international data flows.

The final section outlines economic, policy, and technical alternatives for identity, looking into pragmatic alternatives to preserve trust, privacy and data ownership in today's big data environments. It looks into access to data, economic strategies to manage data as commons, consent and licensing, tools to control data, and terms of services. It also looks into policy strategies such as privacy and data protection by design and trust and ethical frameworks. Finally, it assesses technical implementations looking at identity and anonymity, cryptographic tools; security; decentralisation and blockchains. It also analyses the future steps needed in order to move into the suggested technical strategies.

The development of detailed guidelines to be used by D-CENT collaborative platforms is out of the scope of this project, however this work directly impacts the implementation of specific tools developed by the project (in particular D5.4 and D5.6) following the W3C guidelines for the management of personal data, data portability, identity management and security (see D-CENT publication for a general introduction to the topic: http://dcentproject.eu/wp-content/uploads/2014/01/D4.1-State-of-the-Art_new_2.pdf).

Throughout this journey, in the six sections different crucial aspects relating to the forms of regulation open to guarantee a new approach to the management of identity that is privacy aware and distributed. This report will guide developers of collaborative democracy tools understand their

position in the wider socio-economic system of personal information and digital identities. In particular, this analysis offers the D-CENT project possible models of democratic and distributed management of data and common infrastructures that are at the base of the experience of shared democracy in Spain, Iceland and Finland, with the aim of achieving middle and long-term sustainability. Specifically, this research into the market of identity formulates an opposing claim of social data as a digital common good and the need for developing public and common infrastructures of information and communication not based on the logic of the market and surveillance.

Obviously, the success of a new approach to manage identity and personal and social data as a common good is a complex process that must rely on an integrated technical, legal, economic and policy approach. In this context, it becomes more and more essential and urgent to define the terms of an alternative model of producing, managing and regulating knowledge commons in collective awareness platforms.

1. Introduction: The identity ecosystem in the big data and mass surveillance paradigm

1.1 Identity as a complex collective issue

"*On the internet, nobody knows you are a dog"*. This is one of the most repeated memes online,[1] and the one that best grasps the problems and promises of digital identity. As we digitise many of our social and economic activities into networked computer systems, simply transferring our basic ideas of identity in the physical world does not work. There are many ways to understand what a digital identity is.

Since the beginning of the development of inter-networking protocols, digital identities have been a critical component for these systems to operate. The most basic digital identity on the Internet are the Internet Protocol (IP) addresses that traditionally would uniquely identify a computer. But **the technical aspects** of online and digital identities have evolved into a huge and very complex field, including authentication, and authorization[2] well beyond network engineering and technical interoperability[3]. There are many technical committees in international standards organisations such as the World Wide Web Consortium working on identity issues[4] in order to keep the internet and the web running. For an overview of the technical questions around identity please see the D-CENT paper on the State of the Art of identity systems, social networking and social data stores[5]. An update on the technical aspects of identity will be also given in section 6 of this report.

Being able to guarantee that a system is dealing with the right person behind the computer is a basic requirement for different types of transactions: social, economic, and administrative. Hence governments, banks, social media platforms and many specialist communities building particular types of digital services, from health to e-commerce, have been driving the quest for secure identities. For example, proponents and detractors of online voting for digital democracy elections have their own set of issues and requirements in trying to tell a dog from a genuine voter.[6] The developers and citizens using the D-CENT digital platforms will have to grapple with many of these issues as well.

At some level, the need to **establish a digital identity becomes a legal requirement**, for example on e-government platforms that allow you to apply for a driving license or passport; or simply in online commerce. The elements of identity involved have been termed "transactional identities", the information required to perform a certain operation.[7]

Here is one of the areas where discussions about identities diverge more sharply. On one side we have the view that we must adapt traditional ideas of identity, such as the ID card and ID number, to the digital age. This means that national governments should be the main providers of digital identities. On the other side are the proponents of using multiple identity sources and other more decentralised systems that do not rely on state assurance. These could provide a better balance between publicly controlled resources and personal data owned by citizens as commons.

For individual Internet users, the current idea of digital identity appears very differently, sometimes simply as a "the permanent collection of data about us that is available online. Each time we post a picture, a blog, a status or a tweet, we are adding to that digital identity".[8] This description will

instinctively appeal to most Internet users, particularly those who have only experienced the net mainly through social media platforms. This capacity for **self-representation** is what has made fluid online identities a key aspect of the promises for freedom of expression brought by the Internet.

But there are growing concerns over what happens with that personal information and the potential to cause harm. For example, it is now established that social media postings can affect employment chances.[9] There is also the widespread practice of "doxxing" - sharing someone else's personal information online in order to discredit them, normally under the argument that the person is some bigot or reactionary[10] - which raises huge ethical (and legal) issues.[11]

In addition, the picture above is not complete, as there is much more data out there than what we ourselves put out intentionally. While it could be argued that ultimately everything about us online is derived out of our behaviour - e.g. our clickstream or detailed internet history - much of it is derived through complex analytics. The potential lack of control over that information is even more concerning on self, mind and society, so much so that scholars are looking into ways to prevent us from becoming slaves to Big Data[12].

There is a clear conundrum around digital identities in their potential for both **freedom and control**. These complex social issues have developed into a huge field. There are now whole academic programmes dedicated to digital identities, such as the Nottingham based Horizon centre, with funding for 80 PhD candidates working on this topic. They are looking at themes such as digital identities centred on particular geographical places, personal movement profiles for more sustainable buildings, and digital identities of vulnerable people, such as adult social care.[13]

In addition to those newer themes, there are some recurrent discussions around digital identities that have been around since the birth of the web and constantly resurface. **The Anonymity and Accountability** debate pits concerns about the dangers of anonymity on one side - from bullying and trolling to crime and terrorism - and in the opposite camp those worried about the impact on freedoms of the reaction towards real identities and calls to make anonymous internet usage impossible. But internet security experts such as Bruce Schenier caution against presenting the dichotomy in too simple terms where trust always requires a persistent identity and anonymity will always lead to a social breakdown.[14]

The debate rages on. It seems that as the number of internet users grow and interactive web and social media platforms allow for more participation we see a corresponding growth in online abuse, in many cases this is directed at women or vulnerable groups. This is why juridical and legal scholars together with technologists of the likes of Tim-Berners Lee are advocating for the need of a new declaration of Internet Rights[15]. This new Magna Carta is now being shaped as a foundational document that should include the protection of personal data and the right to the informational self-determination. It should also include access, neutrality, integrity and inviolability of IT systems and domains, mass surveillance, development of digital identity, rights and guarantees of people on internet platforms, anonymity and right to be forgotten, interoperability, right to knowledge and education, and control over internet governance.

The security and anti-terrorism agenda is another major driver in this debate. The recent rise in Islamic radical online activism of ISIS supporters have led to renewed calls, e.g. by David Cameron, for the security services to be able to monitor personal information online[16].

The leaks by US security contractor Edward Snowden of unprecedented information about the extent of online mass surveillance have sharpened attitudes to these issues, which are some of the defining aspects of digital identities[17].

Finally, but not less important, is the consideration that the information we leave all around in the physical and digital world is now being exploited through a **new economic and technical revolution**[18]. Since the first descriptions of the information society, the economic importance of data in the digital economy has grown exponentially. Big data[19], the Internet of Things, business intelligence, are all terms associated with the attempts to capture the wealth released by the explosion of data generated by devices such as mobile phones or sensors, and by people using social media and the web[20].

But it is not simply a matter of sophisticated analytics building more complete personal profiles based on our data, which is the traditional realm of privacy, our personal information is now used to generate insights - and corresponding monetization of the life of the users - about a myriad other things, ranging from improving internal business processes to finding the cure for cancer. The recent breakthroughs in artificial intelligence by companies such as Google and Baidu hinge on machines being able to learn by being fed masses of information[21], which we provide when we use these services and generate future economic predictions, while being able to monitor, inform and nudge citizens in real time.

The issues of privacy and freedoms that we introduced above also apply here, but in a more complicated way[22]. For example there are concerns about potential discrimination based on class profiles that do not necessarily require a fully personal identification. Elements of our digital identity, such as ethnicity, age, gender, social media profiles or internet history could be enough to make us receive differential treatment as regards our ability to get a mortgage, a private insurance, or to be able to access public benefits.

There are also issues of **economic justice**. Ordinary people provide the data that fuels this big data economy, but it is unclear whether the benefits are distributed fairly.[23] Today a handful of non-European internet giants control the entire digital infrastructure from data centres, to Cloud, to social networking and App ecosystems. This raises some shadows about the positive and emancipatory nature of the internet, as the quest for more data has generated unprecedented levels of economic surveillance that have been defined by critiques as "surveillance capitalism"[24].

Almost everything we do on a computer is now expected to be recorded, analysed, and eventually monetized. In this context there is a growing movement of people trying to reclaim the economic, societal and ethical value generated by these processes for the public good. We discuss these issues in more detail, throughout the text.

1.2 Identity, privacy and surveillance by business and governments

1.2.1 Historical background

Concerns about surveillance and privacy started well before the existence of computers. Since the introduction of the population census in the nineteenth century, **governments have continued to amass and process unprecedented amounts of information**, but not without some reaction. For example, attempts to introduce a national population register in the UK after First World War were resisted as "Prussian" in nature, although eventually the British government would achieve most of their aims through a combination of alternative schemes. Part of this process was the attempt to give each individual a unique name.[25] Government information systems became increasingly mechanised through the use of punch hole cards and associated machines to control the population and its identity. But another trend was the increasingly close collaboration between

information technology companies and governments. The grimmest example of this systematisation of information was the well-known case of **IBM's collaboration with Nazi Germany**, which helped in the extermination of European Jews. Recent documents have shown that IBM went as far as to create a "code for death by Gas Chamber"[26].

Since the 1950s there has been an exponential growth in the **collection of data by businesses**. The increasingly competitive business environment has led to more aggressive marketing with a focus on perfecting the understanding of customer needs. Putting the customer at the centre has paradoxically driven the quest to know more about them through **the creation of large databases**[27]. In the 1970s credit databases and geographical demographics models paved the way for the modern customer segmentation systems in use today, such as **Experian's Mosaic** that aims to "treat people like individuals"[28]. The growth of internetworked computers since the 1980s has led to the linking of personal databases into interconnected data systems. Since then there has been an explosion in the generation and processing of data, as we discussed in the previous section.

1.2.2 Implications of surueillance on priuacy and data protection

US jurists Warren and Brandeis famously defined privacy as the "right to be let alone" and the basis for all other freedoms in 1890.[29] This sets the idea of privacy in some defined private space - e.g. the home - that stands in contrast to the public space. Private space is where we are meant to be who we really are, our true self, which is an extension of our mind. But as explained by Julie Cohen[30] privacy is not there to protect a static self but the on-going development of our personas: "Privacy shelters dynamic, emergent subjectivity from the efforts of commercial and government actors to render individuals and communities fixed, transparent, and predictable. It protects the situated practices of boundary management through which the capacity for self-determination develops." [31]

In this view, the surveillance by governments and companies that we described in the previous section does not simply cause specific harm, such as political reprisals or denial of credit, but **undermines the basic foundation of the autonomous self** that is required for all forms of social interaction. The 1970s saw the first reactions to these growing asymmetries of information and corresponding power imbalances between governments and companies on one side and citizens on the other. The "fair information practice principles" (FIPPS) which form the basis for all modern privacy laws were first codified in 1973 in the US,[32] and include: (i) no secret record keeping (transparency and openness), (ii) individual right to know what information is kept about him and how it is used (participation), (iii) information obtained for one purpose cannot be used for other purposes without consent (purpose limitation), (iv) individual right to correct or amend a record, (v) organisations must assure the reliability of the data for their intended use and prevent misuse (integrity , quality)

These ideas were further developed in Germany into the concept of **"informational self-determination"**, which has greatly influenced **European data protection laws**, which we discuss in section 2.1. In 1983 the German Federal Constitutional Court issued a seminal ruling - in a dispute about the census - setting out the framework for "the authority of the individual to decide himself, on the basis of the idea of self-determination, when and within what limits based on the principle of self-determination to determine in what information about his private life should be communicated to others and to what extent."[33]

The internet has further complicated the idea of privacy as boundaries, but the fundamental aspects of informational privacy remain valid. The spatial concept of privacy of the nineteenth century has

developed into a discussion not over access to a physical space, such as the home, but as a control over information and identity. And as we can see in the examples in the previous section, **the developments of informational privacy and surveillance are completely linked to considerations of identity:** unique names, ethnic classification, individual profiles, etc.

1.2.3 Today's Identity marketplace: "Surveillance Capitalism"

The collection and analysis of huge amounts of personal information is critical for most digital companies to establish a competitive advantage in the market. But even if the data is not strictly necessary for current processes, most companies feel compelled to collect it. Startup companies will harvest data on demands from venture capitalists, while consultants advise established companies to invest in data analytics tools for economic predictions and the corresponding data collection. This is a cultural shift towards **data hoarding**. The implications of this situation were discussed in a D-CENT workshop, which we summarise in the rest of this section, and in more detail in the workshop proceedings published as part of this research[34].

The huge financial and technical resources needed for managing such massive amount of information, together with complex network effects lead to the formation of global digital oligopolies. For many people the internet in practice is reduced to a few platforms they use most of the time: Google, Facebook, Amazon, etc. This new wave of technology companies from Silicon Valley, with their mottos of not doing evil and the influences of 1960s West Coast alternative culture, appeared to be completely different from the previous incumbents, e.g. IBM and telecoms giants such as ATT. But there is a growing understanding that companies such as Google represent a new form of capitalism that may have improved competitiveness, efficiency, and access to knowledge in certain areas, but is not without problems. In addition to the conflicts with many sectors, such as the creative industries, and traditional publishing, the digital giants are now entering many new markets (energy, education, automotive, health), and engaging on a new form of "enclosure that captures the collective intelligence of internet users as they engage in widespread social cooperation and new forms of democratic organisation"[35]Every activity connected to devices that are linked to the digital platforms become incorporated in the production process and continuously tracked, measured and lastly monetised, mainly through advertising. This new economic model has been defined as "surveillance capital", which according to Zuboff, "challenges democratic norms and departs in key ways from the centuries-long evolution of market capitalism".[36] This describes the latest accumulation logic in the networked sphere, based on intense data extraction, data analysis, continuous monitoring, prediction and the related commodification. The hypothesis is that big data represents the foundation of a new logic of accumulation that can be described as surveillance capitalism. " (.

The open and transparent internet of today is thus growing into a market of citizens' data, (an **identity marketplace**). Behind the big digital brands there are hundreds of lesser known companies building all sort of analytics, trading personalised ads in real time and providing other ancillary services, mainly related to marketing and advertising. Marketing techniques become indistinguishable from surveillance techniques, as their goal is the profiling and targeting of population and the efficient manipulation of consumer demand; attempting to monitor, capture and control the subjectivity of the potential target of their consumer products. **Comprehensive surveillance (political and economic) is the way digital systems operate by default**.

1.2.4 Gouernment Surueillance and the Snowden affair

Governments have also increased the intensity of their surveillance to an unprecedented level. The amounts of information held on citizens continues to grow, and despite the massive amounts of data held by private companies, states still maintain the largest databases of personal information. Much of that data is related to delivering the modern governments' core functions: taxation, regulation and services. But governments are also engaging in a completely unprecedented mass surveillance of internet communications related to the state's core functions: the defence of the realm and the control of its population. The documents leaked by US whistleblower Edward Snowden on the extent of surveillance by the US National Security Agency and its global network of partners has completely changed how informed people see the internet. Mass surveillance has huge implications for digital identities, citizenship and personal autonomy.

Now we know that many governments, such as those in the US and the UK, routinely tap many of the fibre optic cables that compose the backbone of the Internet and collect all the data that passes through these: emails, websites visited and also phone calls as these are routed through the same cables. This data is analysed to look for potentially useful information related to criminals and terrorists, but also politicians, businesses, and human rights organisations such as Amnesty International[37]. Information on everyone else is also processed to discover patterns or other suspicious indicators that can generate new targets. The cable tap infrastructure is used for hacking into thousands of computers and defending from cyber-attacks. This mass surveillance and militarisation of cyberspace, which is perceived as primarily a civilian space by most of its users, has caused widespread consternation, including among technology companies. But the surveillance by the NSA and its growing global coalition of surveillance partners - and almost certainly also China and Russia on the other side - is generally enabled by the data proliferation we described in the previous sections. For a start almost all internet communications are accessed in deals with companies, with varying degrees of compulsion, including the ones that operate the cables themselves.

This symbiosis of corporate systems and government surveillance forms the basis of the infamous PRISM programme, where the NSA and FBI have direct access to data from some of the major tech companies: Google, Apple, Facebook, etc. But even when companies don't cooperate directly, spy agencies can harvest the data we are forced to generate. For example, the NSA has been capturing information transmitted by mobile phone apps, including the advertising marketing profiles used to serve personalised adverts. Some of these profiles include "ethnicity, marital status and sexual orientation"[38].

1.3 Challenges of new technology: Secrecy, Openness and priuacy dilemmas

The current revolution in the creation and use of data has many angles and it would be impossible to cover every single one of them. We will give a basic overview of some of the main aspects of the current digital environment and how they put extra pressure on digital identities.

Big data is the defining buzzword of the times. There is no complete agreement on what constitutes big data but many people use the mnemonic descriptors of the Three Vs: **velocity, volume and variety**. By handling vast amounts of data we can generate completely new insights that cannot be achieved using samples and extrapolation. In their best-selling book on big data Cukier and Mayer-Schönberger[39] explain how big data has generated an epistemological change where knowing why has

given way to discovering apparent correlations and connections among things without necessarily worrying about exactitude and causality. This shift in data practices has raised numerous concerns. A good summary is provided in the US White House report on big data[40]. The report celebrates the positive aspect of big data in improving health, industrial processes and government efficiencies. But it also acknowledges the potential for discrimination, privacy intrusion and a negative unbalance of power between citizens and institutions. One big challenge of big data is that data that was collected for one purpose can end up sliced and diced for use in completely different contexts for other purposes. As we saw in section 1.2.2, this is in complete conflict with established ideas of privacy protection. This means that those who provide the data cannot foresee the consequences. Such lack of transparency makes any ideas of informed consent moot.

Data science - another popular buzzword - relies heavily on statistics and other disciplines to generate useful knowledge out of big data. Data science focuses on discovery and extraction of actionable knowledge from the data[41], and as we saw above explanation and causality are less important than decisions and predictions. One key component of the new data revolution is the developments in **data mining**. This is the name for a variety of **techniques to analyse data to look for patterns, clustering, or possible classifications**. There is also a strong focus on graphs, visualisation and network analysis.[42] Another important development at the heart of data science is **machine learning**. With its links to artificial intelligence, machine learning develops algorithms that enable computers to train themselves to **predict, optimise and classify**[43] data. These sophisticated processes promise to bring immense benefits to humanity but have also raised concerns about potential discrimination and the ethics of predicting behaviour, which we discuss in section 1.5. Another well-known aspect of big data is that it can make it possible to re-identify supposedly anonymised data.

Much of big data is composed of lots of small data generated by individuals, as digital technology becomes ubiquitous and spreads into every aspect of our lives, with all new technological equipment fitted with sensors that constantly generate data feeds. The most important development in computing in this century has probably been **the smartphone**: a powerful and always connected computer full of sensors that we carry with us everywhere. Smartphones are an extension of our cognitive self, allowing us to avoid remembering birthdays, numbers and navigation routes. The concentration of personal information is such a small device allows anyone able to tap into it, whether commercial companies or security services, to gain an intimate picture of our lives. Smartphones also collect and potentially transmit our physical location, which adds a new dimension to any other data collected.

The next wave of technological development promises to connect to the internet most electronic gear in order to exchange all forms of data with users, manufactures and third parties. The **Internet of Things** very soon will have access to a wealth of data from cars and home appliances such as thermostats and fridges. A particularly concerning development is the emergence of wearable technologies and health sensors which can track not just minute movements but also a broad range of physiological information. The Article 29 Working Party[44] has raised concerns about potential inferences derived from such data: "Apparently insignificant data originally collected through a device (e.g. the accelerometer and the gyroscope of a smartphone) can then be used to infer other information with a totally different meaning (e.g. the individual's driving habits). This possibility to derive inferences from such "raw" information must be combined with the classical risks analysed in relation to sensor fusion, a phenomenon which is well-known in computer science."[45] In addition to the above privacy issues there are considerations on who is the primary beneficiary of these sensors' data, the user of the device or the companies that can analyse the data.

The back end side of all those devices connected to the internet is the massive growth of **data centres**, as computing operations are distributed from the relatively low powered end points of the internet to distributed groups of machines sitting in massive refrigerated warehouses. Web technologies that started as a simple way to combine information from different sources into a single webpage have evolved into sophisticated tools that enable this shift towards distributed computing. These technologies have also triggered a revolution in the accessibility of information under the banner of **open data**, which is based on the premise that information should be made available online in machine-readable form and without restrictive licences.[46] What started as a niche movement of web scientists, access to knowledge and freedom of information activists has now become a major phenomenon involving thousands of people and organisations including the World Bank.[47] In 2013 the G8 published an **Open Data Charter**, asking for data to be "open by default".[48]

While most government and companies are still far from following this call, every day there is a lot more information accessible online, from public records to financial information. There are indisputable benefits in opening a lot of data, particularly when it provides unique references required for other services, as is the case with maps or public sector information. Making information publicly available to everyone is one way to avoid creating any large data monopolies. But there are concerns when these principles are extended to personal information, even if this could bring certain benefits, as is the case with health. The US Federal Trade Commission published a scathing report on organisations that gather any available information to produce commercial profiles on citizens, so-called data brokers.[49] We discuss data brokers in detail in section 3. The availability of data sources also contributes to make it **increasingly difficult to anonymise data** as the possibilities for triangulation or so called "mosaic re-identification" grow. In addition, the availability of public data adds to on-going concerns about identity theft and other forms of criminal exploitation.

1.4 Liquid surueillance

The distributed nature of modern digital surveillance has shifted away from the concept of a monolithic surveillance system controlled by a centre, the classic panopticon of Jeremy Bentham that formed the model for much modern surveillance. Haggerty and Ericson introduced instead the idea of the "surveillance assemblage". This new composite system operates by *"*abstracting human bodies from their territorial settings, and separating them into a series of discrete flows. These flows are then reassembled in different locations as discrete and virtual 'data doubles'.*"*[50] Zygmunt Baumann and David Lyon call this new state of affairs **Liquid Surveillance**,[51] and add the important observation of the complete separation of morality from the design and operation of these systems, or *adiaphorisation*. This new surveillance regime can be examined in operation in the use of security profiling to control air travel, borders and detention orders, as documented by Louise Amoore.[52] She found that closed watch lists do exist, but increasingly these systems are based on data mining techniques that have been developed in commercial applications, such as at casinos and in fraud detection. Computers perform complex real-time risk assessments of all airline passengers, where those with a substantial risk score are flagged when they cross the border.

This score is not fixed but will be constantly recalculated. Buying a ticket in cash, or having taken a previous trip to a troublesome country, when combined with other factors could trigger an alert. Each individual element of a security risk alert may be completely lawful and innocent behaviour. This has been evidenced in security deportation orders. In the words of a defence lawyer at one such case at the Special Immigration Appeals Commission: "Neither we nor our clients were given the

'ingredients' of the mosaic – we were only given conclusions, expressed in the form 'we assess that X has been involved in attack planning". This is the way it operates, piecing together fragments which in themselves are innocent."[53] The sophisticated mass surveillance programs of the NSA and GCHQ treat all internet users like international air travellers. But as we saw above the techniques were first developed by the commercial sector and they rely on access to the masses of personal information held by private companies.

The distributed nature of digital identity has some important consequences for how we understand the governance of data systems. Our data trail is completely mixed up with other people's data, our friends but also people who share the same classification categories. Our identity is not simply about ourselves as individuals. This view of identity as a socially constructed and disembodied composite instead of a whole tied to the person is not new. For example, anthropologists have long understood people primarily as social beings distributed across a web of practical relationships, such as labour and gift exchanges; and embodied in material objects full of social meaning, e.g. a warrior's weapons.[54] These insights are being carried out into the digital world by researchers and academics, yet they are slow to permeate into the regulatory realm of privacy and identity, which remains mainly focused on individual persons. But the latest wave of technological developments we discuss in the next section have brought renewed calls by people such as Antonio Casilli to **treat privacy as a social issue, not just an individual right**.[55]

Breaking up identities into disparate data spread on distributed databases can lead to alienation and loss of control over information, but it is also seen by some as potentially liberating. This fragmented aspect of digital identities, called "unbundling", has been discussed for a long time as presenting an opportunity for people to control the information attributes they release in different contexts.[56] Attempts to control those distributed datasets that make up the composite ***data doubles*** has been the main focus of the battles over identity. For example, the campaign **Europe vs Facebook**[57] aims to make the social network more accountable. But the information Facebook holds on us is put to use by other companies, mainly to sell us advert and services, but increasingly as part of job screenings and other more serious profiling. There have been some attempts to control the direct effects of the uses of these data analytics. For example privacy organisation EPIC has a campaign calling for **Algorithmic Transparency**[58]. But the battle over the assembly of identity components, the analysis and creation of meaning that takes place in the background is still in its infancy.

1.5 Freedom, automation and algorithmic regulation

As discussed in the previous sections the advent of big data and data mining raise some new issues. Computer algorithms play an increasingly important role in our daily life. They filter our communications with family and friends, determine what properties we see in online searches for housing, give us driving directions, and increasingly determine critical decisions about our employment, education, health and financial wellbeing. But most people do not understand how they work and how they influence their lives. Many of these algorithms are so complex that they cannot be interpreted simply by reading them, and not even the author can fully predict what results an algorithm will produce without experimenting with some data. Part of the allure of computer decisions is that they are supposed to be inherently fair and free of human bias, but this is now being decried by a growing number of critical computer experts such as Moritz Hardt: "a learning algorithm is designed to pick up statistical patterns in training data. If the training data reflect existing social biases against a minority, the algorithm is likely to incorporate these biases."[59] Hardt goes on to explain that even if the data is not biased then minorities will always get different results. If

nothing else because there is less data available about minorities, so "our models about minorities generally tend to be worse than those about the general population"[60] As explained by Barocas and Selbst, it can be hard to identify these biases and explain them to a court in anti-discrimination cases.[61] In addition, they believe that data mining could also support intentional discrimination by masking intentional exploitation, for example through "purposefully bias in the collection of data to ensure that mining suggests rules that are less favourable to members of protected classes".[62]

As discussed in section 1.3 one of the main concerns with big data is its alleged capacity to predict human behaviour. But hundreds of years of debates over free will, predestination and the predictability of humans have not settled this issue. In addition to the ethics of taking pre-emptive action there are some problems of implementation. As we explained in the previous sections big data is concerned with messy general connections, not with the detail of individual cases,[63] In any case these concerns are not theoretical. There is a drive to move from a criminal justice system that struggles with overcrowded prisons to a system that seeks to use historical data and algorithms to prevent crime from happening altogether, thus turning citizens into potential suspects to fight crimes before they happen. This is the hope behind "predictive policing" – a technique that is already widely adopted in America and is spreading across Europe as well. New Zealand security firm Wynyard has developed "predictive" software that allegedly can suggest when serious criminals will strike again. UK police forces are considering its implementation, according to the Sunday Times newspaper.[64] This technology is used by the police and government agencies in New Zealand and works by analysing emails, text messages and social media files to alert of abnormal behaviour. Predicting behaviour is also an important issue in political processes. Political parties increasingly use sophisticated methods to predict who their voters are in order to focus their efforts on those more susceptible. The centrality of data in these processes has led to the modern political party to be described as a database.[65]

Once we believe that we can predict behaviour the obvious next step is to try and change it. The combination of data science and behavioural psychology is a growing field[66] but the main driver of data driven behavioural modification appears to come from within ourselves. The availability of sensors now enables large numbers of people to engage in constant self-tracking to monitor their habits and health. This behaviour has been promoted by smartphone manufactures and normalised in relation to physical activity - e.g. counting daily steps - or tracking sports performance. There is an organised vocal movement of people around the banner of the **quantified self**, a movement that promises "self knowledge through numbers"[67]. Whilst apparently harmless, these behaviours have raised concerns about the normalisation of self-surveillance. The argument put forward by critics such as Evgeny Morozov is that people who monitor and hare their personal behaviour and conditions make it harder to preserve the privacy of those who don't want to follow suit[68] by delegitimising these positions.

In addition Barocas and Niseembaum explain that if enough people from a certain category or group disclose their information, big data systems may be able to generate insights about the whole group, including those who didn't provide their data. They have described this as the *dictatorship of the minority*[69]. Most people engaged in self tracking put their data in commercial platforms, and have little control over what happens with that data. Some quantified self-enthusiasts try to find ways to download their data form the platforms to carry out their own analytics.[70] Companies such as Fitbit are already working with employers in the US to assess employees' health with a view to lower the insurance premiums companies have to pay for their workforce.[71] Self-tracking is part of a wider trend towards what law professor Frank Pasquale calls **"the algorithmic self**", where we engage in

strategic self-promotion to game the algorithms, but without a full understanding of the rules if the game.[72]

Morozov and other authors are developing also a more political critique to algorithmic governance, stating that the Silicon Valley ideology of "technological solutionism" [73] that embraces of the outsourcing of problem-solving to technologists is very much in line with the neoliberal project. The growing appeal of **data-intensive techniques** allows policy-makers to pretend that problems are being solved. Furthermore, instead of tackling the actual structural causes behind problems like unemployment, inequality, or poor health, governments prefer to remind citizens that most of these problems are the result of their own irrationality and undisciplined behaviour.

There's a growing interest in using real-time surveillance to shift governments to a pre-emptive mode of governance – what Tim O'Reilly refers to as "**algorithmic regulation**". This could be observed in various aspects of daily life, as well as in practices of government where increasingly the emphasis is on aiming to anticipate events in order to either prevent them from occurring, or indeed try to encourage specific events to occur or specific collective behaviours. This involves the pre-emptive forms of intervention we discussed above, whether this is in the forms of anticipating consumer behaviour (information consumerism), risk analysis, or predictive policing, which would allow to avoid problems before they happen. Thus, for example, we would witness a shift from "healthcare" to "health", so rather than heal us when we become sick, health services are likely to give us "healthy living" advices, together with practices of audit, inspection, and review to enhance accountability and value for money across a variety of public services. These programmes inspired by the work of Thaler and Sustain[74] are nudging techniques that encourage the citizens to be self-fulfilling and self-developing as if it were a social obligation. One critique of these programs is that there is a withdrawal of the state and public institutions from fields which under the old welfare model were collective, rather than individual responsibilities. The problem with "algorithmic regulation" is that in eliminating spaces for friction and conflict, it also risks to block the numerous channels – from civil disobedience to less obvious kinds of refusal – through which the system that we are seeking to optimise could be reoriented towards different values[75].

2. Socio-Economic Framework: Reality vs Myth in the Digital Economy

2.1. Setting the context

Since the end of the 20th century, the continuing proliferation of information and communication technologies and their progressive incorporation into globally networked socio-technical infrastructures has led to the emergence of the so called *digital economy*. According to a document released by the *Organization for Economic Co-operation and Development* (OECD) in 2012, the digital economy presents "an umbrella term used to describe markets that focus on digital technologies. These typically involve the trade of information goods or services through electronic commerce. [The digital economy] operates on a layered basis, with separate segments or data transportation and applications."[76] According to the OECD report, "the digital sector has been a key driver of economic growth in recent years"[77]. This finding coincides with a report the McKinsey Global Institute released in 2011, which found that the internet (economy) had contributed to about "21% of GDP growth in the last five years within mature countries" and had been able to create 2.1 jobs for each one that has been lost.[78]

When looking at the world's most successful companies, the importance of the digital sector within the current global economy becomes undeniable. According to *Fortune,* companies like *Apple*, *Google, Amazon* and *Microsoft* are among the 500 most important firms worldwide.[79] Moreover, the consultancy *Price Waterhouse Coopers* placed *Apple* as the first company in terms of market capitalisation between 2009 and 2015, with *Google* ranking second, *Microsoft* fifth and *Facebook* seventeenth[80]. Consequently, the growing importance of the digital economy has contributed to a new mode of economic development that is differentiated from the economic paradigm of the Industrial Age. In contrast to industrialism which is "oriented toward economic growth, that is toward maximising output, informationalism[81] is oriented towards technological development, that is, toward the accumulation of knowledge and towards higher levels of complexity in information processing".[82] As the digital economy takes place online and operates within a network of globally interconnected information and communication technologies, "the level of connectivity between actors and ideas is increasing dramatically. […] What is really new in [this] new economy is the proliferation of the use of the Internet, a new level and form of connectivity among multiple heterogeneous ideas and actors, giving rise to a vast new range of combinations."[83] Digital markets are enabled by new information technologies and applications; at the same time, the products they generate are themselves often new technological products, applications or software. Therefore, "digital markets are characterised by high rates of investment and innovation, which lead to rapid technological progress in the sector". [84]

This digital economy has led to the transformation of many economical niches and the emergence of new business models based on ICT applications and big data. Often, these new models are a response to changes in consumer behaviour, enabled by the internet and other communication technologies. For instance, when it comes to e-commerce, new business models have to be explored and conceptualised; in contrast to traditional commerce, e-commerce "operates under totally different principles and work rules in the digital economy. A general rule in e-commerce is that there is no simple prescription and almost no such thing as an established business or revenue model for

companies even within the same industry."[85] Moreover, digitalisation has for example significantly altered the music and film industry through the emergence of services and applications like those provided by iTunes, Netflix and Spotify. These businesses models do not deliver "traditional" material products like CDs, VCRs and DVDs, but operate on a business model that works through monthly subscriptions and individual streams of virtual content. Within the digital economy, businesses can create economic surplus through a variety of different business models, offering different kinds of products and services such as network infrastructures, mobile devices, cloud storage (i.eg. Google, Amazon, Dropbox), and online services. They can employ information – encoded within digital data – as a resource to increase the added value of their operations, through for example data analytics and comprehensive market analysis, profiling and targeted advertising.

Another big part of the digital sector is the provision of the **material ICT infrastructure**, which allows the transmission, storing and processing of the physical signals of digital data streams within the global network. Infrastructural components include undersea cable networks, Internet Exchange Points, (cloud) servers, data storage facilities, etc. At the moment, companies and Tier I Internet Service Providers (ISPs) like AT&T, Verizon, KPN, British Telecom and others,[86] but also Google and Amazon, are making great profits by providing data infrastructures as well as **over the top applications and services**.

The app market is another market niche that is exclusive to the digital economy, as it delivers (often) uni-functional, integratable software products for information technologies. These software products can then be downloaded to mobile phones and tablet devices, where they provide a great range of useful services for our technologically mediated everyday interactions, such as instant messaging, picture editing, navigating and booking accommodation. The app market is hence closely related to another business model the digital economy's internet environment enables: the model of *cybermediaries* which do not themselves sell products, but provide services and digital platforms for connecting customers and suppliers, demand and supply, and offer services that base on networking and connecting people and technologies, enabling them to communicate and exchange.[87]

Finally, **Digital Platforms** like the transportation network company Uber, the lodging site Airbnb and the online market place Ebay are built on a similar business model. What is interesting to note is that their business model does not require any material assets or capital in the sense of ownership: Uber does not own any cars, Ebay does not own any (material) goods, and Airbnb does not own any apartments. The economic model of these businesses is based on the idea that information technologies can be employed more efficiently and on a much bigger scale to connect people and bring together those who have an interest in exchanging goods and services. They do not themselves create any new products, but believe that the products are *already* there and only need to be *connected* to those who are in demand of them.

While this categorisation is not exhaustive, it points to the new niches that are shaping the digital economy, its services and business models.

2.2 Data as the oil of the 21st century

In this world where new practices and rewired models coexist, it is also common for companies to have dual business models. Companies like *Microsoft*, *Amazon* and *Apple*, for instance, provide both material products and proprietary information technology and online services to users; at the same time, they collect comprehensive data on the way people use these services and hence extensively engage in the big data economy discussed further below. This characterises the **two-sided**

character of many digital markets:[88] on the one hand they provide services and products to customers online, and on the other hand they harvest data from their customers' use of these products and services, which they can then in turn sell to online advertisers.

Scholars have investigated **industry platforms as technological building blocks** (that can be technologies, products, or services) that act as a foundation on top of which an array of firms, organised in a set of interdependent companies develop a set of inter-related products, technologies and services.[89] There is a relevant economic and innovation research stream on **multi-sided platforms**[90] that is very useful to explain the rise and proliferation of digital platforms, in which user mobilisation is a priority to maximise their profits based on the exploitation of the network effect created by the huge user base and the production and sharing of social data. For instance, *Facebook* encourages a critical mass of adoption, while monetising the installed base through advertising. The secret for profitability and growth is thus to activate the Facebook "social graph" by keeping linkages among members active and facilitating and imposing engagement and interaction on the platform. In addition, *Facebook* has designed a marketplace for ecosystem innovation based on applications built by a community of 600.000 developers. Many applications and widgets build on the *Facebook* platform are inherently social in nature, because they lead to building, activating and refreshing the social graph, by enhancing **network effects** and attracting new members to the platform.[91] In the digital economy, a significant factor for the success of online service providers, cybermediaries and new social media and exchange platforms is precisely this *network effect.*[92] The network effect describes a dynamic in which the value of an online service or platform increases with the number of its users, who consequently attract new users, and hence exponentially increase the market success of the service or platform. The network effect can be advanced by several factors. Social media platforms for example become more attractive to use for people if they already have a large number of users, as their purpose is to connect people and let them interact and communicate. Another major contributor are online rating systems which give users of a service the possibility to "like", rate or recommend a service; other people in turn rely heavily on such rating systems and consequently will be inclined to use services with a high number of (positive) evaluations. The great importance that **rating and recommendation systems** have in the digital economy has itself led to the emergence of a new online business model, commonly known as "**click farms**". Click farms generate fake social media accounts which can then sell online "likes", "clicks" and "followers" per thousands to the best bidder.[93] As some platforms like Facebook and Twitter can easily detect computer-generated algorithms and consequently delete the fake accounts, click farms employ human workers with wages as low as 120 dollars per year, most of them based in Asia.[94]

As data storage capacities increase together with the number of online-integrated everyday services and their users, data volumes rise exponentially. And the amount of digital data sets can be expected to continue to grow in the future and to incorporate ever more types of information; already in 2011 it was estimated that the world wide data volume doubles every two years.[95] Looking at the estimated sizes of the data held by major internet players can help us to recognise the extent of the data sets they have at their disposal: *Google* has about 10-15.000 petabytes (pb) of stored data[96]. According to Sullivan (2012)[97], *Google* has seen 30 trillion URLs, crawls over 20 billion of those a day, and answers 100 billion search queries a month. *Facebook* has about 300 pb[98] and *Ebay* about 90 pb[99]. To make the numbers more imaginable, 1 petabyte equals 1000000000000000 byte (=10^{15} byte).

Revenue models, however, are not yet established. A service that is offered for free at the beginning as a way to get attention and market share can seriously jeopardize its success when changing its business model later on; trying to avoid this problem, *Facebook* promises in its homepage "*It's free and always will be"*, offering most of their services to end users free of charge. Companies like *Google*,

Amazon, *Facebook*, and many other internet companies and service providers base a big part of their business model on gathering and maximising the utility of the personal data they collect on their customers. As a result, **a significant part of their stock value is based on the expectation that their data gathering abilities will translate into monetary value in the near future.**

2.2.1. Addicted to data

The increased value of data encourages "digitalised" and "datafied" payments through e-services, mobile phone payments systems and contactless credit cards, for instance. This will become more pervasive with the rise of the Internet of Things, where objects will have IP addresses and the flow of data and identity attributes will be widespread across people's daily activities. While physical currency is not traceable, virtual and digital money, credit cards, and cellphone payments can keep track of numerous metadata linked to the buyers. When paying through these means, consumers do not only transfer money, they also provide added value for the companies with their data. When it comes to digital payments and other services where authentication and information security provide a crucial aspect, e-identification or eID emerge as key issues in order to avoid identity theft and fraud. This business, which is concerned with online or electronically verifiable identity, constitutes a different kind of *Identity Market*,[100] geared towards the conversion of personal identity data to the access of services, online and offline. A great part of this identity market is concerned with anti-fraud and identity management services. According to a European Union report, a functioning e-identity system is important for the security of many (online) services and to ensure access to public and private services, through for example e-tickets. Consequently, the referenced report goes as far as stating that actually, personal identity data is "becoming an enabler of the digital economy". However, they acknowledge "there is a well-known tension between the collection of personal identity data in business-to-consumer transactions and the respect for users' privacy." [101]

Hence, we are seeing an increasing mediation of everyday activities by internet-related applications and technologies within the Digital Economy and the Internet of Things paradigm. The great potential for networking and connecting different individualised digital data sets that this offers has led to a downright **explosion of data volumes as well as data storage and processing capacities**, as for example many "freely" downloadable phone apps already do,. This development in turn invites the harvesting of comprehensive and exhaustive information on internet users and promotes the total transparency of individuals' lives. The creation of vast amounts of digital data boosts the value of data turning daily activities into a stream of bits that feed the "quantified self", tracking who we are, what we say, where we go and what makes us who we are. Within the digital economy, the quantified selves appear as data doubles, digital data sets that incorporate the digital traces of all internet-related activities. In this context, "metadata appears to have become a regular currency for citizens to pay for their communication services and security"[102].

Identification and the correlation of online data to individuals also opens up totally new possibilities for market research, business intelligence and targeted advertising, which is a flourishing business (see section 3). However, the data-driven economy is still more of a promise than a reality. According to a survey conducted by the International Institute for Analytics[103] in 2014, while companies recognise the important role of informational inputs, advanced analytics (the "extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and add value") are still a challenge for them. 71% of companies surveyed indicated their company is actively using, or has near-term plans to use, even the simplest forms of analytics in everyday decision-making, but only 1 in 5 companies using advanced analytics

reports actually use high volume or high velocity data. Most firms seem to have their hands full with their own internal, "small" data. Nonetheless, in that year, two thirds of mid-market organisations invested over 100,000 dollars on analytics programs, this is greatest in the financial services sector where the technologies are more mature, and 98% percent of the companies surveyed consider analytics of strategic importance to their organisation. Only one-third, however, expected to gain a competitive advantage in the future due to successful data mining.

In this context, the notion that **"If an online service it's free… you're the product being sold"** has also been gaining momentum, and concern over people's privacy and control of how their data is used is a matter of growing concern.[104] Such a trade-off could possibly be agreed to be a mutual advantage, as customers can use services "for free" and receive targeted advertisements that suit their needs and desires, while companies can learn about market characteristics, user demands and customer profiles while advertisers increase their advertising success. However, on a fair market basis this is only fully justifiable if awareness of the exchange of data for services is guaranteed and the relationship between customers and companies works as a consented trade agreement. Based on these principles, companies like Datacoup who have decided to directly buy users' data against money instead of services, acknowledging the economic potential of data as goods or assets.[105] Even though exact numbers are difficult to determine, targeted advertising has proven to increase sales conversion, which describes the relationship between advertising and product sales.[106] The business strategy of such companies relies on "individually catered advertisements based upon the content of the website, location of the user, browsing history, demographics, the user profile, or any other available information."[107] The presupposition of **targeted advertising** is that "using information about online behaviour, including sites visited and interest in particular types of content, **behavioural targeting** seeks to serve advertisements that particular groups of consumers are more likely to find interesting. If advertising better matches consumer interests, consumers are more likely to respond to the message, and advertisers will be willing to pay more for ads delivered to such an audience."[108] The economic promise of the potential of data collection and the creation of vast consumer databases leads businesses and companies, as well as other organszations, to create, acquire or access sophisticated systems to gather and dynamise data flows which turn into vast information reserves concerning millions of individuals.

At the same time when they provide a valuable source for understanding consumer behaviour and global markets can successfully perform targeted advertising, the databases which the digital economy, and more generally our increasing use of internet-related technologies in all situations of life, offers great potential for population surveillance and control. The exploitation of commercial databases collected by participants of the digital economy by governmental surveillance agencies builds **the surveillance-industrial complex**.[109] The incentive for governments, companies and individuals to have access to more and more information about their environment transforms personal data into a desired new resource and an economic promise, a "new oil" of the 21st Century. In contrast to natural resources however, the global network presents a virtual and infinite source of information and contents. Consequently, scarcity of resources – on which traditional economies and business models rely – is replaced with over-supply, which has been defined by Rifkin the "**zero marginal cost society**"[110]. This problem generates what has been referred as "**attention economics**" – given a situation of information overload, attention is a scarce resource and actors have to compete for it.[111]

In this advertising-fuelled business environment, where attention is a scarce resource and paying services find a world of apparent "free offerings" (in reality financed by the personal data marketplace) they have to compete with, online services have to refine their strategies and find

solutions to make their investments profitable. One of the most widespread options is to make use of the websites and applications as customizable advertising and data-mining platforms. Mayer-Schoenberger and Cukier[112] define this "**datafication**" as the transformation of social action into online quantified data, enhancing real-time tracking and predictive analysis in the hope that this will boost profits and create new sources of revenue. In order to monetise these comprehensive complex databases, companies and data brokers try to obtain quality information that is relevant, timely, usable and complete. These resources are used by organisations to cut costs, reduce risk, achieve compliance, and improve business performance. But data can also be monetised through its commodification, being directly sold as a raw product. The collection of this vast amount of information is the basis of what has been termed "Big data", which is data characterized by the "3 Vs": Volume, Velocity and Variety[113]. This trend has a reciprocal nature: "Datafication", i.e., the digital tracking of an increasing number of individual facts raises the importance of data as a source of value; since it becomes a richer resource (it increases its quality due to more than ever complete, comprehensive, accurate and updated available data). On the other hand, the increased attention to data mining and big data analysis promotes the interest of organisations to broaden their informational scope, the growth of data brokers, and thus, an ever-increasing interest in data. In this context, R. Clarke[114] introduces the term "**dataveillance**", the "systematic use of personal data systems in the investigation or monitoring of the actions or communications of one or more persons" and remarks their potential to affect the individual's behaviour, "by means of four classes of actions: 'recorded observation'; 'identification and tracking'; 'analytical intervention'; and 'behavioural manipulation'"[115].

2.2.2 Big Data algorithms as refineries

If Big Data is to become the oil of the 21st Century, then data analysis programs and Big Data algorithms are its refineries. Data algorithms and digital data analytics offer the possibility to sift through huge amounts of data, and to find the information needed for specific purposes and specific (consumer) profiles, and finally to create the correlations between different data and data sets which make Big Data so appealing. Algorithms and computer-based data analytics are especially important within the evolving big data paradigm, as data algorithms allow the use of big data sets in a way that would not be feasible by a human analyser, and hence to discover new patterns and correlations on a large scale and through the surveillance of whole populations. Whereas digital Big Data sets contain a lot of *information* about people and their online activities, it is data analytics and data algorithms that can turn this *information* into *knowledge*.

Note that information and knowledge should not be confused, as information is not equivalent to knowledge but builds its basis. Knowledge is the purpose-oriented interpretation of information, which can provide the basis to act upon the information gathered. In times of the internet, the digitalisation of everything and the constant expansion of digital markets, comprehensive information is there in abundance and available to everyone, and so the economic advantage depends on accuracy and interpretation. On the ability to generate intelligence on the basis of the data gathered, and not just data overload[116] or *infoxication*. Consequently, in the digital economy, we can observe an economic transition from competitive advantage based on information to an advantage based on knowledge creation. Algorithms, as well as visualisations, can contribute to this transition.

Algorithms for example play a big role in contemporary stock market practices, especially when it comes to a new, algorithm-enabled form of trading called **High Frequency Trading** (HFT), where algorithms sift through data in order to make economic decisions and, based

on these decisions, buy and sell stocks and derivates within milliseconds.[117] When it comes to data collection on consumers' and citizens' online behaviour, the comprehensive data sets that governments and companies collect can also only be purposefully searched, ordered and employed with the help of data algorithms. Within the digital economy, a big part of big data analytics is used for creating profiles about people's psychology, preferences and (online) behaviour, which can then in turn be used to create comprehensive consumer profiles and exercise targeted online advertisements with great success and revenue. **Hence, if big data is the new oil and algorithms are its refineries, then, consumer profiling and targeted advertising are the trucks which this new oil fuels**.

2.2.3 The Big Data Bubble

The idea of big data as the new oil of the 21st Century is not uncontroversial. A big part of big data's economic potential relies on the belief that there will be more and better "cars" – that is profitable applications of big data – in the future, for which we are mining the fuel – big data collection – today. Today, big data is "in", promising a new and vast field for future business revenues. Consultancies such as McKinsey and Deloitte are eager to reassert the value of (big) data in their reports.[118] However, the belief in the intrinsic value and future potential of big data, which itself fuels the big data industry that builds on comprehensive (personal) data mining could very well present the next economic bubble set to burst, the ***big data bubble***. This hypothetical bubble is thinkable, as there are quite a few open questions with regard to big data. For example, one question is whether big data presents a significant advantage over small data - **can one actually make better, more accurate predictions by constantly expanding the data sets?**[119] Is information alone the key to good decision-making? These questions point to the distinction to be made between correlation and causation. Where causation is an explanatory framework that builds on a causal chain, correlation only explains a statistical relation between two events. For economic purposes, causation is a very useful tool, since it can be reliably employed to predict future developments and influence consumer behaviour. Building a business model on coincidental correlation however is a risky business, as the anticipated effect might not actually come about.

The value of big data and the potential of comprehensive data collection can be questioned when we take a look at how big internet companies make their money. For *Facebook* and *Google*, the overwhelming part of the revenue comes from advertising; in the case of *Google*, advertising is responsible for about 95% of the business' revenue[120], while for *Facebook* it is about 82%, with another big part coming from online gaming.[121] In the case of *Facebook*, which works through "like-systems" and the recommendation of *Facebook* pages and suitable products, it is conceivable that by improving data algorithms, a better understanding of consumer behaviour through collecting big data sets on individuals might lead to better advertising and consequently higher sales conversion. However, a great extent of *Google's* advertising is currently simply based on the search terms entered into the search engine. For this type of targeted advertisement, it is not really clear how creating comprehensive personal data doubles presents a major improvement or contributes to greater profits. Therefore, it is interesting to note that companies like *Google* have been making some attempts to enter into the hardware industry, where the internet giant *Apple* already makes most of its revenue (88%). Microsoft, on the other hand, focuses its business model on selling software products and licenses,[122] not identities.

Therefore, aside from targeted advertising, there is not much evidence to support the case that data and identities are or can be the oil of the 21st Century. And it is indeed questionable whether an

expansion of big data sets and an improvement of big data algorithms will lead to a major increase in revenues from targeted advertising, above the already existing numbers. Hence, the main incentive for collecting ever more user data and creating increasingly expanding identity databases on behalf of internet giants is not totally clear. The bubble may burst. As some authors have already pointed out, it may appear as if big internet giants are collecting and correlating comprehensive personal data just because they *can*, as data storage has become quite inexpensive in the last decades. Holding such comprehensive data sets can then make the businesses appear as major players in a future, hypothetical market, thus increasing their stock value, but not their actual revenue. **This value and investment can only be maintained as long as the big data bubble persists and the belief in the value and future potential of big data is widely shared**.

2.3 Data as currency

Thinking in terms of metaphors can be useful to think through the various dimensions of the digital economy and the role of digital data sets within it. The idea of big data as the oil of the 21st Century points to the economic cycle of the digital market. Further, it can elucidate significant changes that come with the economic transformation that the information society induces. The industrial age relies on the conversion of resources and kinetic energy into the production of material goods, which in turn then employ more resources, whereas the digital economy relies, in part, on the conversion of personal data into digital services. These, in turn, produce more comprehensive data sets that can be used to create further services. According to a report by Deloitte, in order to understand data as a currency, "we must rethink our conception of currencies. **Currency is how we create and exchange economic value across geography and through time**. It is anything that can serve as a medium of exchange, something that can be "cashed out" for goods and services, or used to pay debt or to store value for future use. Data has each of these essential characteristics."[123]

Whereas "data as oil" points to the idea that data, as the basis of knowledge, has an intrinsic value and can itself fuel the production of new (future) goods and services, "data as currency" moves data to the realm of the symbolic, where it becomes a unit of account and an agreed-upon standard of exchange. Just as a 100 Euro bill, as a small piece of paper, does not carry its intrinsic value in the medium itself, but attains its value by its acceptance as a universal monetary standard and the guarantee of its value by governments independent of its "real" or intrinsic value. Similarly, data as a currency works as long as parties agree upon its value equally and hence can trade data as a means of exchange. The trade with data, as exercised by data mining and data analytics companies, then becomes the trade with an asset that builds on future expectations, just as brokers' trade with stocks and derivates relies on making a bet on the value of an asset in the future. This trade – and the digital economy that connects to it – is held up as long as the data bubble is intact and enough parties believe in the value of data so as to use it as a means of exchange between them. Hence, **the industry that trades data relies on the belief in the "myth of big data", namely that there will be future money to make through the exploitation of data sets for specific economic purposes**. If, however, in the future it turns out that, or rather it becomes the dominant belief that, big data cannot offer more than a few surpluses made by targeted advertisement, economic parties may lose trust in the fulfilment of big data's future promise –and in its currency. Consequently, the exchange rate of the data currency will drop, losing its attractiveness for investments.

Companies already use data as a form of currency. For example traffic app *Waze* expanded into Latin American swapping data generated by its customers while using the service in exchange for high quality maps.[124] But the idea of data as currency has more profound implications for digital identities and society at large, as well covered in the work of John Clippinger and the ID3 on identity and the future of money.[125] In the same direction, in his book *Identity Is The New Money,*[126] digital money expert David Birch proposes that digital identities can lead to a cashless society. For Birch the main reason we need money is to give us enough trust to trade. In the past we could have achieve this with face to face personal trust or letters of credit, and now we start to use non-monetary mechanisms such as cards and mobile payment systems. New digital identity systems can build the trust that until recently required a national bank. This is why digital platforms and digital retailers are now becoming the model for the evolution of payment systems and banking. Lately, *Amazon* has announced their intention to enter into retail banking launching their own lending algorithm. As emphasised by Catherine Bessant, head of technology for *Bank of America* "*Amazon* was conceived around the use of data and the customer experience. Banks have not grown up that way."[127] At least not yet.

Thinking of data in terms of a financial system is revealing, as it seems that where data has replaced traditional money in the digital economy, we can find elements of the economic systems that preceded the development of paper money and the modern financial system.[128] Such economies were for example gift economies, in which products were given away for free by trusting in reciprocal social relationships that would work to the benefit of all. The idea that companies like *Facebook* offer their services to us "for free", gaining financial profit only from advertisements, can mediate the impression that they would engage in such a gift, or "collaborative", economy. By offering these services for free, they gain customers who, in turn, provide the company with advertising revenue. This could be seen as a barter economy. However, barter economy presupposes a "coincidence of wants", which means both sides want something the other has, and then, in order to exchange those wanted goods successfully, there needs to be equality of value, which means that the goods exchanged have matching socio-economic values.[129] As the economic value of big data is yet very unclear, and might change significantly in the future, it is difficult to present the trade with personal data as a fair and transparent barter. It is impossible to evaluate what exactly it is that we own (if we own our data at all) and hence how well we can trade it. This uncertainty about the value of data, in the present and in the future, is the main weakness of the data-currencies that depend upon socio-political, but in this case also technological, developments even more so than established currencies that are backed by central banks.

If personal data and identities are more of an informational resource for the creation of knowledge than a good then parallelisms can be drawn between data and early monetary coins which were made out of silver, copper and gold, and whose value directly corresponded to their value as a resource or raw material (hence, one "pound"). In the emergence of material money, using gold and silver as means of exchange was based on the promise that these raw materials held intrinsic value, which would be sustained. Similarly, today the collection and storage of huge data bases is often built on the assumption that the data will have a sustained value and provide a profitable source for different kinds of knowledge, of whose exact application we might not even be aware of yet, and hence be able to generate economic surplus in the future, specially to early adopters (investors). However, the intrinsic value of the collected data is still an open issue –if the data bubble bursts, the gift economy might have presented the right account all along.

Where it is not directly employed for targeted advertisement, but collected as an asset or investment, data as a currency is based on a promise of future value and so **it resembles more the**

emergence of the debt system[130] than a currency, where debt certificates replaced monetary means of exchange that held intrinsic value such as gold, silver and even cattle, holding a promise of future compensation. Interestingly, in many accounts the debt system is said to have contributed more to the emergence of our contemporary financial system and the development of paper money than the monetary coin system.[131] The emergence of large-scale financial systems and symbolic means of exchange, such as our contemporary money, also led to a government interest in regulating the means of exchange and trade.[132] Here, the parallel with data, also collected by governments that, at the same time, attempt to regulate the field, is worth noting.

To consider data as a currency means to move away from the intrinsic value of data into the realm of symbolic representation, where it functions as a standard for exchange and trade. In the history of money, the transformation from the exchange of cattle, gold and silver as a standard means to paper money and debt bills, moved money from being a unit of exchange into being a unit of account.[133] Such a monetary system does not require trust in the intrinsic value of the means of exchange, but in the stability of the system and the continuing function of the unit of account within it. This means that as long as the data bubble is intact, data can function as a currency even if its intrinsic value is unclear. When enough stakeholders buy into the data promise and accept it as a valid means of exchange, data becomes a currency that can be used for trade. Here it can fulfil the features of money as a "means of exchange, method of payment, standard of value (and store of wealth, and unit of account)".[134] As we move from the material realm to the virtual world of digital data, and as financial systems are increasingly digitalised, data could eventually become a currency.

The promise of the data economy, however, does not run unchallenged, even in its own terms. The potential of (big) data and the role of data within the digital economy have been based on the assumption of identity and correlation; in order to use digital data and people's online traces and activities for consumer profiling, targeted advertisement, data needs to be identifiable and correlatable. This means that collected online data must be matched to single individuals so that they can be profiled. Due to the need for personalisation and correlation, privacy concerns have become one of the biggest ethical challenges of the digital economy, and also the reason behind the development of alternative digital markets that function according to a completely different paradigm, based on anonymity and free trade.

Here, Bitcoin as an alternative data currency is the prime example. As a *cryptocurrency,*[135] Bitcoin is a "digital, decentralized, partially anonymous currency, not backed by any government or other legal entity, and not redeemable for gold or other commodity. It relies on peer-to-peer networking and cryptography to maintain its integrity."[136] Interestingly, the principle of Bitcoin is indeed based on data as a currency. In fact, Bitcoin radicalised the approach that data holds intrinsic value in terms of processing power: bit coins are 'mined' through pure computer capacity, where standardised computer algorithms solve complex mathematical equations. For every equation solved, a Bitcoin is mined.[137] What is unique to Bitcoin is the absence of governmental control, as the currency operates through a globally distributed, anonymised computer system. The anonymous payments Bitcoin enable, as an approach radically opposed to the e-ID and dataveillance approach of mainstream digital economy, open up the possibility of alternative digital markets that allow goods to be traded globally and outside governmental control –the most famous example here being the anonymous trading platform Silk Road, whose founder recently got sentenced to two life terms in prison for founding and running the website.[138] However, there also alternative approaches as the one explored in the D-CENT project, that will use cryptographic blockchain technologies, on the model of Bitcoin for decentralised data commons, community value exchange and management of trust[139].

Next to anonymous digital currencies like Bitcoin, alternative and anonymous internet platforms are enabled by privacy enhancing technologies and software such as encrypted e-mail, the Tor browser or VPN tunnels. Next to providing ways around internet censorship and protection from political persecution based on information collected through digital surveillance, these technologies offer people the chance to hide their online traces and hence to avoid, to a certain extent, their involuntary participation in the digital economy's personal data market. Many of these privacy-enhancing tools are provided by developers using free and open source software (see Section 7). As is the case with non-anonymous forms of data as currency, however, cryptocurrencies will at some point have to face the challenge of fulfilling a social function comparable to that of current systems that allow citizens to pay taxes and take part in large-scale redistribution mechanisms. Some of these issues are being addressed by the D-CENT project in its digital currency pilots.

2.4 Data as property

One way to understand how the benefits of data should be distributed would be to look at data in terms of property. One reason for the interest in this approach is that in many big data projects, personal identifiers are removed from the datasets. If it is deemed that the data cannot be related to an identifiable person, this would remove many of the legal protections personal data currently enjoys. This is counterintuitive to what most people would understand it is "their data", but technically it would cease to be personal information as such. Could property protections provide an alternative form of protection instead?

Discussions about **data as tangible property** first arose in the US in the1980s, in the context of insurance protection and whether data would be subject to "loss or damage", but there was no conclusive agreement despite a string of cases.[140] These arguments resurfaced when the US government shut down the cloud storage company *Megaupload*, which held data from many thousands of users. The government rejected claims that it should help legitimate users recover their "data property" because the terms of service of the company included clauses severely limiting any property rights.[141] In a blow to the idea of data as property, the UK's Court of Appeal has agreed that there is no common law right to keep possession of data belonging to another person until a debt owed by that person is discharged.[142] A similar ruling in New Zealand supports the same view by finding that a computer file is pure information and has no separate protection as a property. Laws against computer crime - hacking, etc. - still apply though.[143]

If not akin to physical property, maybe personal data could be another form of intellectual property? Traditionally a lot of data has not been protected as copyright because it would not fulfil the required criteria of being someone's original creation. As we discuss in section 2.4 the EU provides certain protections for databases right, and according to the European Commission's own review this "comes close to protecting data as property."[144] But, ultimately, in EU law there is no property right for data as such. Even if this was the case, would a property approach help solve the conundrums around the distribution of the benefits of data? Would giving people ownership of their data allow them to get a fair share of the benefits or allow them to stop what they perceive as misuse?

In the -admittedly quite different- context of health data in the US, Barbara Evans has concluded that "creating property rights in data would produce a new scheme of entitlements that is substantively similar to what already exists, thus perpetuating the same frustrations all sides have felt with the existing federal regulations."[145] She bases this on what she terms a mythical view of private property

of absolute control over an asset. But different types of assets can have different forms of ownership involving different forms of control over the asset. Evans uses the example of the owners of land bordering on rivers, which can use the waters but must not interfere with navigation. Another problem identified by Evans is that raw original data in many cases is not in itself a valuable data resource. Creating useful data resources requires investment and effort, and in this sense "owning" data is not enough without the means to make the most of it in the form of analytical capacity. The right over databases in the EU is ostensibly designed to protect the investment of those who build the database, not the rights of individuals. Owners of platforms could well have a claim for joint ownership of the database right, together with those contributing their data.

As we saw above, ownership of data may not be the right approach in all cases if what we want to achieve is control over the access and uses of data. As we saw in section 1.2.2, data protection attempts to provide such control but it just focuses on the legitimacy and fairness of the uses of data. Companies can use my personal data to make money as long as they do it in a way that is not completely inconsistent with the original purpose for which they obtained the data, they cause me no harms or distress, and they are transparent and follow due process. There is nothing in data protection about the fair distribution of economic benefits from the exploitation of data. Control over personal information requires more than relying on basic legal protections. Advocates of a user-centric approach to personal data believe that individuals must manage their own information, collect their own data and get value from that in new markets.[146] People maintain different aspects of their life separate online (personas) for different contexts, such as work and family life. For many companies it is more beneficial to have accurate information for the context of a specific transaction even if they don't know anything else about the customer. In addition, different classes of information (financial, health, government records, social, etc.) require different degrees of control.

This is to be achieved through a combination of legal, commercial, and technological tools. Innovative agreements for the use of personal data can give end users more control above the letter of the law.[147] Complex technical systems are being designed to give third parties just the restricted access to individuals' data required to perform specific services. For example, the idea of Vendor Relationship Management turns on its head the concept of Customer Relationship Management software, which allows businesses to keep track of their customers. Dozens of organisations have been built around these principles, although they remain a tiny minority among millions of online businesses.[148] In section 7 we look in more detail at some of these user-centric tools.

The idea of giving individuals absolute control over their data is very appealing and surely a step in the right direction, but there are some issues. Even the most user friendly systems will require an effort to keep data under control, and fine-tuning access could become cumbersome and an on-going endeavour. For example, a fair amount of people have changed their *Facebook* privacy settings,[149] but as the company constantly changes its defaults, users can't keep up.[150] And most people have data scattered around dozens of online services. Having a central point to hold all the data would make things simpler, but it would require immense trust on the organisation holding the data, and this organisation would in its turn create new points of data vulnerability. We cover these aspects in more detail in the next section. In certain circumstances, giving people control over their data could end up having detrimental effects. For example, since April 2015, the UK's National Health Service gives patients online access to health records[151] but doctors can refuse this access if they believe that the patients may be "forced or misled into providing access to their information" to third parties.[152]

2.5 Data as an asset class

The World Economic Forum (WEF) synthesised much of the current thinking around the value of personal data in their very influential 2011 report on "Personal Data: The Emergence of a New Asset Class".[153] Companies such as Infonomics have developed this theme into concrete methodologies for the evaluation of data as an asset that can be included in the balance sheet of companies.[154]

Presenting personal data as an asset class has some important implications. The report called for the alignment of the interests of organisations using data, regulators and those who provide the data in the first place in a balanced ecosystem around personal data. But despite all the explanations above, it squarely frames the discussion in the worldview of investors. An asset class is a group of financial assets with similar characteristics and regulations[155], such as bonds, equities or real estate. As Evgeny Morozov has pointed out, the fluctuating value of personal data could not only generate speculation and hoarding of "data assets", but also lead end users to anxiously monitor their self-worth.[156] Any asset is vulnerable to "bubbles" of over-valuation.

The WEF aims is to produce a triple win situation where everyone - citizens, businesses and regulators - trust each other and share the benefits, but in practice this may be hard to achieve. One of the recurrent memes in this area, as already mentioned previously is that "data is the new oil". The original quote did not relate to personal information, but was referring to the need to add value to raw data; the same way that crude oil requires refining into products.[157] But in any case, and as numerous critics have pointed out,[158] this is far from reassuring for those whose personal data is refined and data can be perceived as toxic and risky as oil.[159]

2.6 Price discrimination

Price discrimination is a long-established economic practice, defined as "the practice of a firm or group of firms of selling (leasing) at prices disproportionate to the marginal costs of the products sold (leased) or of buying (hiring) at prices disproportionate to the marginal productivities of the factors bought (hired)".[160] Price discrimination is common in cinemas, for instance, that offer regular discounts through coupons or special days to reach the price-sensitive customers, or in the airline industry, where companies adjust the price of seats depending on the demand of certain routes and times. In those instances, companies present a whole set of strategies and it is the customer that decides whether to choose the hassle-free option (no constraints on times, no need to plan, no restrictions, etc.) or the cheaper, constrained alternative.

In the context of the identity market, the hope is that having access to large sets of personal data, companies will be able to assess a client's financial situation and willingness to purchase a specific product and tailor the offer to those circumstances. In this scenario, Big Data would optimise price discrimination by not offering cheap products to affluent customers or not attempting to sell expensive products to those who cannot afford them, and benefit both companies and customers. In a recent study on First Degree Price Discrimination and Big Data, Schiller[161] estimates that a combination of demographic personal data and website browsing history can boost profits by over 12%, with some customers paying as much as twice the amount others do for the same product. However, it is still unclear to what extent price discrimination is an extended practice –a study on e-commerce websites the authors found that just 9 in 16 used some sort of price personalisation.[162]

Notwithstanding its financial impact, price discrimination can have externalities that need to be taken into account. The White House's report on The Economics of Big Data and Differential Pricing stresses how "differential pricing in high-stakes transactions such as employment, insurance or credit provision can raise substantial concerns regarding privacy, data quality and fairness. In these settings, big data may facilitate discriminatory pricing strategies that target consumers based on factors outside their own control, or run afoul of antidiscrimination provisions in existing laws such as the Fair Credit Reporting Act or Civil Rights Act."[163] Moreover, the use of online privacy settings on price tracking and comparison websites could increase if customers realised that their personal data may be putting them in high-end brackets in relation to the pricing of some products, and also ultimately impact on a company or sellers' reputation.

2.7 Trust and the new reputation economy

Trust is a concept that permeates much of the discussion about the current data revolution and digital identities, but it's in danger of becoming devalued, as many organisations take a purely instrumental approach that focuses on getting consumers trust to give them their data.

Notwithstanding the above, the discussion on data as currency in section 2.3 makes clear that digital identities always require some form of trust. In the absence of face-to-face interaction and faced with the limits of traditional word-of-mouth dynamics, the digital economy strives to find alternatives that provide people (clients, users, citizens, prosumers, etc.) with the necessary guarantees to engage in online interactions, whether to submit their data, to take part in sharing economy schemes or to dare to change their habits and dynamics to embrace the possibilities of the online world. The report on digital identities by the World Economic Forum sees trust and interoperability in digital transactions as some of the key enablers of the identity ecosystem.[164] We need to know that the person we are dealing with is who they say they are, and that they are entitled or authorised to engage in the transaction. Traditionally this meant personal contact, but in modern times, identities are built and shared relying either on the institutional assurance from the state through the use of databases and ancillary tools such as cards, or on crowd-driven reputation systems based on people's opinions of services, experiences, or on other people.

In the digital realm, reputation is money, and in many different waysit means money for the online reputation management firms that help people and companies manage their online presence or optimise search engine results related to a particular person or product. Leading companies in this field are *BrandYourself* and *Reputation.com*, the latter claiming 1.6 million customers in over 100 countries. But it also means money in terms of market share. The companies that manage to get the trust of their potential customers will see their client base increase, and in an economy where usually the winner takes it all (or a large part of the sector), being able to make people feel comfortable and secure in an online environment can make the difference between success and failure.

Technology is central to the development of trust in identities. As discussed in D-CENT paper D 4.1,[165] the growth of the internet is connected to a proliferation of incompatible technologies to identify, authenticate and validate users of services. The paper documents in detail the complexities involved in developing interoperability and trust across systems. But, as the paper also explains, many of the issues are not technical in nature, as the interests of different institutions and stakeholders are not completely aligned. Ultimately and despite the innovations in social aspects, organisations are the main gatekeepers of the components of digital identity and it is only natural that those incumbents in

a position of power would like to keep newcomers out. Personal reputation may well take the place of government assurance as the basis for trust. According to David Birch this is inevitable, as a social graph is a better predictor of someone's identity than any institutional scheme.[166] But given that currently these social mapping of relationships take place within corporate networks, this change could simply shift power from states to companies without an increase in personal control over information and identity.

The internet and new fragmented identities have brought new possibilities to build trust, such as the use of social media to validate identities. The peer-to-peer accommodation platform Airbnb is one example that requires a minimum number of social media connections[167] as part of its ID verification process.[168] In platforms such as eBay, a bad reputation as determined by your previous customers can significantly affect the chances an online vendor has on the platform. The old saying "the customer is king" gets a new life online, as one dissatisfied customer can leave a lasting opinion in a crowd-driven reputation platform and drive potential new customers away. There is, however, a twist. In the digital world, customers rate vendors and providers, but providers also rate customers. Contrary to the interfaces where crowdsourced opinions are public and visible to everyone, the rating of customers usually has a life away from the eyes or control of the data subject. Taxi drivers can rank passengers in the new app-based e-hailing services, for instance, and employers can rank former employees without their knowledge in platforms such as LinkedIn.

3. Mapping the Identity Industry

3.1 Identity Industry

Identity management is not a new industry. "Offline" collection of personal data has been carried out for decades to conduct research on consumer patterns and enhance marketing campaigns. Back then companies would use demographic information like zip codes to help marketers find out where to send catalogs, or area codes to figure out which towns to telemarket to. However, as fully analysed in Chapter 1 and 2 of this research, with the development of online-based services and its possibilities in terms of data mining, the sector has entered a new era. The **"datafication" of individuals' lives**, **thoughts and moves** through their mobile phones and portable devices, computers and online activity, self-tracking applications, financial transactions, social media and networks, sensors, and the growing "Internet of Things", where home appliances become transmitters of our daily chores provide a vast volume of detailed records of digitalised (personal) information is now routine. Nowadays, and through the integration of the different bits of data we produce, third parties can assemble a **"data double"** of every connected individual and sell it to the highest bidder. The existence of these data doubles, and the specific form and shape of one's digital identity is often unknown to the data subjects that provided the information in the first place. While privacy policies and cookie notices provide some information on the future lives of digital activities, these are hardly read or understood.[169]

In the Information Society,[170] this personal or re-identifiable information has become a key resource for all kinds of organisations. But the way information flows circulate in the digital era is complex and may vary considerably. While public actors have so far promoted the availability of open data to enhance the measurement and understanding of our societies and environments, and to enhance transparency and accountability, private actors have focused on the value of personal data, promoting the commodification of identities with the hope of developing personalised services that can be charged at high premiums. This scenario may be changing, and the emergence of **Public-Private data partnerships**, specifically in the field of health data is increasingly blurring the lines between the goals and practices of the public and private actors.[171] This has translated in the emergence of an "**identity market**" where personal data emerges as a valuable commodity, and where new actors such as "**data brokers**" have a major role to play.

Data brokers are companies that "collect and aggregate consumer information from a wide range of sources to create detailed profiles of individuals. These companies then sell or share your personal information with others,"[172] "sometimes without consumer permission or input."[173] They are also known as "information resellers", "list brokers", "data vendors", "information brokers", or even "independent information consultants".

These companies emerged at the end of the 90s, when incipient online data brokers appeared in the global marketing scenario. Even though back then their techniques and actual reach was not as good as conventional "offline" mechanism, their potential grew rapidly, parallel to the spread of online networks, and regulation was not able to keep up with technological developments[174]. Companies like DoubleClick (currently belonging to Google) or Engage (now part of Microsoft) increased their massive consumer profiling capacities and therefore, their economic value.[175]

As the Federal Trade Commission's report on Data Brokers shows, data brokers gather data from the public and private bodies that collect it from individuals. This information may be volunteered by

the data subject (with or without explicit consent), but also inferred (a credit scored based on consumption patterns, or future needs) or observed (in the case of browser or location history for instance). However, data brokers don't just sell collected or modelled data. They can exchange it or provide it at no cost (e.g. through advertising or referrals) to a myriad of potential customers in different business sectors, including other data brokers, organisations, government agencies or private persons.[176] The products exchanged in this market also have different levels of data complexity, which may range from simple e-mail lists to comprehensive datasets or personal dossiers.

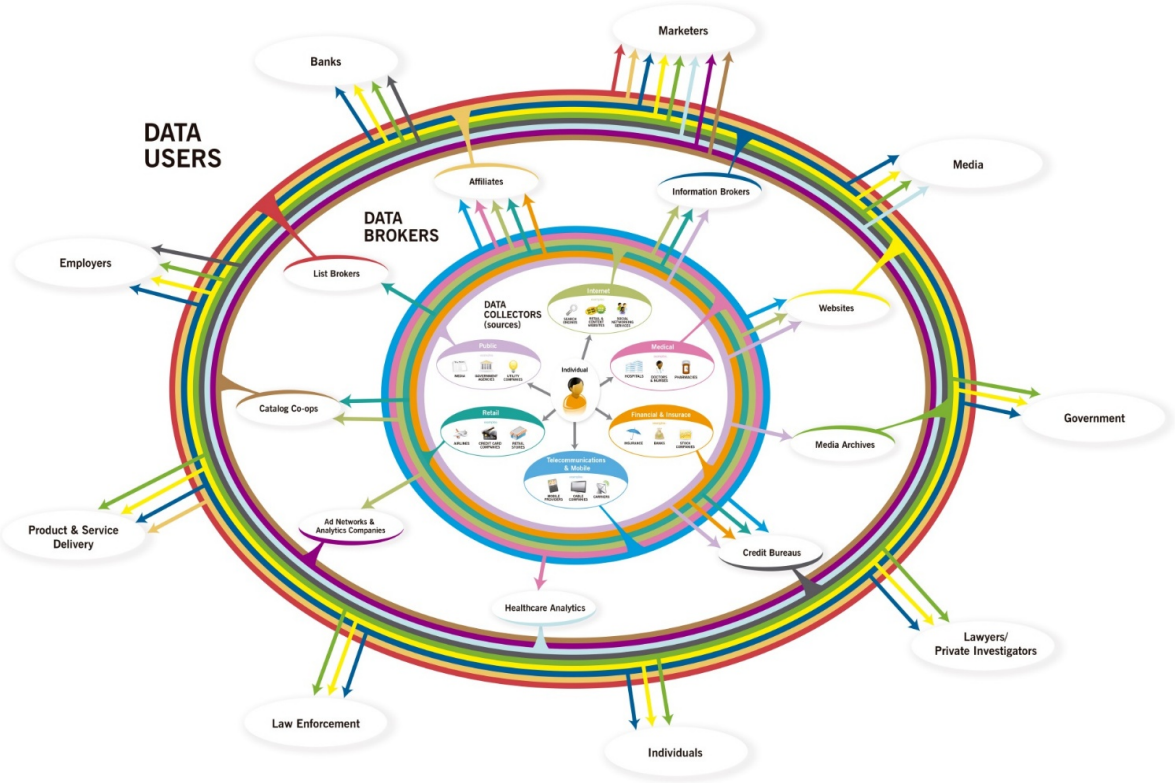

Figure 1. Personal Data Ecosystem. Source: FTC.

Personal data collection always has a comprehensive aspiration, in a quantitative but also in a qualitative sense. The value of databases increases when their volume, accuracy and number of variables grow, as this provides more possibilities for segmentation. Likewise, profiling services are more valuable if they can provide a faithful portrait of each individual –hence the emphasis on "quantified self" products and technologies.[177] Just one US–based company, ID Analytics "holds information on more than 1.4 billion consumer transactions and 700 billion data elements".[178]

In this industry, embracing a vast amount of features increases the added value of the product, as being able to offer all the features possible, even sensitive data, gives a better position in the market. For data brokers, virtually everything recordable and traceable can be commodified, including emotions and future predictions. In this context, traditional offline tracking also has a role to play, as data brokers complement the bits of information gathered from the digital world with other databases developed on the basis of offline activities. These can include property transactions or

traditional shopping, as data collectors digitalise this information into databases that can be used by data brokers, and even linked to digital activities using unique identifiers such as name, ID or address.

Existing research on data brokers has been conducted mainly by two opposite types of contributors. On the one hand, the increasing economic potential of data collection, processing and sharing has attracted the interest of industry stakeholders who have issued white papers and articles on the potential of (personal) data in the field of business intelligence. This includes the set of strategies, techniques, tools and aspects, which are relevant to optimise the management of organisations through business environment knowledge.[179] Business intelligence makes use of all the relevant data available, with the goal of obtaining useful information for business analytics. Information brokerage in the field of business intelligence has gained technological and economic potential due to the rapid development of big data and the ability to process large quantities of complex information in a relative short period of time. Large amounts of data, gathered and analysed in real time, from a multitude of sources hold the promise of extensive monetisation opportunities for unexploited assets. For data brokers, data emerges as a service in itself, and identities are its more compelling asset.

On the other hand, social concerns have led some parties[180] to commission and develop policy documents, reports and initiatives to explore the impact of the identity market and data brokers on privacy and fundamental rights and values. This is due to the fact that data brokerage is a complex field due to its secretive and unaccountable nature. As mentioned before, (personal) data often has a life beyond the control of the data subjects and beyond the terms of the consent agreement, in the cases where there is one. It is therefore virtually impossible to keep track of the data flows, their uses and the roles each actor plays in the data life cycle. Citizens have growing concerns on how their data is being tracked and used while buying, surfing the net or making use of their mobile devices. Media have fuelled the complaints, revealing the reach of marketing-based techniques, raising awareness on privacy violation and the inference capabilities of these data mining companies and business intelligence departments[181]. A widespread illustrative case is that of retail company *Target*, which started sending coupons for baby items to customers according to their pregnancy scores. A man had first notice of his daughter's pregnancy thanks to the company's marketing strategy[182]. Especially controversial practices are Wi-Fi and mobile devices tracking[183], showing acceptability difficulties, in stores as well as in the street[184]. Concerns about governmental tracking abuse are being now also reflected in the area of private consumer and business management.

Privacy concerns from experts, privacy advocates, public office holders, consumers and society in general had been always contested with self-regulation measures. Giant data broker *Acxiom* had to issue a response to the EU's consultation on European Data Protection Legal Framework[185] and has also tried to improve its transparency perception by setting up a website, www.aboutthedata.com, that allows consumers to check which information records are owned by the company. In Europe, the FEDMA (Federation of European Direct and Interactive Marketing) has issued a Charter on Ethical Person Data Management, as part of their strategy to achieve compliance of legal and ethical standards[186]. In the US, data collection practices are now in the scope of regulators, legislators, and the White House itself. In 2012 both the Federal Trade Commission (FTC)[187] and the White House issued insight reports in this field to enhance consumer privacy legislation through key factors like transparency, security, and choice[188].

3.2 Key players in the "Identity Marketplace"

It is not easy to draw an accurate and reliable picture of the scope, structure and connections of the identity industry, not least because of its secrecy.[189] Based on their experience, most people would assume that the main actors in the data industry are *Facebook*, *Google* or *Apple*, as the breath of their services and the scope of its client base makes it apparent that they have access to an unimaginable amount of personal data, produced by their clients while browsing the net, sending e-mails, taking pictures or updating their work calendars.

However, the biggest name in town is *Acxiom*.[190] This global marketing and information management company has databases that include data on 126 million households and 190 million people in the US, and about 500 million active consumers worldwide through 23,000 servers with 1,500 data points. *Acxiom* is headquartered in the US with offices in the UK, France, the Netherlands, Germany, Poland, Portugal, China, Australia and New Zealand and processes data from over 150 countries. *Acxiom* has about 6,000 employees worldwide and a global annual revenue of nearly $1.3 billion. Its current chief product and engineering officer, Phil Mui, developed Google Analytics, and the company also partners with *Facebook* to develop solutions to improve the reliability of our digital personas even when part of our data is incorrect or missing.

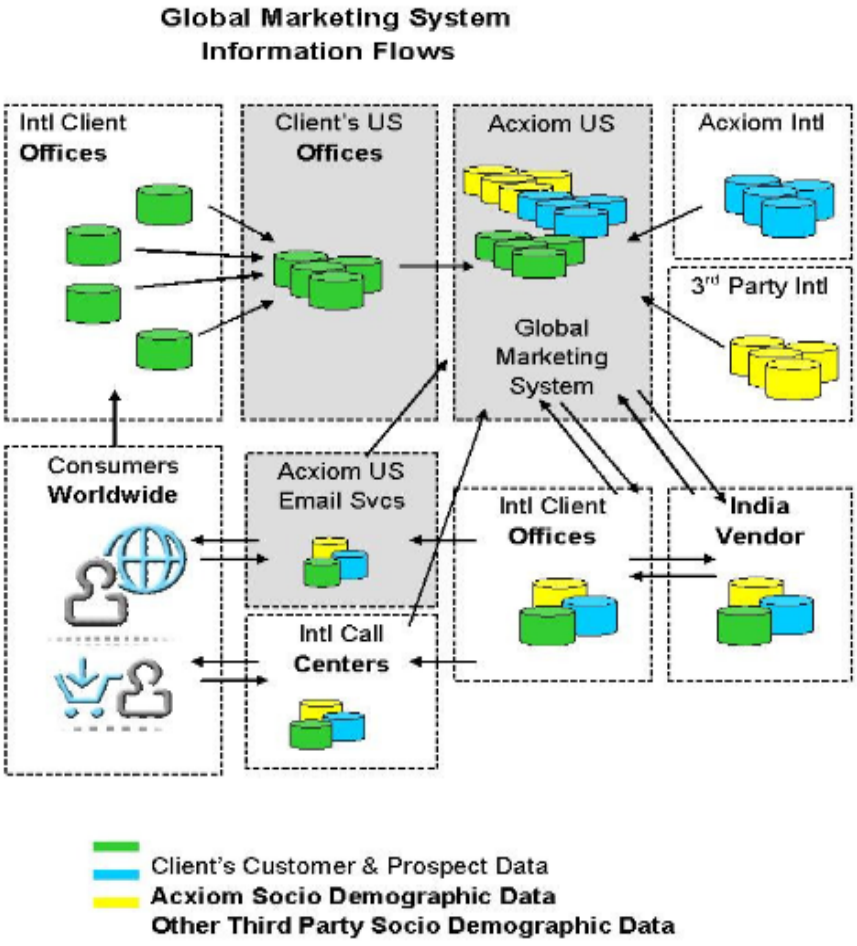

Figure 2. Acxiom's global data flows example. Source: Acxiom.

Household names like *Google* and *Facebook*, therefore, act as massive **data collectors**. The information they gather, however, is only as valuable as the services they offer. By matching their data with data coming from public records, offline activities or alternative digital services, data brokers bring added value and more accurate data doubles to their information-thirsty clients and providers –including *Google* or *Facebook*.

FEDMA, a lobby of the **direct marketing industry**, estimates that the sector of direct and interactive marketing strategies represents an annual expenditure of over 60 billion euros and employs over 2 million people directly within the EU alone. This group has more than 100 company members that use dialog marketing techniques, integrating advertising, public relations and marketing into one strategy.[191] In US, this industry is estimated to produce 300 billion dollars every year. On their part, Privacy Rights Clearinghouse has identified 270 data vendors in the US, but the World Privacy Forum and the Federal Trade Commission have estimated that the industry might reach up to 4000 companies worldwide. This suggests a field where the big players and multi-national corporations such as Acxiom are just the tip of the iceberg, with many small to mid-size companies struggling to become the next big thing or to generate enough revenue through the collection, analysis or reselling of data to justify their existence. It is common to find that data flows from the larger name-brand companies to the smaller companies, who then turn around and resell the data to a third parties of "affiliates", who then market the information themselves or sell it to another downstream affiliate. The term used to describe this process is "affiliate storm", and results in a situation where a consumer at the end of all of the data reselling chain finds it almost impossible to find the original compiler and seller of the data.

Most data brokers engage in multiple online and offline activities and have a range of core expertise, from list brokering to data analytics, including predictive analytics and modelling, scoring, customer relationship management, application programming interfaces, cross channel, mailing preparation, campaigns and database cleansing. This makes the analysis and mapping difficult. Moreover, many of their activities are not disclosed. Some data brokers host their own data and are significant purchasers of original data, such as *Acxiom*. Others, on the other hand, primarily analyse data and come up with scoring and return on Investments proofs. The best example of this second category is another major player –*Datalogix*.[192] *Datalogix*, part of *Oracle Data Cloud*, is a consumer data collection company founded in 2002 that manages loyalty card data, connecting offline purchasing data to digital media to improve audience targeting and measure sales impact. This firm aggregates and provides insights on over 2 trillion US dollars in consumer spending to deliver purchase-based targeting. Over 650 customers (mainly advertisers and digital media publishers) use *Datalogix*, including *Facebook* and *Google*.

A third group of data brokers **sell or resell consumer information online**. This is the case of *Intelius*.[193] Founded in 2003, it specialises in public records information. They offer services to consumers and businesses, including background checks, screening services, people search, customer solutions, public records, criminal check, e-mail lookup and identity theft protection. The company has access to many of the world's most extensive databases and public record repositories, gathering billions of public records annually from a multitude of government and professional entities and assigning them to more than 225 million unique people. *Intelius* services 300 million monthly requests for access to its databases. In addition to retrieving historical and current data, *Intelius* leverages proprietary genomic technology to identify connections between people, places and things.

Overall, there are four core business services that appear repeatedly in the company description of the most well-known data brokers. These are:

- **Identity and fraud services**

Companies like *Experian*, *ID Analytics*, *Equifax* or *Choicepoint* help organisations manage credit risks and prevent fraud. They offer credit monitoring and ID theft products, credit reports, credit scores and credit models, and help clients manage their commercial and financial decisions. Some of these companies also provide risk management and fraud prevention information services, or pre-employment drug screening solutions, shareholder searches, credential verification services, and background checks. Checkpoint, part of the Elsevier group, offers in addition underwriting and claims information services such as motor vehicle reports, claims histories, policy rating and issuance software, property inspections, and audits.

- **Customer relations and care**

Loyalty cards and schemes are both one of the main systems to gather consumer information and part of the core business of many enterprises in the data brokerage environment. Companies such as *Epsilon* and *Bluekai* specialize in helping companies get and retain customers. They provide list marketing data, insights & strategy, marketing technology, creative services and media reach. Epsilon alone, with 7,000 employees and 70 offices worldwide, manages more than 500 million loyalty members and more than 4,000 databases in areas as diverse as the financial sector, retail, consumer packaged goods, insurance, automotive and healthcare. *Bluekai's* services enable companies to personalise online, offline and mobile marketing campaigns with richer and more actionable information about targeted audiences.

- **Predictive analytics**

All Big Data companies argue that data analysis can contribute to predicting the future and thus making better decisions. Only some of them, however, present this as their main focus of expertise. Such is the case of *Corelogic* and *eBureau*. The first provides consumer, financial and property information, analytics and services to business and government and develops predictive decision analytics by combining public, contributory and proprietary data. *eBureau* offers a suite of predictive analytics and real-time big data solutions to consumer-facing businesses, delivering instant insights that help make decisions throughout the customer lifecycle and provide solutions for Business-to-Consumer (B2C) and Business-to-Business (B2B) companies.

- **Marketing and advertising**

Closely linked to customer care, the companies that are specialising in marketing and advertising help their clients find customers and present their products to audiences likely to buy them. *Criteo*, founded in 2005, focuses on digital performance advertising and performance display. They generate millions of high-quality leads through dynamically generated personalised ads, and its success is measured on the basis of post-click performance using a pay-per-click model that includes extensive real-time bidding tools and category and product level optimization in 32 countries across 5 continents. In 2014, it analysed 430 billion US dollar sales transactions, served more than 740 billion ads and reached, according to their website, 1.06 billion unique users globally.

- **Other**

There are many other models in a field that seems to be ever expanding. Some data moves from online to offline and back, some through social media and back. Some companies, as *Intelius* (see above), specialise in public records. Others, such as *PeekYou*, in people searching by analysing content from over sixty social sites, news sources, homepages and blog platforms to identifies the actual people behind it and make sure that using a maiden name or a fake address has no impact on the data-matching to elaborate a profile. *Rapleaf*, on the other hand, finds its area of expertise in e-mail, providing data on 80% of US email addresses and assisting marketers to understand whom their customers are and what channels they can be contacted on. They conduct e-mail intelligence, e-mail validation and e-mail append. A case worth highlighting is that of *Recorded Futures*, a company founded in 2009 that provides organisations with real-time threat intelligence, allowing them to proactively protect themselves against cyber-attacks. With billions of indexed facts, the company's "web intelligence engine" continuously analyses the open web, including news publications, high-calibre blogs, social media platforms, financial databases and government websites to give insight into emerging threats. They offer their products under four categories: cyber threat intelligence, corporate security, competitive intelligence and defence intelligence.

The point here is that the business models and data flows in the data brokerage scene are complex, use many sources, and differ between types of data brokers. Moreover, in the identity market, actors can play different roles. Data collectors may transform the collected input into information and they can make use of it as well, thus refusing the need of intermediary companies. Data experts may also develop analysis software to help organisations achieve a situation of "data self-consumption" where they do not need further intervention of external advisers or providers. In other cases, the complexity of the data managed (volume, number of sources, type of variables and indicators…), the kind of information desired (profiles and market scores, tables, segmented databases, charts…), and the ambition of the results pursued (risk mitigation, consumer loyalty, etc.) my affect the needs of the organisations, and it might be necessary for external data scientists to play a role.

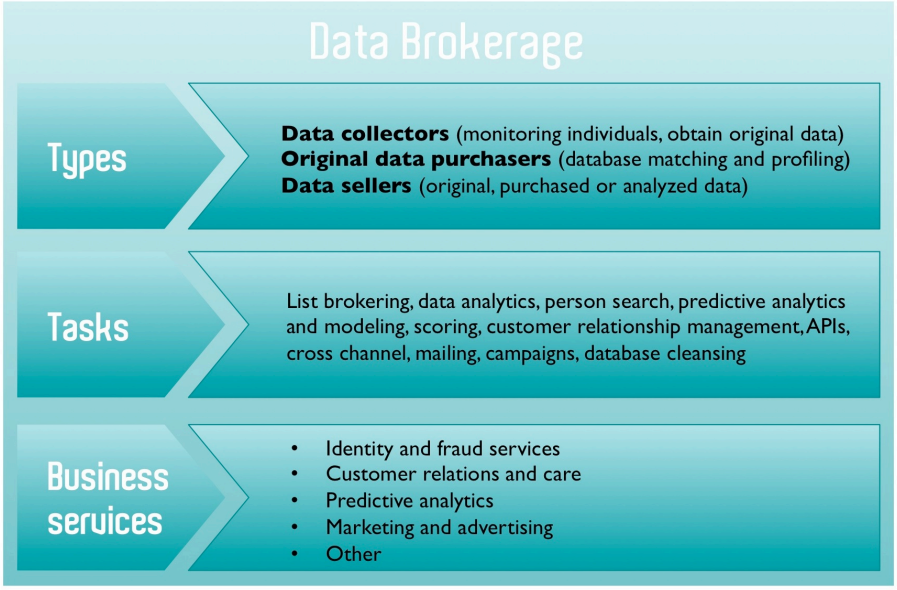

Figure 3. Overview. Source: Own elaboration.

3.3. Value chains and business models

The purpose of collecting, storing, processing and analysing big data is to obtain added value for companies through different acquired or enhanced capabilities.[194] As shown above, that can be offered Business 2 Business (B2B) or Business to Customer (B2C). Data brokers increase the value of their information assets on the basis of two key factors that bind together the wideness of global dispersed information with the particularity of the sought piece of data:

1. **Exhaustiveness and broadness:** data brokers are expected to have "all" the information possible and encompass "all" the populations through comprehensive databases. They are supposed to reach millions of records, to monitor trillions of dollars in sales transactions and to carry out the continuous tracking of a vast amount of human actions.

2. **Segmentation and specificity.** Big data is useful if it is segmentable and easy to exploit. Throughout the vast ocean of data, customers expect to be able to find the specific individual profile or segmented dataset demanded. The larger the number of data points recorded, the higher the value of the information assets.

Technological advances allow to go from exhaustiveness to segmentation, sometimes even in real time. This does not mean that data brokers aspire to embrace all the possible services, but that in that specific area, a specific company is able to provide the most comprehensive scope and reach. For instance, people search services decrease their value if they are not able, on the one hand, to have access to a huge amount of persons, and on the other, to find and obtain the information demanded. When data brokers have comparatively large amounts of personal features quantified and, ideally, regularly updated, as the big players do, they are meaningfully more competitive than others. Nevertheless, it does not make sense to reach such large populations if the data brokers are not able to translate these bits into useful, operative and interpretable information.

Data brokers monetise the information they compile through several ways: selling it to other companies (e.g. other data brokers), organisations, government agencies, or to private persons. But they also might exchange this information under a cooperative arrangement rather than sell it (e.g. *iBehavior*, a data "cooperative" that allows retailers to pool data about customer transactions).[195] Another way to make profit from this data is providing the information at no cost, and making money through advertising or referrals.[196] A 2013 study from the Tata Group estimated that half of firms producing big identity data sets currently sell their digital data, producing an average sale of 22 million US dollars in 2012.[197] Companies like *Acxiom*, *KBM Group*, *Bluekai* and *Datalogix* have been increasingly making use of marketing data for resale.

Data collectors and brokers, online advertising companies, web publishers, and marketers are key actors in the current data brokerage scenario. They create added value from personal data, which is transformed and managed into several products and services that help to outperform the competition by generating wider audiences and reaching more potential customers or reinforcing actual customers' loyalty, improving the efficiency and utilities of internal databases, and increasing sales. Increasing the informational inputs indiscriminately, however, does not increase added value. The quality of monitoring and matching is a crucial factor –putting together credit card payments, geolocalization, and online searches might shed light on the health status of an individual, or about the potential impact of previous advertising exposure.

One aspect that has enlarged the scope of the services and activities of data brokers is mobility. APIs are currently outstripping web browsing, and mobile devices can be equipped with almost 20 sensors monitoring information.[198] The progressive expansion of thse kind of devices in developing countries opens new perspectives and markets, as shown by the interest of many data-related companies to expand their business activities beyond the Western world.

Due to the current lack of governance in the sector, the value distribution issue raises concerns about the equitable impacts of value exchanges and the achievement of a global trusted flow of data.[199] The question of the ownership of the data is another issue that may impact on the future development of the identity market. Even though individuals should control who can access, use, aggregate, edit and share their personal data, this is not always the case in practice.[200] The consumer value exchange is currently limited to better consumption experiences and, when available, certain levels of control over the data (consultation of existing records, right to opt-out, etc.).[201] Nevertheless, some data brokers are considering the possibility of providing a tangible benefit back to consumers, involving them in the value chain as a more active element (*Datacoup*, for instance, already offers this option).[202] Experts, like Paul Jacobs, executive chairman of Qualcomm Incorporated, suggests that data "is not ours like the dollar bills in our pocket" which you can choose to give out based on what you get in return. But this may soon be a growing trend -to give up specific information in exchange for services, products or money.[203]

3.4 The identity industry through the data lifecycle

In a context of changing categories, one way of looking at the identity industry is through the data flows, following the data lifecycle, which includes Data collection and/or access, Data storage and aggregation, Data analysis and sharing, Data exploitation and use, and Data deletion or removal.

3.4.1. Data collection/access

The first stage of the data flow cycle implies harvesting as much data as possible from every identifiable individual. An identifiable individual is not someone from whom a name or identification number is available, but a distinguishable unit, a unique user. For certain purposes, data collectors may be interested in linking datasets to a certain combination of name, surname and date of birth (e.g. for people search engines), but for others, the key value can remain in other variables like the e-mail address, the IP address or the postal address, among others.

The key players in the data collection process include the household names most would associate with the business of identities (*Google*, *Apple*), and also social media providers such as *Facebook*, *Twitter*, *Instagram*, *LinkedIn* and a long etcetera. These companies offer services that usually function as bait to get their customers' information. For most citizens, for instance, *Google* is an e-mail provider (gmail) or a search engine. However, well over 95% of Google's revenue comes from advertising via its AdSense program which places ads on millions of websites. The more information *Google* has on its customers, the better it will tailor the search results to their needs, thus reinforcing advertising, and not end-user services, as the backbone of its business. In the case of *Facebook*, the revenue coming from advertising was 85% in 2014. Again, in this case customers do not perceive this social network as an advertising company, but a company devoted to connecting people. As companies dedicated to optimising the match between their users and their clients, the more information they have on the users, the better their matching services. The size and granularity of

these companies' databases is the main basis of their success and prominence in their respective fields.

Original data brokerage was based in generalised actions, routines and preferences, and focused in consumer patterns -credit card records, retail loyalty programs, registration profiles, etc. Nowadays, it is possible to perform a thorough tracking of each person's life in real time and with accurate positioning details. Due to their individualized use and multiple sensor equipment, mobile devices have made it possible to enhance the detailed collection of personal records linked to an identifiable person (who), with specific information about placement (where) and time records (when). Data collection practices have shifted from "offline"-based forms and surveys to "online" practices based on the subtle caption of human behaviour, with the intermediary step of credit card payment monitoring.

The World Economic Forum has listed three categories of data on the basis of the collection methods. Two relate to primary data and one to secondary data:[204]

- **Volunteered data:** created and explicitly shared by individuals, e.g., social network profiles.
- **Observed data:** captured by recording the actions of individuals, e.g., location data when using cell phones.
- **Inferred data:** data about individuals based on analysis of volunteered or observed information, e.g., credit scores.

Concerns have arisen about online surveillance and the ability of these data collectors to mine personal information and derive intelligence from it. These have put the focus on web surfing habits and use of apps, although data collectors use offline sources as well. Thus, it is important to bear in mind that data collection and brokerage is not exclusively an internet-based activity. "Offline" conventional records (which may range from social security numbers to trade union affiliations) are still important for data brokerage, but online data collection offers a vast horizon of possibilities for real-time individualised tracking. For instance, an online retailer like *Amazon* is able to easily keep track of actual purchases as well as of product searches inside their platform. A "physical" retailer can easily report its daily purchases through widespread technologies like barcodes and systems like customer loyalty programs. But in order to achieve an approximate idea of the attention dedicated to each product, these business have to deploy a more complex set of sensors in their buying area (presence direction sensors, smart cameras, etc.) and systems or tools for their interpretation (e.g. "heat-maps"). Disney, for instance, recently introduced a bracelet with sensors to track its visitors through the park.[205]

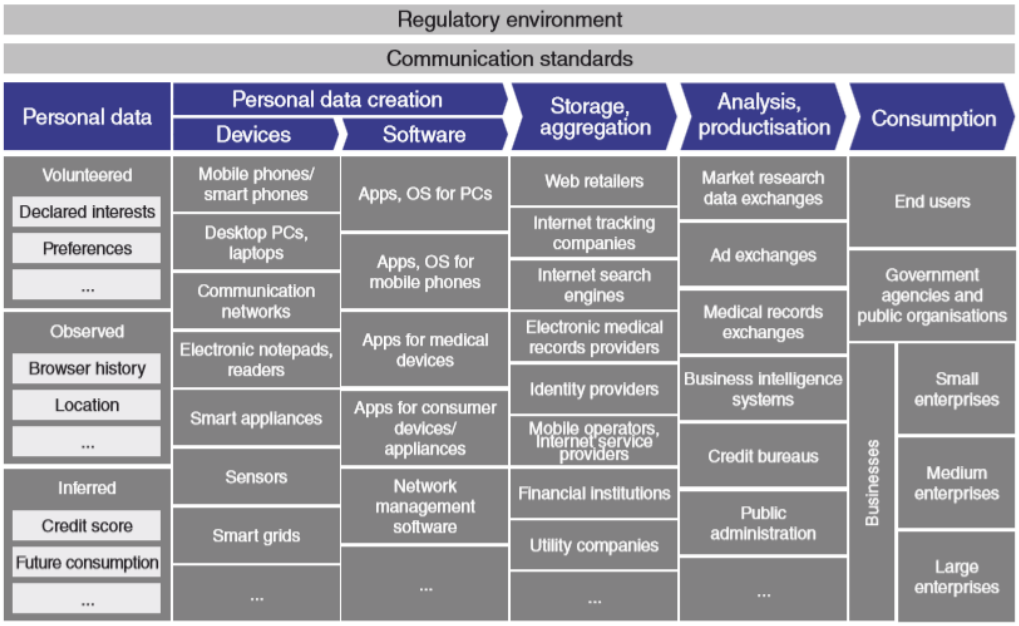

Figure 4. Personal data ecosystem. Source: WEF (2011)

Some examples of the types of data collected online and offline include Personally identifiable information (Name, ID number, date and place of birth, etc.), device identifiable information (IMEI, IP, MAC, etc.), public data (Information retrieved from administration like criminal records, civil records, property ownership, and media and public data), online activity (browsing and web searches, social media activity, time spent in a website, clicks, e-mail content), geographical information (postal address or ZIP code, geolocalisation of mobile devices), transport (Type of private vehicle/s; public transport use tracked using smart cards), leisure activities (sports, cultural activities, musical and movie taste, press reading, etc.), consumption and lifestyle (stores visited, acquired goods and services, average spending, drug use and smoking habits, etc.), financial details and insurance (estimate income, debt reports, credits and mortgages), medical data and health (medications, health-related searches and medical history, etc.), telecommunications and mobile use (Incoming and outgoing calls, texting, mobile data use) and other features such as gender, religion, sexual orientation, ethnicity, marital status and offspring, education level, political leaning, etc.

Some of the tools to collect data (e.g. registration forms) allow for an opt-in or opt-out of (depending on the option marked by default) the legal cession of personal data. However, recent experiences such as that of the National Health Service e-records in the UK have exposed that the guarantees behind opt-out can be difficult to enforce.[206] Additionally, studies show that basic data protection regulations such as access rights are often undermined in practice by the difficulties citizens encounter when trying to exercise them.[207] It is therefore difficult for citizens to access, rectify or remove the information owned by private organisations, even when the data controller can be identified and the relationship between the data subject and the data broker is straightforward and clearly stated. Additionally, many of the sensing systems currently installed in urban areas usually operate without the knowledge of the citizen, who is thus unaware of its rights, the practices of the data brokers or what his or her devices are revealing about their activities.

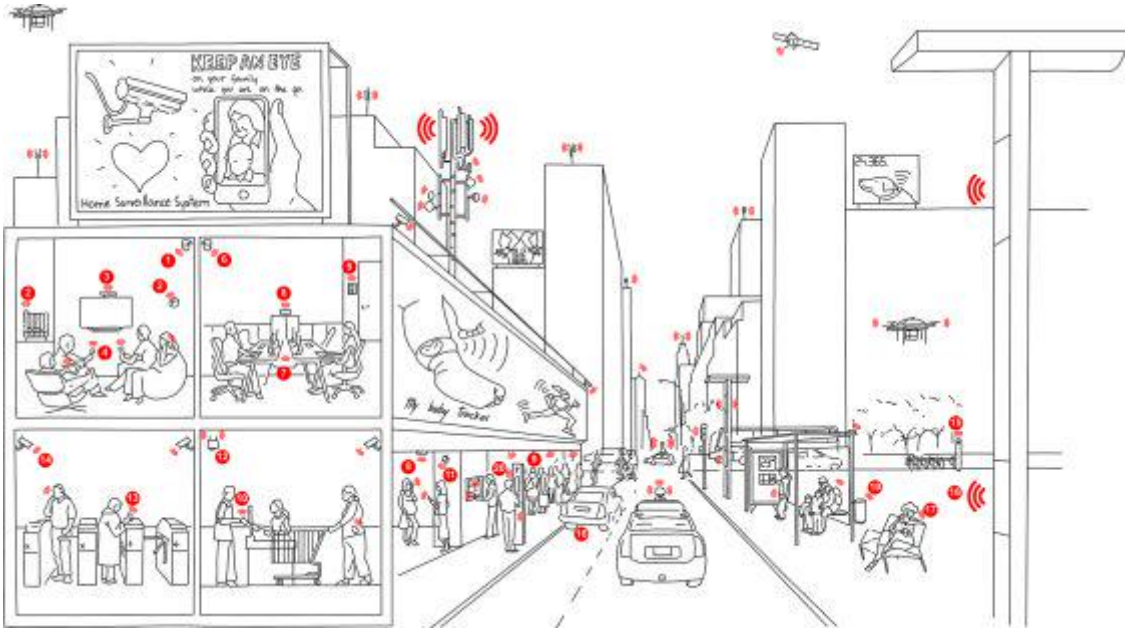

Figure 5. Sensors in urban spaces. Source: Own elaboration.

The specific case of *Google* and *Facebook* is a case in point here, as these large companies promote the integration of several platforms by allowing users to log in third-party applications using their social network profile. ***Facebook Connect*** is a cookie that enables *Facebook* to act as an authentication mechanism as well as liking, commenting and sharing contents outside the social network. According to SimilarTech, this integration feature that allows *Facebook* to keep their cookies working even if a user is not browsing its site is present in more than 7 million websites.[208] Tracking cookies for marketing purposes have caused controversy as they could potentially inform your social media contacts about browsing routines you may want to keep private, and because while most of them last between one or two years in one's computer, some can outlive the device and last up to 10, 100 or even nearly 8000 years.[209]

Future challenges in the field of data collection are linked to the success of APIs in the face of "traditional" web browsing, the proliferation of the semantic web and the development of the "Internet of Things". The Internet of Things is a concept that turns daily objects into tracking devices, contributing to the increasing "datafication" of people's lives and the "digital depiction" of the quantified self. An energy meter may reveal the consumption patterns of a household. This is relatively uninteresting for a person but it can bring valuable information for marketers and analysts. Sports monitoring apps, in their turn, are based on the measurement of health-related data, which, in combination of other users' details is a very valuable information source for data brokers.

These technological trends allow collecting a higher volume of personal information, implement data-analysis and take advantage of the multiple sensors present in the devices. For instance, the semantic web makes use of more than plain text and keywords, and is based in languages specifically designed for data. This means that web resources can be linked to data-revealing elements. Developments in these fields increase not only the ability to collect more and more data, but also the ability to interpret more precisely what kind of data it is, even without the intervention of humans. Moreover, this situation allows machines and interfaces to exchange these data points more easily and effectively, since the pieces of data are better recognised.

3.4.2. Storage and Aggregation

Personal data collection requires large storage capacities, especially if historical datasets are included. Storing large databases has become an increasingly outsourced asset, even transferred to colder climate countries to save on ventilation costs. Maintaining a data warehouse can be an unnecessary cost for a company. However, firms focused in data management may prefer to have their own databases (e.g. *Acxiom* or *eBureau*). It is not surprising that the main cloud computing providers include *Amazon*, *Microsoft*, and *Google*. These companies need to develop large storage and data transfer capacities for their own purposes, and take advantage of the deployed infrastructures to offer these capacities as outsourcing services for third parties.

Cloud machines under test			
	Virtual CPUs or cores	RAM	Cost per hour
Amazon m1.medium	1	3.75GB	12 cents
Amazon c3.large	2	3.75GB	15 cents
Amazon m3.2xlarge	8	30.00GB	90 cents
Google n1-standard1	1	3.75GB	10.4 cents
Google n1-highcpu-2	2	1.80GB	13.1 cents
Google n1-standard-8	8	30.00GB	82.9 cents
Windows Azure Small VM	1	1.75GB	6 cents
Windows Azure Medium VM	2	3.50GB	12 cents
Windows Azure Extra Large VM	8	14.00GB	48 cents

Figure 6. Cloud machines test and cost per hour (Feb. 2014). Source: Infoworld.

Teradata Labs, the global leader in data warehousing hosts several big data storage solutions for big companies. Among their main customers (those with petabyte-level capacities) one can find *Apple*, *Walmart* and *eBay*, as well as other important names like *Verizon*, *AT&T* and *Bank of America*. Other remarkable big data storage solutions are provided by *Quantum*, a company specialising in scale-out storage, archive and data protection that provides services to more than 50,000 customers -from small businesses to multinational enterprises, including *NASA*, *LSI* (which now belongs to *Avago Technologies Company*) and *EMC*, which produces a range of enterprise storage products, including hardware disk arrays and storage management software (its flagship product, the Symmetrix, is the foundation of storage networks in many large data centres).

Concerning In-memory databases for Big Data analytics, according to Markets and Markets this market will enjoy a 43 percent compound annual growth rate (CAGR), leaping from 2.21 billion US dollars in 2013 to 13.23 billion in 2018,[210] led by companies like *Aerospike, Altibase, Couchbase, Datastax, Exasol, IBM DB2, Kognitio, McObject, Memsql, Microsoft SQl, Oracle, Parstream, Pivotal, Quartet FS, Redis, SAP Hana, Teradata, Unicom Systems and VoltDB*. A widespread key resource for database management is an open-source framework and file system called *Hadoop*.[211] An interesting

phenomenon to optimize big data storage are the so-called "data lakes",[212] storage repositories that use a flat architecture and hold a vast amount of raw data in its native format until it is needed. *EMC* is currently developing competitive solutions for data lakes. As data brokers match and merge partial databases to develop more exhaustive profiles and infer new information through weighed indicators so they can offer more data points, storage solutions are a key and necessary player in the identity industry.

3.4.3. Data analysis and sharing

Once the information is adequately compiled, it has to be organised in a way that is useful for the customers that will acquire it or to be analysed. The analysis and sharing of the harvested data is the part of the process that is most invisible to the eyes of the data subject. Here, the names of the key players (*Acxiom*, *Bluekai*, *ID Analytics*, etc.) remain a mystery to most people, as are their business models and data practices, as shown in the previous section. However, it is important to bear in mind that most of the daily personal data compilation processes are not aimed at making a direct financial profit from the harvested information. For example, for Internet Service Providers (ISPs) it is legally compulsory to retain specific metadata from users. Public bodies also manage vast volumes of personal data for administrative and security purposes, not commercial. Private organisations have to deal with employees' or members' files to carry out their routine functions. In these cases, simple datasets with the minimal required information are generated. However, there are many ways to take financial advantage from this kind of information, both directly (selling datasets) and indirectly (optimising services through a better understanding of a problem through the intelligence generated using data).

General datasets and dossiers showing consumption trends and patterns are useful for organisations butut databases with identifiable entries (names, user names, e-mail and post addresses, IP, MAC, etc.) allow feedback actions as well, like directly addressed marketing campaigns. Personal data analysis and sharing enables end-users to make the most of the acquired product. It is not the same to monetise a detailed set of identities describing patterns (more or less raw databases) than monetising the results of processing that information (e.g. showing ready-made reports and charts). Simple datasets can be easily managed by a single organisation and do not require specialised intermediates, but obtaining additional or external sources and managing them to achieve a clear image of the desired target audience may require the intervention of a data broker specialising in marketing and advertising.

Data brokers provide clients with a wide range of products in the field of data analysis, depending on their needs, including collected "raw" marketing databases, segmented datasets, business intelligence tools, customised marketing platforms, geomarketing solutions, customer analytics, one-time campaigns, etc. *Bluekai* even offers a platform to manage your own data, additional data from 700 million profiles, and pre-integrated media partnerships. One of the key challenges data brokers face is precisely the need to offer comprehensive products. This requires that they combine different sources, mixing several datasets and matching databases, ensuring that they can depict a quantified self of thousands or even millions of individuals or households, define segments and thus be able to provide quality, reliable information. A single entry may be the combination of both online and offline data mining, showing both online and offline patterns. This process also allows inferring identities from potentially anonymised or pseudonimysed environments, thus rendering some simple anonymisation techniques useless as unique identifiers like a MAC Address or an IP address can lead to the re-identification of an individual.[213] Due to the continuous distribution, transferring and

exchange of datasets, it is virtually impossible to accurately know who owns files with personal data affecting a specific person, or what combination of existing databases may lead to re-identification.

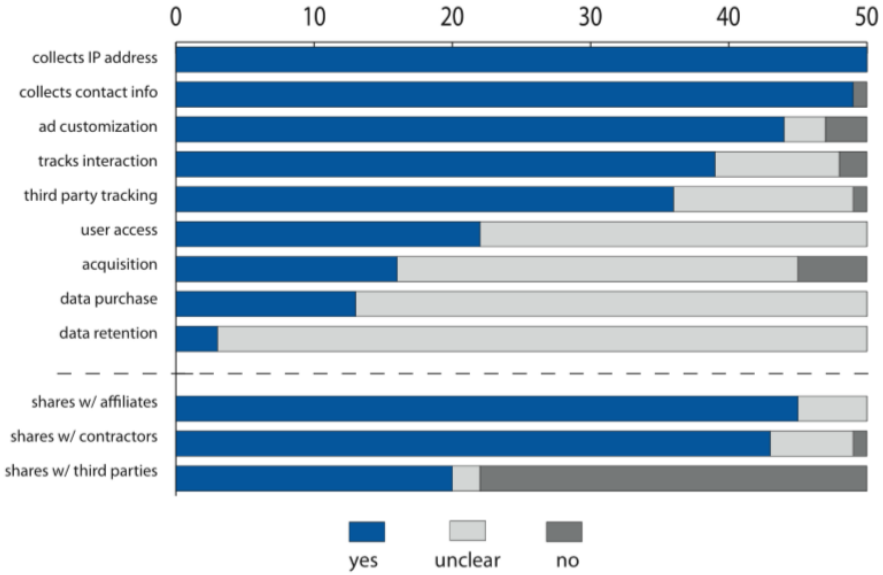

Figure 7. Privacy Policy coding for Top 50 most visited websites in 2009. Source: Knowprivacy.org

Once the data is combined, personal data services create artificial segments (clusters) with common characteristics that are the basis of most of the services offered by data brokers. These clusters classifications follow diverse models and are aimed at depicting potential consumption patterns - "Personix" is the consumer cluster categorixation sold by *Acxiom*, for instance. It includes up to 70 categories like "working & studying" and "affluent households". Other examples of cluster categorisation are "Prizm", sold by *Claritas* or "P$ycle" by *Dataman Group*.[214] In a more simple categorisation, marketing and e-mail lists are offered according to a unique variable, like "people over 65", "people who just acquired a credit card", "green & hybrid vehicle owners" or "cat lovers". These lists are provided by websites like Mailing List[215] finder or List Giant.[216]

Figure 8. List Giant first level of categorization. Source: List Giant.

A new way to make the most of the collected data is to implement machine learning technologies. Machine learning refers to the "construction and study of systems that can learn from data."[217] *Facebook* uses machine learning algorithms to recognise faces in photos, and *Netflix's* leverages them to improve its recommendation engine. This capability allows them to identify past patterns and anticipate events and needs on that basis, to provide a better service. Combining geolocalisation data captured by sensors or transmitted via device applications with individual tastes, companies may offer personalised advertising in real time and according to one's location at any given moment through smartphone applications. The alleged lack of transparency attributed to the development of these algorithms raises concerns on privacy due to the unknown inputs that are used.[218] A key company in the field of machine learning is *Skytree*, which provides deep analytic insights and future trends predictions, allowing them to identify untapped markets and customers. India is currently specialising in this area, but for the development of algorithms freelancers are commonly being hired.[219]

3.4.4. Data exploitation and use

End-users need high-quality data that is clean, processed, and adequate for their goals, but also free from potential lawsuits or customer backlash. The right kind of data analysis may allow companies to cut costs, reduce risk, gain market share or improve business performance. Creating intelligence on the basis of personal information and being able to cater better to different consumer segments can have a direct impact on profit. In the long run, it is hoped that big data will allow companies and governments to predict user/citizen behaviour and thus cater to their needs before they are made explicit.[220] However, it is not easy for data customers to check the quality of the acquired product. They have to rely on providers' ability to obtain accurate data and profiles. Nevertheless, data brokers develop ever more refined systems to improve the accuracy and veracity of the collected information and to provide updated records. In the context of the big data boom, it is possible to find that the leveraging of complex datasets it is not only reserved to large companies. Even single individuals might take advantage of processed databases with personal data (e.g. through person search websites). Also mid-size and small business might have better insights of their actual or potential customers to minimise risks or to explore new markets. *eBureau* targets this category of businesses to sell their products, since large companies' demands are being already covered by other data brokers or by their own data analysts. According to a 2015 survey, nearly all the small businesses that contact software advice are looking for dedicated marketing automation software[221].

Companies might leverage internal data through external software or analytical support. They can also combine and enrich their collected data with external data obtained through data brokers, or they can just obtain external data or data analysis to drive their strategies. According to the FTC,[222] the main clients for processed data include a wide range of categories. Attending at the most common uses of big data for each category, it is possible to establish an approximate usage classification:

- Credit risk assessment, accounts receivable management and ID products (ID theft prevention, fraud prevention, verification and authentication solutions, people search, e-mail validation, etc.) are products acquired by customers belonging to areas like alternative payment, attorneys and investigators, insurance companies, government entities, lenders and financial services, and also individual consumers.

- Big data exploited for business day-to-day management (sales impact measurement, lead management, customers loyalty and retaining, customer relationship management), is used by common goods and services companies in the automotive industry, consumer packaged goods manufacturers, educational institutions, energy and utilities, telecom companies, travel and entertainment services, pharmaceutical firms, real estate services, retail companies and technology companies.
- Big data oriented at marketing long-term strategies (data assessment and data-centric strategies, customer acquisition, Analytic insights, Automate decision making, predictive analytics and market niches detection) are carried out by marketing companies and also other data brokers.
- Audience analytics (advertising impact measurement, advertising impact measurement and audience targeting improvement, digital performance advertising, performance display) is the area where big data is leveraged by marketing and advertising firms, media companies and non-profit entities/ political campaigns.

In an attempt to improve transparency, control and corporate responsibility, *Facebook* stated publicly that it would dedicate special care to the selection process of their third-party partners. For this, they have established guidelines for their partnership policy:[223]

- Inline Transparency. Each *Facebook* advertising display shows the message "About this Ad" that explains the company that was responsible for including the users in the audience. The company also offers a list of the third parties that collaborate with advertising and other efforts, such as measuring the effectiveness of Facebook ads.[224] Some of them are *Atlas, Bloom Digital, DoubleClick, Flashtalking, GroupM, Mediamind, Mediaplex, Pointroll, TruEffect* or *Weborama*.
- Comprehensive Control. *Facebook* offers the choice to avoid a certain ad, or not to be shown ads from that partner. *Facebooks'* partners also agreed to provide on their information page a comprehensive opt-out of future targeting.
- Enhanced Disclosures. *Facebook's'* partners will give details on how they collect and use information, including the kind of information collected and the policies related to the sharing of that information.
- Data Access Tools. Partners are committed to develop tools to assist people to see audience segment information that the partner has associated with them, and to exercise control over that data.

Even though most of the data brokers fulfil their legal requirements (sometimes thanks to the lack of regulation in certain areas), some companies and actors do not observe the legality or just make unethical brokerage. However, it is difficult to accurately identify bad practices due to the complex web of data brokering that blurs the track of data transferring. Many daily actions imply the collection of personal data that can potentially be transferred to third parties. Nevertheless, the Electronic Privacy Information Center has detected and listed 40 websites offering telephone calling records and other confidential information.[225] These companies offer for an approximate price of 100 US dollars all the calls made and initiated from a wireless phone, or toll calls from wireline phones. The World Privacy Forum has denounced the online offering of sensitive information and lists that should not exist, by websites like ExactData Consumerbase, DM Databases or Medbase 200.[226] These include police officers' home addresses, rape victims, domestic violence shelters, genetic disease sufferers, seniors who are currently suffering from dementia, patients with HIV/AIDS,

people with addictive behaviour, alcohol and drugs, people identified by their illnesses and prescriptions taken, Hispanic payday loan responders and derogatory credit consumers.

3.4.5. Data deletion/ removal

Interestingly, data deletion is the part of the data-flow where it is difficult to find industry actors. This points to one of the current trends (and anomalies) in the identity industry –the push to accumulate data at all costs and in all contexts, regardless of its need or specific use. In a process akin to primitive accumulation, data is gathered, kept and analysed in the hope that it will have future, profitable uses. Data deletion, therefore, does not currently exist as a market sector, even though deleting data permanently is no easy task.

There are, however, actors that have been paying attention to this fact. Privacy advocates and consumer rights associations, for instance, push companies to carry out best practices in order to facilitate personal data control and privacy. Even though the ideal data collection, according to these associations should be based in the opt-in model, they have managed to identify which data brokers allow to be deleted from their lists, or for their records to be marked as not usable information. According to a research published in ProPublica in 2014, 85 out of 212 companies allowed a complete opting-out, less than half (92) accepted opt-outs (85 of them, complete opt.outs). And most of them required submitting some form of identification.[227] Although there are no companies that offer individuals to be removed from data brokers files, some groups like World Privacy Forum[228], Stopdatamining[229] and Privacy Rights Clearinghouse[230] have developed and published FAQs and/or their own lists where they indicate how to opt-out from the most known data brokers: The data brokerage global leader *Acxiom* has its own opt-out website, where each individual can check and control which information had been collected by the company. However, the website itself demands personal data inputs and opt-out processes to avoid new data collection derived from this service utilisation. Moreover, the terms of use do not clearly explain the details of data collection that take place during the registration.

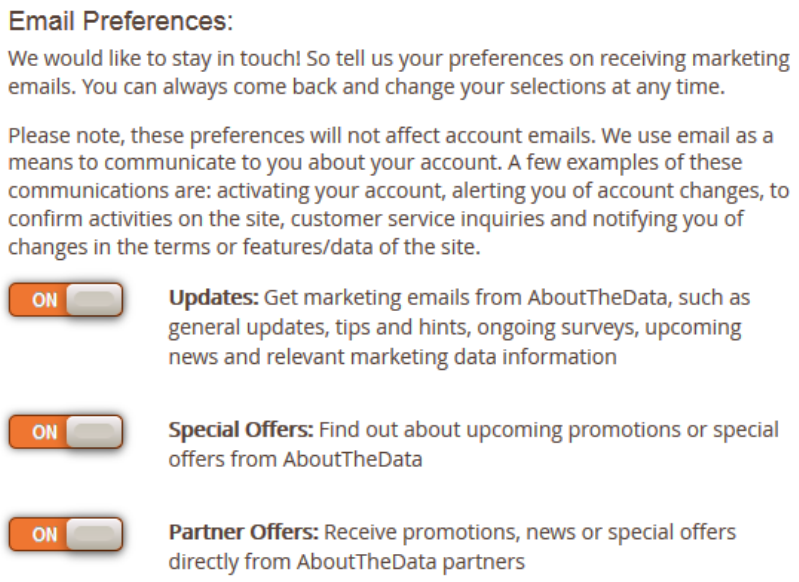

Figure 9. Privacy settings at Acxiom's Aboutthedata.com

Other remarkable initiative is the Do Not Track (DNT) header, a suggested HTTP header field that requests a web application to turn off its tracking or cross-site user tracking features. Currently, all browsers support this function (for Technology strategies see Section 6).

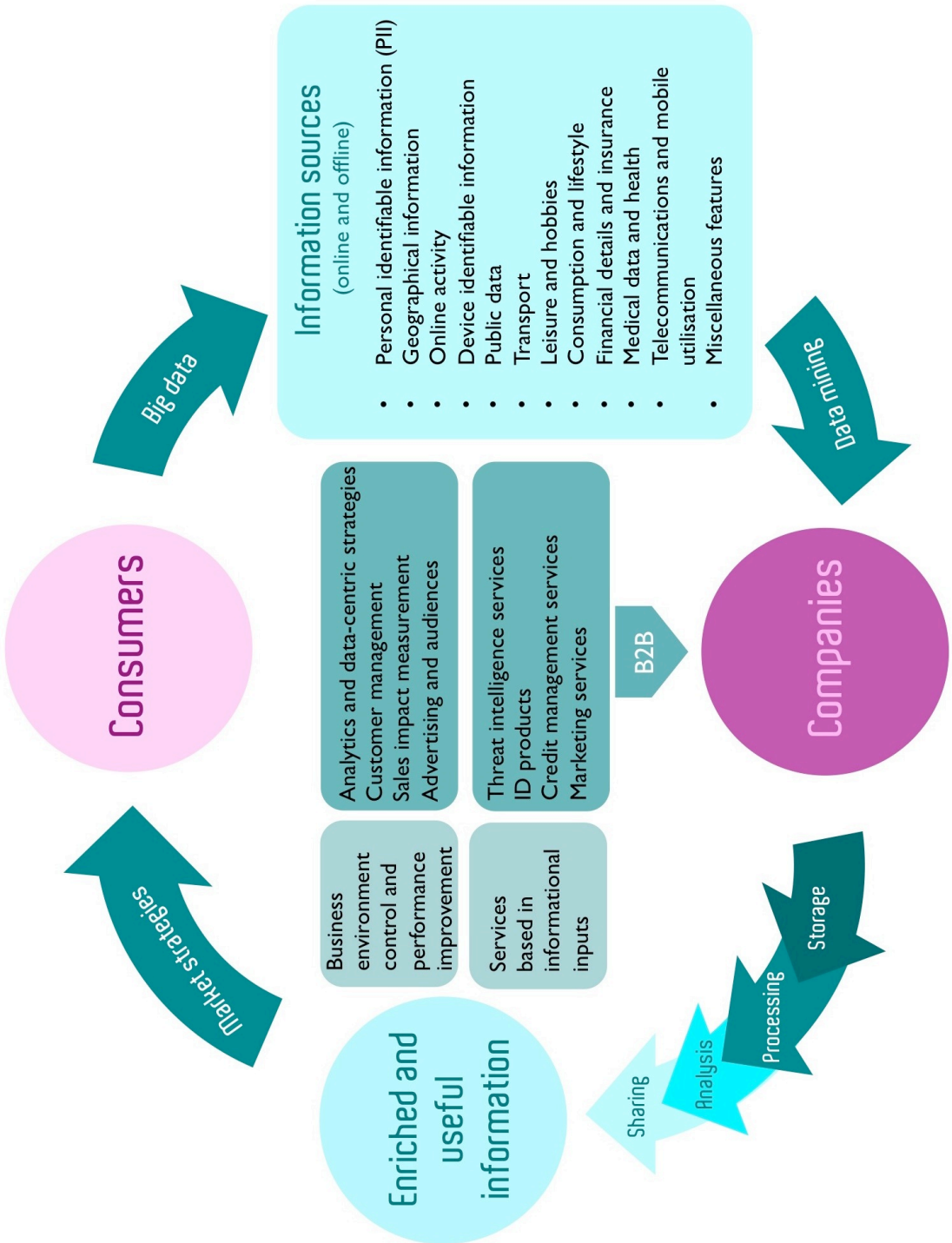

Figure 10. Value chain and data flow. Source: Own elaboration.

4. Empirical Case studies

4.1 The "sharing economy"

The so-called "sharing economy" can be defined as "a set of practices and models that, through technology and community, allow individuals and companies to share access to products, services and experiences".[231] It implies the use of digital platforms to organise and optimise the acquisition and exchange of products and services. In this context, an organisation or company develops a digital platform (website, app, etc.), which is offered to match supply and demand with other relevant users. This model shifts the focus from production to consumption and increases the relevance, initiative and decision power of consumers. Already in the 1980s, Toffler coined the term "prosumer"[232] (a mix of producer and consumer) which acquires a new meaning in the new context where a networked society interconnects users (potential producers, distributors and consumers) in a decentralised way.

The sharing economy eliminates or minimises intermediaries, as products and services are distributed from a collectively managed demand. The model goes beyond customisation and individualised consumption, and users acquire a more active role to become more than just final payers. Exchange overtakes money as the main rationale behind the sharing economy, as shown by the cases of initiatives as diverse as *Wikipedia* (an online encyclopedia built collaboratively and on the basis of donations), *BlaBlaCar* (a ridesharing community), *Couchsurfing* (a hospitality exchange site where guests stay in people's home for free), *Foodsharing* (a platform where people exchange food), *Bookcrossing* (an online book club where people give away books in public places), *Kickstarter* (a crowdsourcing platform), Craigslist (classified ads platform) or *Uber* (a transportation company) or *Airbnb* (a website to rent out lodging). The sharing economy works thanks to globalisation and the internet, allowing people who do not know each other to pool and share resources (from cars to skills), and crowdsourcing logics, delegating to the users the responsibility to take care of organisational, operational or financial resources, among others. Commonly, the economic model is based either in donations from people who chose to contribute to a service that is useful to them, or in charging for the provision of the technical resources where the exchange take place.

The sharing economy is a key player in the identity market as it is often based on a reputation and trust model. As these companies often put in contact people who did not know each other before, mechanisms have had to be built to ensure their identities and establish their reputation as trustable users. This means that those participating in the sharing economy often need to reveal a great deal of personal information –or at least to make it available to the service provider. The personal data shared may include names and addresses, car number plates, telephone numbers or personal traits. While this data may not be made public by the service provider, the users of the sharing economy will most likely need to create an online persona that will expose a history of their activity with a particular service (contributions to Wikipedia or books crossed) and publish other user's opinions on one or one's property (house, car, etc.).

Other than donations and charges per exchange, the actors in this field may also chose to track or analysing through machine learning algorithms different aspects linked to the users' digital identities and sell the data to data brokers as a way to get revenue to maintain the networks, storage, systems, software, etc. Data from private and public profiles could also be shared with third parties interested in analysing the behavioural patterns of the individuals that interact in a certain platform. It is precisely the use of rating systems and the accumulation of personal data that has led the Federal

Trade Commission to examine internet businesses that facilitate peer-to-peer transactions for transport, accommodation and ecommerce, with a report due later in 2015.[233]

There are a wide variety of areas covered by the sharing economy, from transport to cooking, including accommodation, knowledge, clothing and even money lending.[234] There are also different models and approaches, to the point that a categorisation is nearly impossible. In order to introduce some logic, some chose to speak of the "collaborative economy", describing those activities that are enabled by internet technologies, connect distributed networks of people and/or assets, make use of the idling capacity of tangible and intangible assets, encourage meaningful interactions and trust, and embrace openness, inclusivity and the commons.[235] This definition, however, only captures a fraction of what today constitutes the sharing economy.

The possibility to share among equals in a networked world, bypassing costly intermediaries, can conjure dreams of exchange, solidarity, trust and generosity. However, it can also destroy established business models, as happened with peer-2-peer exchange in the music industry, or commodify previously uncommodified resources. In the sharing economy people give away their knowledge on social networks, workers share their daily ride to work and groups that were previously outside the traditional circles of funding can now reach funding parties through crowdfunding platforms (crowdfunding). In the economy of equals, however, every resource becomes an asset that must be mobilised. If a car is parked, its value is not being maximised. If a house is empty, it is inefficient. In this definition of the sharing economy, even free time becomes an asset, and dreams of solidarity and altruism become obscured by value maximisation and profit seeking.

In 2010, Botsman and Rogers[236] proposed a categorisation for the then-emerging phenomenon of collaborative consumption. They distinguished between "Product service systems" where companies offer goods as a service rather than sell them as products; "redistribution markets", where used or pre-owned goods are moved from somewhere they are not needed to somewhere they are, with or without charging money for it; and "Collaborative lifestyles", where people with similar needs or interests band together to share and exchange less-tangible assets such as time, space, skills, and money. This categorisation, however, fails to capture the criticisms and controversies causes by some of the actors in this field, who point at the need to differentiate between the for-profit and the non-for-profit branches of the sharing economy. As some authors have pointed, some actors in the sharing economy emerge as new "middlemen" trying to avoid taxes, regulations, insurance and the obligations related to holding a business activity. These voices denounce how, in the sharing economy, the original gift economy is being subdued by market capitalism.[237]

4.1.1 Uber

Founded in 2009, Uber is a transportation company based in San Francisco (California), Uber has become the poster child for the sharing economy, even if in its recent court cases it has chosen to abandon that label to define itself as a company "providing information services".[238] Uber is available in more than 270 cities and 50 countries, and the company has recently expanded their products, offering premium services (*UberLUX*) or food on-demand (*UberEATS*).

Uber makes use of a mobile app that allows users requesting specific trips or offering a drive in real time to get in touch and pay for the service. It is very similar to a traditional taxi system, and it even uses a metering system calculated on the basis of distance or time, with the difference that Uber drivers do not need special permission to carry out their activities and that all payments are handled

by Uber –not the driver. Since the drivers are not licensed, the system builds trust on the basis of a rating system in which users can rate drivers, and viceversa. When downloading the app that allows the system to function, drivers and users agree to let the company access the individual's identity, contacts, geolocation, SMS, telephone calls, multimedia, camera, WiFi connection and device ID.

Uber was launched after its founders raised almost 50 million US dollars in venture funding. By early 2015, the total number of attracted funds were reaching the phenomenal figure of 3 billion US dollars, with a market capitalisation of 50 billion or higher. It is expected to make 10 billion in revenue by the end of 2015,[239] on the basis of charging drivers up to 30% of the cost of the rides. Contrary to traditional taxi companies, Uber does not have fixed fares nor takes responsibility for insuring the drivers or passenger, and all indirect costs are passed on to the "independent" drivers. Using surge pricing algorithms to match supply and demand, Uber changes its fares every 3 to 5 minutes to maximise profit, taking advantage of adverse weather conditions or states of emergency to increase their fares by up to 300%, as determined by their algorithms.[240]

This has caused controversy, which has been added to their conflictive coexistence with traditional taxi services, subject to laws and regulations often ignored by Uber. Beyond the complaints related to its treatment of drivers, tax evasion and regulatory arm-twisting, and surge pricing, Uber has also attracted attention due to its questionable activities to undermine the competition,[241] and its CEO's threatening of journalists.[242]

However, Uber's data policies and practices have also raised concerns. The ability of drivers and Uber staff to track customer's data and the news that this is a common practice, even for recreational purposes[243] has put Uber under the privacy spotlight, especially when one considers the amount of sensitive information, linked to financial and geolocalisation data that Uber amasses.

On February 2015, the company confirmed a data breach affecting 50,000 of its drivers[244]. A few months before, it was discovered that Uber was reporting data back without users' permission, like malware apps do.[245] *GironSec*, a security systems analysis blog, decompiled the code of the Uber Android app and published the complete list of data being collected by the firm,[246] which includes accounts log, app Activity, app data Uusage, app install, battery, device Info, GPS, MMS, NetData, PhoneCall info, SMS, telephony info, WifiConnection, wifi neighbors, root check and malware Info. Uber has also raised concerns on their corporate attitudes towards privacy since its senior vice president of business, Emil Michael, made some controversial comments suggesting that Uber should consider hiring a team to look into the personal lives of its critics in the media. In order to compensate this unfortunate statement, Uber's spokesman N. Hourdajian said that Uber has clear policies against illegitimate use of personal data (journalists' travel logs in this case): "Access to and use of data is permitted only for legitimate business purposes. These policies apply to all employees. We regularly monitor and audit that access."[247] However, in November 2014 a US senator expressed his worries about the company's "troubling disregard for customer privacy," suggesting that there is evidence that Uber's practices may be "inconsistent" with its revealed policies.[248]

Similarly to other identity key players, Uber gathers and analyses an enormous amount of personal data, and it is unclear what the limits to its use are, whether they may end up in the hands of data brokers or constitute an alternative source of revenue in the near future.

4.2 Consumer financial data broker industry

Credit and debt appeared as key elements in the "democratization" of consumption that broadened the access to goods and services after World War II. In the late 80s, capitalism was assumed to be the "best" system for the allocation of resources, while the post-cold war consumer society was fuelled by all forms of credit products, from simple loans to complex mortgages. As financial institutions had to deal with risky decision-making, they started collecting different pieces of information about credit applicants. In order to prevent abuse, however, in the US measures were taken to limit the capabilities of credit lenders to collect personal data for their financial assessments, namely the Fair Credit Reporting Act (FCRA) in 1970 and the Fair Debt Collection Practices Act (FDCPA) in 1977.

However, already in 1989 several US companies in the field teamed up to create the "FICO score", an assessment based on credit files from *Equifax*, *Experian* and *TransUnion*, in order to optimise the eligibility of credit applicants. The ability to match and compare multiple sources of complex data, however, only emerged with the development of ICTs. With them, the digitalised data flow boosted, and vast amounts of useful data were now available, traceable and, in some cases, linkable to specific identities. Now, "More than ever before, big data allows companies to zero-in on ideal consumers, identified through personal information composites, allowing them to predict future consumerist activities."[249]

Nowadays, financial data brokers provide credit-scoring services based in the analysis of complex datasets. This information is provided to the potential customers to increase their performance in the areas of risk minimisation and fraud prevention through the exploitation of scoring and rating indicators. These indicators are not only composed by tax fraud or defaulting information. Internet searches (keywords), and lifestyle data (consumption patterns, etc.) might be included as well to infer the financial status of an individual. This information also enables targeted advertising for financial products: loans, mortgages, etc. aimed at potential customers. Some of the major data brokers offering credit risk products are *eBureau*, *Equifax* and *ID Analytics*.

The collected personal data comes from sources both online and offline. "Thousands of data brokers keep tabs on everything from social-media profiles and online searches to public records and retail loyalty cards; they likely know things including (but not limited to) your age, race, gender, and income; who your friends are; whether you're ill, looking for a job, getting married, having a baby, or trying to buy a home."[250] In exchange, the data broker industry claims that consumers benefit from the data collection with reduced debt through better screening of potentially bad or high risk customers.[251] Credit scoring has become more complex and has shifted from indebtedness records and potential income levels that reveal the creditworthiness of a person, to accurate lifestyle reports used to guess the likelihood that a person will pay his or her debts. For credit assessment, everything counts and might be weighed in the final score.

The emergence of a consumer data market has both stimulated and expanded debt-financed consumption. As of September 2012, total consumer indebtedness in the USA stood at 11.31 trillion US dollars, more than doubling the 2000 figure of 5.06 trillion.[252] Some authors claim that the increased credit demand has been fostered precisely by the consumer data broker industry,[253] which in turn would have had effects in the subprime mortgage crisis in the US. In fact "from 2005 to 2007, the height of the boom in the United States mortgage and financial-services companies were among the top spenders for online ads". All the major online marketing companies, including search engines such as *Yahoo*, *Bing*, and *Google*, are significantly involved in generating revenues through online financial marketing.[254] Moreover, in a self-reinforcing loop, the financial difficulties provoked by the

subprime mortgage crisis would have encouraged credit lenders to refine their decisions through improved and more data-rich analysis and tools.

These practices are not exempt of controversy. Credit scoring has raised concerns because using scoring systems to minimise risk could produce discrimination through a sort of "digital redlining."[255] Consumer valuation or buying-power scores rank citizens according to their potential value as customers, using financial variables but also other inputs that can include ethnicity, gender, religion, place of origin, age, etc. and make automatic assumptions about these. As the resulting scores are private digital rankings with no official oversight that use their own algorithms, the client would never know what their data double reveals. Banks, credit and debit card providers, insurers and online educational institutions regularly use these kinds of scores to make decisions.

As these are private tools, there are no guarantees that they will be used ethically. If the consultation of this kind of scores became a generalised practice, it could introduce unfair practices in the market, with financial institutions avoiding people with low scores, denying them access to home loans, credit cards or insurance. This might put some consumers at a disadvantage, especially those under financial stress. Moreover, financial scores can inform marketing scores, and vice versa, leaving the citizen unable to escape the judgement of the score. The law does not properly cover the data brokers "digital" evaluation systems and the FCRA "does little to ensure that consumer data broker companies protect consumers' personal financial information, and do not call for any penalties in the event of data breaches."[256] While all companies must have a legally permissible purpose if they want to check consumers' credit reports and must alert them if they are denied credit or insurance based on information in those reports, these regulations are not fully applicable to the new valuation scores because they leverage nontraditional data and promoted for marketing.[257]

Inaccurate scoring (e.g. through outdated information or inaccuracies or mistakes in data collection) could unfairly make more difficult the access to goods and services for certain people. Credit scoring companies could make use of questionable variables and the lack of transparency and specific regulation makes it difficult to exert controls over score-based assessments. Furthermore, trying to opt-out of such databases is currently virtually impossible. In the worst case scenario, differential service and unequal attention could threaten the principles of the free market and "vicious circle" effects might appear. If scoring becomes a common practice (due to the sophistication of this kind of products), customers who already received a low score could have problems to overturn this situation -an individual under financial stress would have more difficulties to find credit for investments, which could in turn make their business less competitive in the market, and the low revenues would keep their scores low. They would in turn be offered other kind of products (e.g. subprime loans), with the risk of worsening their financial situation even further.

4.2.1 eBureau (eScore)

eBureau is a provider of predictive analytics and information solutions founded in 2004, It uses big data assets to help businesses acquire customers, manage risks and maintain customers' loyalty. They have access to vast amounts of predictive data, managing insights that help make critical decisions throughout the customer lifecycle. These services are addressed to Business-to-Consumer (B2C) and Business-to-Business (B2B) companies, in order to improve their profitability, boost efficiency, reduce losses and increase revenue.

Gordy Meyer, founder and CEO of *eBureau*, acquired his expertise in Fingerhut, a company specialised in marketing to mid- and low-income customers. In the 90s, he leveraged his spotting

patterns of fraud and founded *RiskWise*, an analytics enterprise. After selling this and other two companies to LexisNexis in 2000, he founded *eBureau*. Big companies used to hire data analytics to rate consumers, so he focused the goal of this firm in providing customised scoring systems to midsize companies. Every month, *eBureau* scores about 20 million American adults for clients like banks, payday lenders and insurers, looking to buy the names of prospective, reliable, creditworthy customers[258].

eBureau assess companies through data-driven decisions about their customers in aspects like which groups are more likely to become customers, which customers are likely to pay their bills on time, when are there elevated fraud risks and how to most efficiently collect past due bills. *eBureau's* patented technology offers several ready-to-use solutions and has the flexibility to customise a solution in the areas of marketing & lead management, fraud prevention, credit risk assessment and collections and recovery.

eScore[259] is their "flaghship product", a customised predictive scoring tool aimed at increasing revenues, reducing costs, improving profitability and gaining a competitive advantage offering services related to marketing, lead management, fraud prevention, credit risk assessment and accounts receivable management. This tool transforms *eBureau's* informational input (a vast data network that integrates billions of records across thousands of databases) into useful information to make decisions. In order to integrate the datasets, eScore has access to critical information like summarised consumer credit data, real property and asset records, household demographic information, multiple files containing name, address, telephone and date of birth, internet, catalogue and direct marketing purchase histories, and various public records such as bankruptcy and deceased files. A key value source of this tool is the combination of online updated data (including historical records) with retrospective data to improve accuracy. The functioning of *eScore* is based on datasets matching and variable inferring: a customer submits a dataset with names of tens of thousands of sales leads previously bought, as well as names of leads who went on to become customers. Then *eBureau* introduces additional details from its databases to each customer profile: age, income, occupation, property value, length of residence and retail history, etc. At this point, the system extrapolates up to 50,000 additional variables per person and the data is analysed in search of rare common factors among the existing customer base. Prospective customers are detected based on their resemblance to previous customers. eScores might range from 0 to 99, with 99 indicating a consumer who is a likely return on investment and 0 indicating an unprofitable one.[260] *eBureau* charges clients 3 to 75 cents a score, depending on the industry and the volume of leads.

eBureau's credit risk solutions are aimed at helping companies make better credit decisions on applicants interested in the products and services offered by a company. The obtained information optimises the customer acquisition process maintaining or lowering bad debt losses, either alone or when used in conjunction with other credit resources.[261] This company lists 3 credit risk assessment product applications: *Thin & No-File* (for consumers who do not have a scoreable credit file with the major credit reporting agencies), *Non-prime & Underbanked* (up to 60 million consumers need alternative credit products that are not "mainstream", and also have very similar credit scores from the major credit reporting agencies, so it's harder to differentiate good risks from bad) and *Credit Super Scores* (*eBureau* data combined with other credit data sources to generate a credit "super score" that may result in a 20-40 percent improvement in credit risk segmentation).

Along eScore, *eBureau* leverages other tools for their assessments, like Income Estimator, a model-driven information append service that helps consumer-facing companies to estimate a person's income, and eLink, a service that helps accounts receivable management firms and departments locate, update, and append information to a debtor record. With eLink, collection departments and

collection companies can obtain up-to-date telephone and address contact information and be alerted to bankruptcies, deceased individuals and litigious debtors.

It is not easy for regulators to know if companies are using financial backgrounds or marketing scores to make decisions. David Vladeck, the director of the bureau of consumer protection at the Federal Trade Commission warns: "The scoring is a tool to enable financial institutions to make decisions about financing based on unconventional methods". E. Mierzwinski and Jeffrey Chester, of the Center for Digital Democracy, state that "the interplay among the traditional consumer reporting agencies, lenders, online data brokers, and interactive digital financial advertising has blurred the line between the traditional definitions of consumer reporting agency and target marketing,"[262] and they recommend federal regulators to ensure that consumers know the way they have been scored or rated.

eBureau won the 2011 Data Champion Awards organised by *BlueKai* in order to recognise companies that are innovating and using unmatched data-driven techniques to drive higher performance audience targeting for their clients. *eBureau*'s winning case featured a for-profit university that improved their online display advertising strategy by better defining and targeting their audience, resulting in an increase in both brand awareness and lead generation activity.[263]

In the summer 2014, *eBureau* and Oxxford Information Technology announced a long-term strategic alliance to improve assessing fraud and credit risk at time of acquisition and measuring the probability of recovering customer debt from small businesses. This alliance implies the combination of Oxxford's business data (almost 97% of all operating companies in the U.S.) and *eBureau*'s coverage to provide fraud, credit risk and collection insights into U.S. small businesses, especially those with less than10 million US dollars in sales.[264]

In relation to the lack of regulation and controversial practices of this sector, *eBureau* "went to great lengths to build a system with both regulatory requirements and consumer privacy in mind". For this purpose, the company established "firewalls in place to separate databases containing federally regulated data, like credit or debt information used for purposes like risk management, from databases about consumers used to generate scores for marketing purposes."[265] According to the company policy, among the measures taken to increase privacy standards, *eBureau* does not sell consumer data to others, nor does it retain the scores it transmits to clients. They also offer clear information regarding their privacy vision and consumer choices in their website, allowing access to a data report and opt-out request forms.[266]

4.3 Digital identities in public service provision

Regarding the data managed by the public administration, it is important to highlight the difference between personal data and non-personal data. A large extent of the information owned by the public bodies, which could be of interest for other organisations or individuals (like geographical or meteorological information), is not related to directly identifiable citizens. Even statistical data is commonly anonymised from the early stages of the research, even if the level of detail in databases like the census can make simple anonymisation techniques useless, especially in the face of personalised searches. Moreover, public bodies hold large amounts of personal information about individuals, mainly for tax and security purposes, which they sometimes have to share with third parties, such as political parties before elections.

In recent years, the "open data" movement,[267] which calls for higher levels of transparency and information sharing for public service activities, has managed to put pressure on public bodies so that they release some of these datasets in raw form. Most public bodies have been reluctant to do so, partly due to data protection concerns but also due to a long-standing culture of secretism when it comes to official data, and therefore most data, if shared at all, is presented in its final form, which makes it difficult for third parties to analyse it further or suggest alternative data analysis approaches.

Nevertheless, some search companies offer a compilation of personal details obtained through searches that partly involve information available through public administration records, which shows that regardless of anonymisation, the matching of numerous datasets, even when done on the basis of non-personal data, can often expose valuable information about an individual and form an unaccountable data double. As shown above, *Intelius* is one of these companies.

There are numerous concerns and complaints around the activities carried out by these kind of companies. It is not clear whether they comply with data protection regulations, especially in what concerns data subjects' rights. Moreover, the recent "Right to be forgotten"[268] ruling in EU could affect the activities of data brokers that retrieve information from internet search engines, public or private. There have also been some complaints regarding the quality and reliability of the data (outdated databases, inaccurate information, etc.), and the commodification of the information publicly available through public files, bulletins or Internet public resources like social media profiles or indexed pieces of data that can be found using other means.

4.3.1 GOV.UK Verify

The spread and penetration of ICT networks in daily lives as a phenomenon has not only affected individuals as consumers, but also as citizens. Public offices have progressively introduced e-government services to facilitate administrative tasks and to spare public capacities maintenance costs. Digital services can allow the automation of procedures that used to imply face-to-face interaction with civil servants. In April 2014, for instance, the UK Government started the Digital Transformation Programme to make 25 major services digital by default. As of June of 2015, 15 of them are fully working, 9 of them in public Beta phase and 1 in Alpha phase.[269] According to a Cabinet Office spokesperson, the digital version of these 25 "exemplar" online services will help save 1.7 billion pounds a year.[270]

In order to use these services, citizens need to prove their identity to avoid identity theft and fraud. Identity assurance specifically refers to the degree of certainty of an identity assertion made by an identity provider by presenting an identity credential to the relying party. Identity claims have been traditionally made through "physical" credentials like identity cards. However, not all countries have developed standardized national ID schemes. An acceptable degree of certainty (assurance level) demands different inputs to prove that the claimed identity matches the identity of the provider. For this, UK's Government Digital service (GDS) launched the **UK Identity Assurance Programme**, with the "GOV.UK Verify" system.

Since UK citizens do not hold ID cards, as all government attempts to develop one have been faced with opposition from broad constituencies, GDS had to look for a decentralised alternative based on public-private partnerships. The assurance programme is thus based on a federated identity scheme that leverages multiple distinct identity management systems at the same time. Citizens need to initially go through an enrolment procedure that should take less than 10 minutes. After that, users are able to log in much more quickly using their acquired digital identity. They are asked to verify

their identity via a credit reference agency -currently Experian and Verizon, and in the future also via Digidentity and the Post Office, and up to a total of 9 providers.

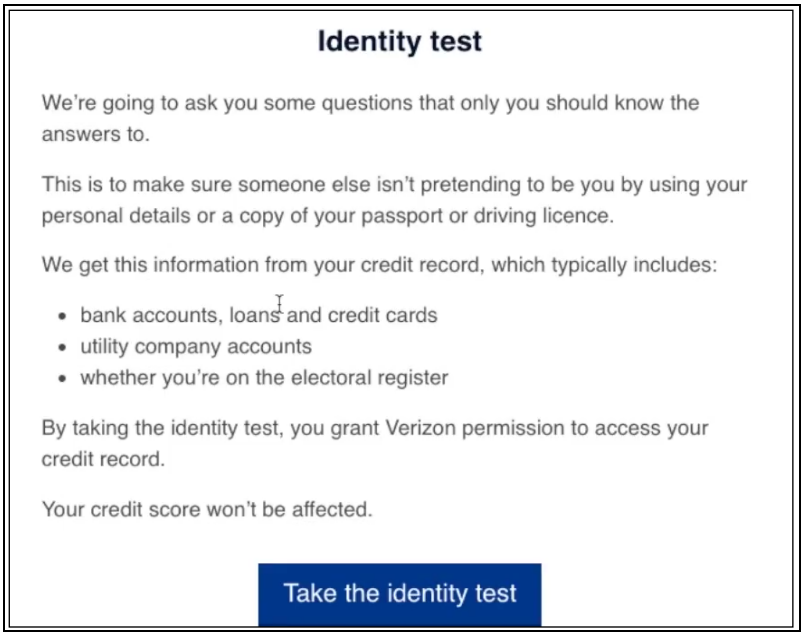

Figure 11. Screenshot with an identity test sample by Verizon.

The confidence of an individual's identity is organised around four different levels. The lowest level is for the creation of simple accounts to receive reports or updates. The second level requires that "on the balance of probability" someone proves to be who they say they are. The third level requires identity "beyond reasonable doubt" (e.g. showing a passport) and level four requires confirmation, using biometrics. Most services require level two of authentication of their identity.[271]

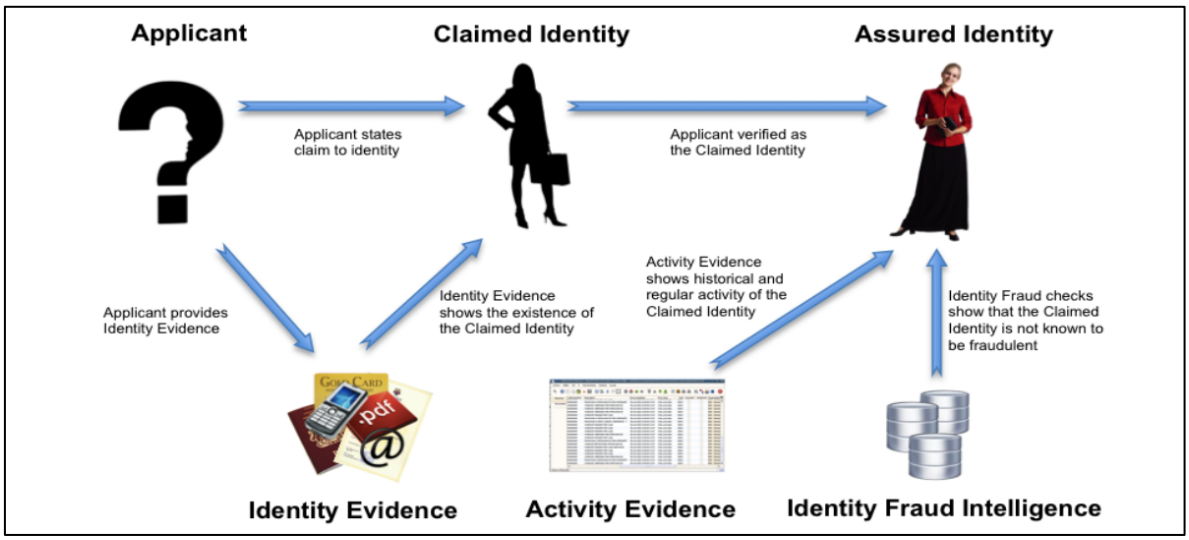

Figure 11. Overview of the Identity Proofing and Verification Process. Source: Good Practice Guide 45.

GOV.UK Verify works with **Open Identity Exchange UK** (OIX UK),[272] a non-profit trade organisation started by the US government an key industry players that aims to enable the expansion of online identity services and the adoption of new online identity products. OIX UK works closely with the Cabinet Office on the Identity Assurance Programme, which is also applied to other, non-government websites where proof of identity is needed.

A key feature of the identity validation system developed for GOV.UK Verify is its use of a federated identity, as mentioned above. The first system developed, called Gateway, was set up in 2001, but in 2011 the National Audit Office (NAO) warned that it should be replaced with a better alternative. "The Government Gateway provides only limited levels of identity assurance and, without further investment, its weaknesses will be increasingly exposed and under attack. Extending the Gateway's life will delay the delivery of the digital-by-default agenda which needs higher levels of identity assurance."[273] The current programme uses a "hub" (a technical intersection) that allows identity providers to authenticate identities without the government centrally storing an individual's data, without breaching privacy by exchanging unnecessary data and by promoting that the transacting parties openly share user details.

As of March 2015, 25,600 user verifications and 55,000 sign-ins have gone through the system, and around 5,000 people a day are currently verifying their identity and accessing services through GOV.UK Verify. In October 2014 the government said that nearly 500,000 users would be on GOV.UK Verify by April 2015, and the plan is for all individuals to use this identification assurance by March 2016. GOV.UK Verify is being tested in public beta with users for the following departments and services:[274]

- Renewing tax credits online, with Her Majesty's Revenue and Customs (HMRC)
- Claiming a tax refund (HMRC)
- Claiming redundancy and monies owed, with the Department for Business, Innovation and Skills (BIS)
- Logging in and filing a Self Assessment tax return (HMRC)
- Claiming rural payments, with the Department for Environment, Food and Rural Affairs (Defra)
- Helping friends or family with their taxes (HMRC)
- Checking or updating company car taxes (HMRC)

In order to fully verify their identity account with a certified provider, everyone using GOV.UK Verify need to have lived in the UK for more than a year, be over 19, have a photocard, a driving licence or UK passport, and have access to their financial records. If these requirements cannot be met, users are only allowed to set up a basic identity account, which they can use to perform relatively low-risk actions online.

The current success rate of the system is claimed to be at 90%. However, **the evidence of the existence for an individual is linked to their financial activity in the UK**, which has proved controversial. To test a user's identity, the system uses payment information and financial services and products like credit cards, utility bills or mortgage information. It implies that private identification resources are used for public purposes, blurring the lines of data ownership and processing. Even though the system uses information that most people generate, it only recognises citizens based on their consumption patterns. This means that those who have a limited credit or no history at all are at a disadvantage -young people and newcomers may find it more difficult to use the

system, for reasons unrelated to their entitlement to use it or relation to the state. Since the system is just an additional way to relate to the state, it does not have severe implications on equality. However, it does provide additional advantages to part of the population. Since none of the services offered are related to political participation issues, discrimination patterns are not yet an issue.

Since the UK government relies on market providers to provide this identification service, the CESG, the UK's National Technical Authority on Information Assurance and Cabinet Office, and the Government Digital Service have issued the Good Practice Guide No. 45 on "Identity Proofing and Verification of an Individual" (GPG 45). The guide establishes that providers need to rely on the breadth of the evidence, the strength of the evidence, the validation and verification processes carried out, and a history of the user's activity in order to determine different levels of assurance when verifying an identity.[275] Interestingly, GOV.UK Verify was developed in close cooperation with privacy advocacy groups like No2ID, Big Brother Watch, the University of Oxford's Internet Institute, the Consumers Association, and the privacy regulator, the Information Commissioner's Office. As a result, a document called "Privacy and Consumer Advisory Group: Draft Identity Assurance Principles" was issued to contribute with additional guidance to the service implementation.[276] In their recommendations, the advisory group put forward 9 Identity Assurance Principles:

Identity Assurance Principle	Summary of the control afforded to an individual
1. The User Control Principle	Identity assurance activities can only take place if I consent or approve them.
2. The Transparency Principle	Identity assurance can only take place in ways I understand and when I am fully informed.
3. The Multiplicity Principle	I can use and choose as many different identifiers or identity providers as I want to.
4. The Data Minimization Principle	My request or transaction only uses the minimum data that is necessary to meet my needs.
5. The Data Quality Principle	I choose when to update my records.
6. The Service-User Access and Portability Principle	I have to be provided with copies of all of my data on request; I can move/remove my data whenever I want.
7. The Governance/Certification Principle	I can have confidence in any Identity Assurance System because all the participants have to be accredited
8. The Problem Resolution Principle	If there is a problem I know there is an independent arbiter who can find a solution.
9. The Exceptional Circumstances Principle	Any exception has to be approved by Parliament and is subject to independent scrutiny.

Table 1. Identity Assurance Principles according the Privacy and Consumer Advisory Group

Nevertheless, scholars and activists recently published an academic paper exposing serious privacy and security shortcomings on this system and its US counterpart, the Federal Cloud Credential Exchange (FCCX). According to the authors, these systems could "link interactions of the same user across different service providers" and might facilitate the undetectably impersonation of users. Despite the encryption of the different parts connected in this federated system, the central GDS-built hub acts as a single point of failure. According to George Danezis, "the hub sits in the middle, despite different parts of the system being encrypted. The hub can decrypt all the information."[277]

Due to these vulnerabilities, the authors suggest a more in-depth technical and public review inspired on a threat model and adopting the corresponding structural adjustments.[278] As a reaction to this paper, GDS emphasised the convenience, security and privacy-protecting design of GOV.UK Verify.

Moreover, privacy is not only about technical standards, acceptability concerns have been pointed out by a member of Big Brother Watch, for instance, who noted that "It feels inevitable that this will happen because of the government's 'digital by default' drive (…). If it's done in a proportionate and secure way, that's good. But it has to feel like it isn't imposed, and it has to be clear how it works. This is the first time that private companies are being asked to verify peoples' identities. How it works might confuse some people."[279]

4.4 Political profiling

Political opinions belong to the set of personal data referred to as "sensitive data", which deserves special protection under de EU data protection directive.[280] To accurately define what is understood as a "political opinion" in the digital ecosystem, however, is no easy task. It is not the same to publish an indexable op-ed in a digital newspaper than to post a comment in a restricted platform. Facebook's "likes" or belonging to a group in a Social Networking Site are subtle issues that may help to guess the political orientation of an individual. Furthermore, recent facial recognition technologies are able to accurately identify persons, which could be used to reinforce the capabilities to determine the political beliefs of citizens taking part in political activities like demonstrations.

Political profiling has two main variants. On the one hand, the analysis of specific sociodemographic variables to identify subsets of potential voters, supporters or targeted audiences for a politically-based initiative like electoral campaigns, signing petitions, lobbying and reputation actions, and so on;on the other hand, the identification of radicalisation processes and monitoring of political activism. Governments have made use of digital platforms to detect signs of networked activism or "offensive comments". Tags (especially Twitter's hashtags), traceable keywords, activity in virtual communities and other digital records may be used by data scientists and social media analysts to detect political behaviour patterns, and political profiling based on digital shared contents may be used for very different purposes such as electoral marketing or law enforcement.

Informed citizens are more likely to engage in politics, and, according to some authors, people online are more keen to participate in political activities.[281] Moreover, online people tend to search for information that reinforce their political views (a phenomenon known as *selective exposure*) and to ignore those that question their ideas (*selective avoidance*),[282] thus making political preferences easy to infer on the basis of your social networks and online interactions. Political organisations make an increasing use of databases and internet technologies for fundraising, volunteers, organising, gathering intelligence on voters and doing opposition research. Already in 2005, Philip N. Howard examined the role of digital technologies in the production of contemporary political culture after analysing four election seasons between 1996 and 2002. He discovered how the diffusion of political information and the ease for people to express themselves politically affect key concepts like democracy and citizenship[283]. On a different note, the existence of detailed political dossiers on every US-voter has been analysed by Ira S. Rubenstein[284], remarking the privacy implications of these practices and alerting that these dossiers suppose the largest unregulated assemblage of personal data in contemporary American life. The author also suggests solutions based on mandatory disclosure and disclaimer regime aimed at improving transparency levels for voter microtargeting and related campaign data practices, and recommends the enactment of new federal privacy restrictions

on commercial data brokers to increase controls on firms providing data consulting services to political campaigns. More recently, Nickerson and Rogers[285] have described the utility and evolution of data in political campaigns: from publicly available files of official voters to purchased data collected by commercial firms (updated phone numbers, estimated years of education, home ownership status, mortgage information, etc.), and the information voluntarily given when citizens sign up at a candidate's website or party website.

In US, even though the secrecy of the vote is respected, the government collects certain details on citizens' political participation. Citizens have to provide their identity and address during voting registration; party affiliation is registered in some districts, and donors of amounts over 200 US dollars have to observe the federal fundraising rules and provide personal information such as their name, address and employment. This information belongs to the public record and its use may not be limited.[286] For this reason, it is used by companies like Aristotle, that develops voter profiling software, manages voter information databases and tracks voter correspondence with elected officials,[287] Even though on their own these databases do not provide much information, this may change if added to larger data repositories. Tools like *voterlistsonline.com* (also developed by Aristotle) may offer information about "super voters, absentee voters, party faithful or any other targeted group of voters you choose". Using more than 4,000 election boards, county clerks and voter registrars, key information such as party affiliation, race, age, voting history and school board districts is gathered. However, this service is exclusively addressed at political campaigns, candidates, and consultants. Due to the security of the personal information it manages, Aristotle is required to verify the data buyer's identity and the validity of the use that will be given to the voter file access requested.

This leads to an unavoidable paradox: the digital resources applied to the democratic processes may enhance the participation and widen the scope of information sources. At the same time, the digitalisation of the political activities and facts brings risks and increases the chances to carry out undemocratic and unethical actions.

4.4.1 Electoral marketing and the 2008 Obama campaign

In 2008, the year Barack Hussein Obama won the US election the development of the internet was in a crucial phase. Web 2.0 platforms emerged as key actors for the net, introducing or reinforcing new trends in techno-sociological aspects (a more embracing and participative network where information flows boomed) and the financial dimension (the data mining potential started to grow, offering real time tracking of personal data at speeds and volumes never seen before). Two of the social network sites that would become household names had just appeared. Twitter (2006) and Facebook (2004) were still relatively recent platforms, but many of their early users, digital natives, were allowed to vote for the first time. Moreover, sites like *Sixdegrees.com*, *Friendster* or *MySpace* had created the scene for new, multidimensional forms in a participative web 2.0.

Obama chose the online strategy since the very beginning. During the Democrat's presidential primaries, he followed a path started by Howard Dean in 2003-2004 in what was considered the "first online campaign",[288] raising over 25 million US dollars from small donations over the internet (the average amount was just 80 dollars) and organising online referendums to make decisions during the campaign.

Obama's supporters were keener to use the internet for political purposes than any other candidate in that year. Among democrats' voters, it was more likely for Obama's supporters to be internet

users (82%) than Clinton's, due to their age and educational level.[289] The engagement of young voters was crucial. Even though they lack of the economic power to be sizeable donors, they have plenty of resources in terms of "digital capital" to mobilise other donors, supporters and voters, and to amplify the campaign through their networks. Obama took advantage this "microcaption" of voters through what David Plouffe, his campaign director, called: a "persuasion army".[290] A centralised communications system allowed the team to mobilise and address their voters and supporters in real time and through means that felt more personal and close than traditional outlets and channels.

The Obama campaign also took advantage of another new trend. The growing reliance on the internet as a news source, at the expense of television and traditional papers –so much so that Obama's team decided against the use of press clips.

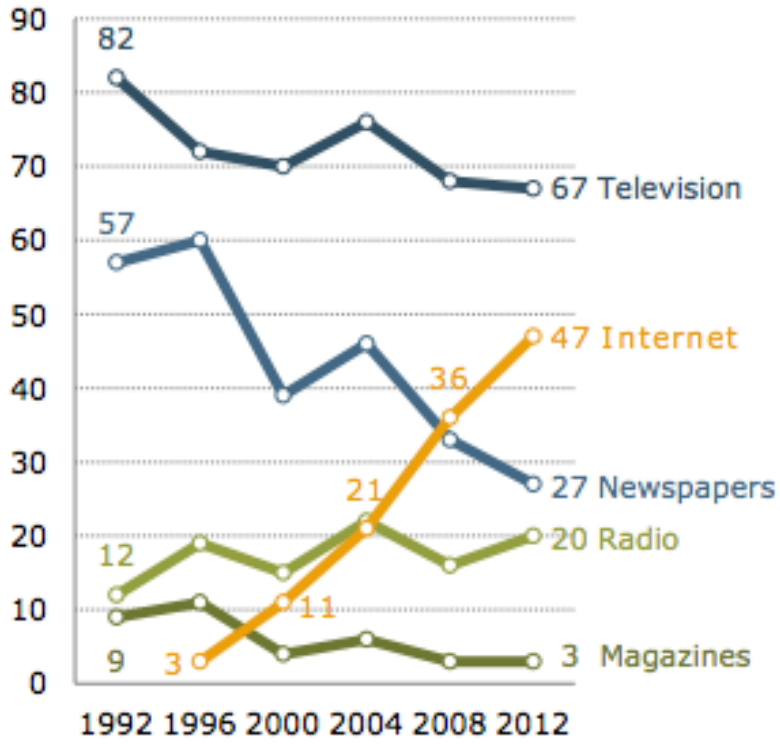

Figure 12. Evolution of different campaign news sources. Source: Bearingdrift.com.

In 2008 Barack Obama spent 16 million US dollars on online advertising (a figure that went up to 47 million in the 2012 budget for re-election) while John McCain only spent 3.6 million US dollars. At the end of that year, Facebook had about 200 million users and Twitter around 6 million. Over 2m people clicked the "like" button to show support for Obama on Facebook, and at the time of the election the future President had 115,000 followers on Twitter. During the campaign, Obama managed to reach 5 million supporters over 15 different social networks, with Twitter, Facebook and YouTube accounting for the majority of these supporters.[291]

Five companies received more than 500,000 US dollars from the online spending budget: *Google*, *Yahoo*!, *Centro*, *Advertising.com* and *Facebook*. Below in the list were the digital versions of "traditional" media companies like *CNN.com* (Turner Broadcasting), *Time* or *The Washington Post*.

Top Recipients of Obama Campaign Online Media Spending in 2008	
Media Company	**Estimated Amount Paid**
Google	$7,500,000
Yahoo	$1,500,000
Centro	$1,300,000
Advertising.com	$947,000
Facebook	$643,000
Turner Broadcasting/CNN.com	$461,000
Microsoft	$405,000
AOL	$313,000
Interclick	$222,000
Pulse 360	$222,000
Quigo	$195,000
Collective Media	$168,000
Politico	$151,000
Blogads	$149,000
Time	$147,000
BET Digital	$138,000
Pontiflex	$137,000
Washington Post	$125,000
Undertone Networks	$110,000
The Weather Channel	$108,000

Table 2. Tope recipients of Obama Campaign Online Media Spending. Source: Klickz.com

Obama also turned to people with experience in the new online world when forming his team. He hired Hans Riemer from *Rack the Vote* and Chris Huges, co-founder of Facebook and designer of MyBarackObama.com, Huges worked full-time for the campaign and coordinated the social network strategy of the future president, which worked better than McCain's.[292] In 2012, the online strategy was even more refined, and Obama's team could "predict which types of people could be persuaded by which forms of contact and content."[293] The communications were segmented, targeted and personalised, and call lists were ranked "in order of persuadability allowing them predict donor behaviours and to mobilize volunteers to get people out to vote, particularly in the critical swing states."[294] Demzilla & Datamart are the names of the databases developed for the Democrat Party to file the names of volunteers, activists, local and state party leaders, and members of the press. The Republican Party uses similar tools, called the Voter Vault. Between the two parties they have more than 150 million entries.[295]

Since Obama's success, political profiling based on online activities and using the resources created by the internet industry has become the norm in political campaigns and also when holding office, as new platforms allow public figures to create different platforms where to interact with their supporters and get their insight before taking decisions. However, using big data for political

purposes, even in democratic states, continues to be a tricky issue due to the sensitive character of the data gathered, but also the level of detail that is often found in or can be inferred from political databases (ethnicity, sexual orientation, health, etc.).

Other risks relate to the potential effects of voter profiling on political participation patterns. The digital divide could create new segments of "informed", "misinformed" and/or "overinformed" citizens, and this could affect their choices and their actual freedom to choose. Despite the participative background depicted in the context of the political digitalisation, the management of campaigns through profile-based targeting could lead to a specialisation of participation, resulting in the pragmatic categorisation of "passive" supporters (donors/volunteers/voters, etc.). The use of big data in politics might also reinforce the application of corporate or managerial principles to political engagement. "Catch-all" party strategies could see voters and potential voters as customers to be lured into making specific decision, instead of active political actors and subjects of sovereignty. This in turn could lead politicians to make decisions on the basis of people's data doubles, and not their flesh-and-bones, offline versions in a sort of "data absolutism" –*everything for the people, with the data of the people, but without the people*.

Finally, the commodification of personal data turns this resource into a valuable asset that not all political parties and organisations can afford. If personal data becomes a key resource for the success of a political campaign, and the access to big data is determined by the economic power of an organisation, the economic bias for political parties during electoral campaigns would be reinforced. Differential mobilisation capacities could thus introduce disadvantages for new political parties with lower budgets.

4.5 Personal data market in e-education

The education sector has embraced the data-driven model as a means to improve pedagogy in the digital technology era. At a time where standardised tests and teaching methods dominate the educational landscape, opportunities to provide personal learning options to students are very attractive to school districts. Personalised learning is offered through e-education, enabled by access to reliable and fast internet connection. As students engage with education technologies (EdTech), the software collects vast quantities of information about each child's actions and progress, which is then tracked and analysed longitudinally with the intention of improving the quality of education for each pupil.

Ben Williamson has widely contributed to the analysis of this topic. He analysed the concept of "Smart-school" (analogue to the idea of "smart city"), described as "emerging "sociotechnical imaginaries" formed of a mixture of technological fantasies and related technical developments."[296] Among his empirical references, he listed commercial initiatives (e.g. IBM's "Smarter Classroom" project and Microsoft's "Educated Cities" programme) and non-commercial projects (e.g. Nesta's Policy Lab and Glasgow City Council). These initiatives are based on the idea that "quantified students" (i.e., pupils that are being measured through an increasing amount of variables that go far beyond conventional evaluation marks) learn better, and that quantified students' technologies can anticipate student behaviour and optimise learning processes. As the same author points out, certain cross-sectoral intermediary organisations are promoting the joint utilisation of network-based communications and database-driven information processing software to optimise the educational decision-making by leveraging socio-algorithmic forms of power. This seeks to increase the capacity

to predict, govern and activate learners' capacities and subjectivities[297] while enacting at the same time new landscapes of "digital governance".

E-education commonly takes the form of apps, online games and learning management platforms, of which the majority are open sourced and free to use. These tools use big data and learning analytics to supply many acknowledged pedagogical benefits. The ability to use big data to provide feedback enables students to understand where their areas of weakness are located, as well as how their performance relates to that of their peers. Students often become motivated to work harder through the process of engaging with their personalized e-learning environment. Efficiency is often optimised with big data, since patterns and relationships become evident when analysed over time. The process of maintaining an effective learning management system requires collaboration between departments within schools, which often extends into improvements in other areas of the school. Tracking of individual students' learning proves to be useful for their individual learning since they can begin to understand their own work ethic and abilities. By tracking the behaviour across an entire course when engaging in online tests and readings, it is possible to review which parts of the syllabus were too easy, which readings spurred the greatest student engagement, and other important information about learning which would not necessarily be available to schools in through any other mechanism.[298]

The EdTech app *ClassDojo*, for example, is designed to aid teachers in promoting positive student behaviour in the classroom, thereby freeing up more time for teaching rather than reprimanding. Each student in a class is assigned a personal profile in the app, to which teachers can award and deduct points, and also view longitudinal trends of each child's behaviour over time. ClassDojo states that 1 in 2 schools in the USA make use of the program. Despite their success in attracting users, ClassDojo has not yet yielded profits and does not have a revenue plan. In 2013, however, it managed to raise over 10 million US dollars from investors.

Moodle, an open source software-learning program used in primary and secondary schools as well as universities, allows students and course administrators to communicate, collaborate, share resources and complete online assignments. Every student action within the platform is recorded by the software, from the exact pages clicked on to the amount of time spent on each test question. This allows administrators to access a back end action-profile of each student, which can then be used to produce trend analyses over time. According to their website, *Moodle* is funded by the contributions of its "partners", authorised companies that help and users with Moodle implementation and give 10% of their earning to *Moodle Pty Ltd* in Australia.[299]

Another popular solution, *Snapshot*, is a free micro-assessment tool which enables teachers to assign quizzes based on subject material to their classes, which the software then marks and analyses for understanding. Based on their test results, pupils are assigned to one of the three following categories: those who have met the standard, students who are borderline, and children who lag behind the established standard. From this point, teachers can use the results to personalsze learning to each student's needs.

In each of the above examples, students' actions and abilities are monitored and algorithmically analysed by the EdTech software, then used to assign the student to categories and inform education decisions and life paths in the future.

These widespread EdTech products are owned by private companies. Therefore, their main drive is profit, and personal data can be very valuable in this context. In an age where data vendors' business models centre on maximising the aggregation of personal information about a person, and selling it to third parties, data about children's learning is highly sensitive, valuable and vulnerable, and the use

of information collected through EdTech to target advertising and marketing to children has the potential to be very lucrative. Students' young age and lack of experience and understanding of the world makes them especially susceptible to targeting advertising and oblivious to nefarious practices. Moreover, profiling practices which categorise students using quantitative variables can be both derogatory and harmful to youngsters, as student identity is transient, based upon social and physical context, and extremely malleable[300] while databases are fixed and permanent.

According to the Software & Information Industry Association's education division, the EdTech industry grew by 5% in 2014, reaching over 8 billion US dollars in sales in the US alone and continuing in a long-term upward spiral. In the first 3 months of 2014, EdTech companies raised over 500 million US dollars in investment capital from venture capital firms. However, the EdTech industry is also known for the amount of start-ups that have had to fold, and the absence of working business models is remarkable. The sector does have characteristics that could explain this fact, as some investors may choose to support companies in this sector as part of their Public Relations strategy, or the funding can come from philanthropists interested in making an impact in the field of education. Nonetheless, the sensitive character of minor's data is a clear drawback for the profitability of this sector in the identity market, regardless of how large their client base or how rich their databases can be. Due to these specific characteristics, this is the only case where we have chosen to review a company that is no longer in operation.

4.5.1 inBloom Inc.

inBloom Inc. was a EdTech company which provided a central database for school boards to store and manage encrypted student records, as well as to provide opportunities for personalised student learning. Its mission statement claimed that the company could "solve a common technology issue facing school districts today: the inability of electronic instructional tools used in classrooms to work in coordination with (or "talk to") one another." Previously, individual schools stored different forms of student data in a variety of databases, which did not facilitate efficient interoperability between schools or with the state, as well as hindering data sharing, comparisons and trend analysis.[301]

The open source non-profit company was funded by the Bill and Melinda Gates Foundation (who contributed 100 million US dollars to the project), the Carnegie Corporation of New York and Joel Klein.[302] At its conception *inBloom* did not charge school districts for the use of their product, however the company had plans to charge between 2-5 US dollars per child by 2015.[303] *inBloom's* business model relied on an Amazon-hosted cloud based storage system, in which schools would aggregate up to 400 distinct categories of student data, which *inBloom* would in turn share with third party vendors selling educational products and services.[304] This data ranged from name and age to extremely sensitive information such as family conditions in the home, learning disabilities and Social Security Number.

At its peak, *inBloom* was operating in 9 states in the US. However, many parents and civil rights associations expressed serious concern surrounding the privacy of student data in the business model of *inBloom*. Ultimately, after lengthy protests, all states retracted their use of *inBloom* in schools, and the company eventually closed in 2014. A number of reasons lead to the demise of the company, but all are connected to the lack of appropriate protection *inBloom* paid to student identity data. There were three components to the *inBloom* business model (see Figure 12). Firstly, schools within participating states shared information about their students with *inBloom*. Then, *inBloom* stored this information in Amazon hosted cloud-storage. Thirdly, *inBloom* likely shared information

from the educational database with third parties such as EdTech companies and other firms who could financially benefit from mining the data.

Figure 12. The process of information movement within the inBloom business model. Source: Own elaboration.

Beginning at the start of the process, school districts and *inBloom* were criticised for not requiring parental consent before moving sensitive student data from the state database to that of *inBloom*. The Federal Educational Rights and Privacy Act (FERPA), the law that oversees the gathering and use of student data by schools in the USA, underwent significant modifications in 2012, and the number of agencies which could access personally identifiable student information expanded from only the education agencies to any representative who the state or local education department assigns to evaluating federally-supported educational programs, including other state agencies and private corporations.[305] Secondly, the breadth of agents who could access personal student information through non-consensual disclosure grew to include any "educational program" involved in educational provision. These modifications resulted in schools having the legal ability to share student records without parent consent to any "school official" within a "legitimate education interest", providing they remained within the bounds of the activities defined in the contract. This included private companies hired by the school – notably *inBloom*. Parents and privacy activists were troubled by the consequences of the novel access of privacy companies to student data, and the Electronic Privacy Information Centre sued the USA Department of Education over the legality of the amendments.

inBloom's lack of transparency concerning who would access data, for what purpose and under what security precautions caused mistrust with the general public. *inBloom's* Chief Privacy Officer, Virginia Bartlett, claimed the company was being transparent in their privacy policy through stating that the company could not "guarantee the security of the information stored… or that the information will not be intercepted when it is being transmitted", and that no company can assure the security of their information.[306] However, ineffective communication about the role the company played in the data flow process, and the lack of effective Public Relations programs to adequately inform parents and students lead to confusion, misunderstanding and eventually distrust.

The greatest public outcry against *inBloom*, however, was inspired by the ambiguous way in which *inBloom* spoke about the way it planned to use the data stored in the cloud. Many believed that the company planned to share highly sensitive information about children with third party vendors and private corporations.[307] Personal data is the currency of our modern information age, and many parents and privacy advocates worried that *inBloom's* aggregation of mass amounts of student data would allow for vendors to market their learning products, apps, games and other services back to schools in order to target the children whose identities were known to them through their data doubles. There was also the fear that information would be shared with data mining companies who would sell information about students to advertising companies.

4.6 Lessons learned from the case studies

The digital economy is still a field in formation. Many of today's actors will likely disappear, and business models will evolve. However, there are several trends and key issues that can already be highlighted.

On the one hand, the role and advantage of private actors: global, flexible and adaptable firms make use of technologies that change and evolve rapidly. The capacities of big data analysis have boosted in the last three decades and the flow of data is currently measured in petabytes. The number of variables monitored grows under the logic of "if it is measurable, it will be measured", and even emotions are being targeted by data brokers, even where there is no clear use for the data gathered. This contributes to the secrecy of the field, as well as the difficulty in researching specific business models. On the other, it is apparent how often **market interests are at odds with regulatory principles**. There is a conflict between the goals of the business models that lead the economic activities of data brokers and the companies linked to them (i.e., obtaining massive data and/or extremely detailed digital depicts) and the privacy principles that guide the corresponding regulations (e.g. the minimisation principle). Other conflicting issues are the differences between regulatory contexts (e.g. for global companies) and the different types of personal data collected (sensitive data generates very attractive information flows for any data broker). The unclear (or non-existing) consent mechanisms and the transferring of data to third parties is one of the main complaints pointed out by regulators and privacy advocates, and it is still unclear how this clash will be resolved.

Another obvious trend is the blurring borders between public and private actors. The introduction of private actors for the management of political profiling, digital identity assurance programs or e-education platforms has contributed to the efficiency of these initiatives, since most of the resources were already developed (databases, know-how, technologies, etc.). Nevertheless, benefit-oriented actors may put in a second place unavoidable guarantees for services that affect such a large extent of population, like those related with the security and the privacy of the data.

Finally, it is worth pointing to the **transformation of the data brokerage market**. Even though data brokers are not a recent phenomenon, their transformation due to the evolving technologies like big data or the Internet of Things might affect the value chain and their business models. The current value cycle of data is unclear and models are based on the speculative future value of massive data collection, but actual identifiable and quantifiable revenue models have not yet emerged. Companies emerge and collapse faster than they can leave a mark or make a lasting impression. However, new frameworks and models are emerging: shared benefits from data brokerage, or a trusted flow of personal data where control (actual data ownership), value (mutually beneficial), trust (identifiable empowered authorities), and transparency (consent, terms of agreement, actual utilisation, etc.) play a significant role in the potential futures being laid out in the context of the digital economy and the identity market.

5. Regulatory Framework for identity, data protection and privacy in the EU

5.1 Overview of regulatory frameworks

Privacy is regulated in very different ways around the world. But generally speaking the OECD privacy principles provide the basic building blocks for most modern privacy legislation.[308] These are in turn based on the Fair Information Practice Principles (FIPPS) we discussed in section 1.1.

The legal implementations are very diverse though. In Europe, privacy and data protections are defined as fundamental rights, as we discuss in the next section, and covered by comprehensive legislation, currently the Data Protection Directive.

The European model of data protection has been adopted by other countries from around the world, such as the Philippines,[309] for various reasons. Privacy advocates tend to favour the EU model because it is strong and tested, but local policymakers also wish to have their country given special status for data transferred from the EU.

The US approach to privacy is completely different from the EU. The US does not have a fundamental right to privacy, as they do not recognise international human rights and this particular right is missing from their constitution. The US does not have a general privacy law either. There is a very limited Privacy Act from 1974 that introduced fair information principles but only applies to federal agencies.[310]

Specific laws cover sectors that at some point have been deemed at particular risk such as health records,[311] or famously video rentals, regulated in a special privacy act passed after a Supreme Court nominee saw his rental collections disclosed to a newspaper.[312] These US laws can be very strong in the context they regulate, but the lack of a general privacy law in the style of the EU severely hampers the privacy rights of US citizens. This was shown in the recent case of New York artist Arne Svenson, who filmed his neighbours inside their homes with a telephoto lens for an art project.[313] The court found his actions "disturbing" and "intrusive" yet had to agree that he had not broken any specific laws.[314]

Mainly, the US approach has been to let industry self-regulate. But this approach has been criticised by many scholars,[315] and also by the body responsible for regulating digital information, the Federal Trade Commission, which has repeatedly asked for "baseline privacy legislation"[316].

In an attempt to make the US more compatible with the EU, in order to help reduce potential trade issues, the Obama administration proposed a Consumer Privacy Bill of Rights in 2012[317]. But the proposals have been criticised by privacy advocates for giving companies too much leeway and consumers too little control[318]. Despite these criticisms, tech companies claim that the bill will place unbearable regulations that will stifle innovation[319].

A very positive aspect of US law that is not generally available to the same level in the EU is the possibility of class action by consumers leading to severe liability compensations. This could tip the balance if the US were to implement proper legislation, as the actual possibility of enforcement in the EU is generally quite slim and geared towards fines that do not compensate consumers for the harms suffered.

There is an on-going process of rapprochement on data flows between the US and the EU involving roundtables and other events. The US is central to any discussions of digital issues, given the prominent role of American internet companies in this sector. The view from many in the EU is that the US is leading a race to the bottom on privacy as companies from other countries are forced to accept de facto lower US standards of protection. But there is also widespread support for a realpolitik approach that proposes to lower protections in Europe for fear that EU companies will not be competitive.

But there are still other approaches around the world; the company Nymity advertises a legal tool to check compliance with 550 different privacy laws[320]. For example, in many Latin American countries the main foundation of data privacy instead rests on the concept of *Habeas Data,*[321] which is the right of citizens to demand access to data held on them and to have certain control, such as correction or deletion. This is similar but more limited to the EU data protection approach, as it does not place any constraints on what organisations can do with the data - purpose limitation - or where it is transferred.[322]

The conflicts between the US and the EU over privacy regulation have huge economic importance, and there are also discussions about data in many international free trade agreements, with calls to enable global data flows as part of economic globalisation.[323] The Asia Pacific Economic Treaty (APEC) includes its own privacy framework in order to advance this agenda[324].

The debates about digital privacy regulation are also part of wider debates on how to regulate the internet more generally. The internet has seen a very strong libertarian drive for self-regulation since its inception, coming from the US but supported by technology experts and advocates elsewhere.[325] This has led to a formal global internet governance model of multistakeholderism, where governments, companies, the technical community and civil society supposedly sit together as equals. But this model is widely acknowledged not to work properly and it is trying to reinvent itself.[326]

More recently, there have been attempts by governments at the International Telecommunications Union (ITU) to bring the internet under state control, in a similar way that telecommunications companies are regulated under formal United Nations treaties. The argument, in fairness quite reasonable, is that this lack of regulation favours the US and American companies. But this push has been fiercely resisted by internet advocates fearful, also very reasonably, that letting countries such as Iran or Russia claim absolute control over the internet would destroy the medium[327] as we know it.

The UN itself is increasingly taking privacy more seriously, and recently endorsed a Right to Privacy resolution calling for stronger privacy protections in the digital age.[328] The Human Rights Council is recruiting a Special Rapporteur on the right to privacy.[329] One important lesson from the experiences of internet and telecoms regulation is the importance of technical details, which is sometimes missing in privacy organisations stuffed with lawyers. The UK Information Commissioner only recently employed a technical expert and many privacy bodies lack the capacity to understand the systems they are meant to regulate.

5.1.1 EDPS vision for regulatory framework

The European Personal Data Supervisor (EDPS), which is the independent supervisory authority that is charged with defending privacy at the EU level, has presented a new model[330] for the regulation of privacy in the age of big data that goes wider than data protection.

The EDPS does not advocate abandoning the protection of privacy, or a softening of the rules in order to accommodate regulation to the perceived reality in the field. Instead the EDPS proposes to enhance protections through an integrated approach that combines rules on data protection, competition and consumer protection.

These rules aim to protect individuals and promote a single European market, and as we saw in the previous section the digital market is a major priority for European authorities. In this context, the EDPS sees major advantages in applying competition and consumer regulation to personal data.

For example, control over data of large numbers of users seems to translate into market power, particularly for free online services such as social media paid for with personal information. The EDPS believes that applying strict competition law to these services will promote the development of privacy friendly practices. Companies being forced to be more transparent about the data they hold may start seeing data as both asset and a liability. These companies may prefer to minimise data collection, delete some of the data they don't use or give consumers tools to exert more control. This requires a shift in the understanding of the value of personal data by EU authorities, for example during the approval of mergers.

These proposals are very positive, but there is a risk that some will interpret them as simply moving away from data protection towards consumer protection. This would be a negative development, as in Europe, privacy and data protection are rights in their own terms, not just ancillary tools to stop discrimination or support fair markets.

In sections 5.3 and 5.4 we discuss these proposals in more detail, after we look at the regulation of privacy and data protection, including e-privacy.

5.2 EU Privacy and Data protection

5.2.1 Legal foundations of privacy laws in Europe

The following is a brief overview of the legal foundations of privacy in Europe:

EU countries are signatories of the **UN International Covenant on Civil and Political Rights**, whose article 17 states "no one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to unlawful attacks on his honour and reputation."[331]

Article 8 of the **European Convention on Human Rights**[332] also gives a right for private and family life, home and correspondence. The Convention is linked to the Council of Europe, which includes countries such as Russia and Serbia,[333] not just the European Union.

The **Charter of Fundamental Rights of the EU** incorporates the rights in the Convention, including privacy in Article 7 updated from "correspondence" to "communications".[334] But it also guarantees "third generation" fundamental rights,[335] such as data protection, which is registered separately from privacy in Article 8.

One important caveat is that the Charter only applies to the EU institutions, or when countries are implementing EU legislation. In all other cases the protection of digital privacy must rely on local constitutional rights (e.g. Germany), and/or on international conventions (as in the UK).

In this context, **Convention 108 of the Council of Europe** for the Protection of Individuals with regard to Automatic Processing of Personal Data[336] from 1981, modified in 2001, is very important.

Convention 108 is an internationally binding document signed and ratified by most CoE countries and beyond, e.g. Uruguay.

These principles are then to be implemented in national laws, regulations and directives.

5.2.2 EU Data Protection Directive

The main regulation of privacy and data protection in the EU is the Data Protection Directive 1995.[337] This directive sets out principles based on Convention 108. Here we see data protection is both a subset of privacy, but also an independent right that gives people control over information about them[338]. Importantly, the Directive set terms of protection, but it also had the aim to promote the flow of data across the EU by creating harmonised common rules. This is a key angle in any discussions about privacy regulation in the EU.

The directive does not cover police and criminal justice, which are regulated separately under Convention 108 and the Cybercrime Convention. There is a proposed new Directive for Data Protection in police contexts[339], which has yet to be approved. The Directive does not apply to EU institutions either, which follow a separate Regulation 45/2001.

The Directive sets out that by default a "data controller" should not process - i.e. collect, analyse, etc. - the personal information of a "data subject" unless they have a legitimate purpose and do it lawfully.

The legitimate purpose is very important in the EU directive. It must be specific and people have a right to know about it before data is processed. Any new processing for non-compatible purposes is illegal, and transfers of data to third party count as a new purpose that just be justified anew.

Lawful processing is normally based on consent, vital interests (getting your medical files in an accident), public interest or some overriding legitimate interest of the processor or third party, the latter being one of the main points of contention as we discuss in the analysis of the new Regulation. There are also exceptions for journalism and other freedom of expression grounds. Consent is not absolute, as some people believe, and generally there is only a right to stop the processing of data that causes severe distress. People can object to some other uses of data, such as automated decisions.

The Directive also sets out obligations on those processing data to maintain data quality - I.e. Data must be relevant, accurate, up to date, etc. Importantly it sets the principle that data should not be kept for longer than needed. People have the right to obtain data about themselves, and modify it or delete in some cases.

Controllers also have an obligation to take measures for the security and confidentiality of the data, and in many counties have an obligation to report data breaches. Fairness is another important principle in EU data protection. This means that organisations must be transparent about the use of data, and people should always know what is happening with their information and how it may affect them, before the data is collected. This normally takes the form of privacy policies or similar documents.

The Directive establishes the principle of accountability, with very clear responsibilities for the so-called data controller, the legal person who decides on the collection and use of data, to ensure compliance with data protection. National laws include detailed requirements for notifications,

registrations, etc. The local laws implementing the directive have to establish a national privacy body, in the UK the Information Commissioner, and clear tribunal routes for complaints.

5.2.3 Balancing privacy with other rights

Privacy is a human right in Europe, but not all rights are equal. Some rights are absolute, such as the prohibition of torture[340]. Privacy is a qualified right, which means that the state can interfere with it under certain circumstances. Interference with human rights - e.g. surveillance - must be prescribed by law, necessary to achieve a legitimate aim, and proportionate to the aim pursued[341].

In practice, privacy will always be balanced with other rights[342] also protected in European legislation, such as:

- Freedom of expression. This includes a special exception for journalism. Generally, public figures and those involved in events of high public interest will see their privacy rights reduced.
- Access to documents, including government transparency and accountability. The publication of public registers, expenses of public officials, court records, etc. can involve interferences with the right to privacy of certain individuals affected.
- Culture, arts and science. For example the Data Protection directive makes scientific research a compatible purpose independently of the original motive for data collection. As with the other balancing acts, safeguards must be provided in law. Archives also have special considerations.
- Property, which has been used in the enforcement of copyright protections.

5.2.4 New General Data Protection Regulation

In 2012 the European Commission proposed a long awaited replacement for the Data Protection Directive from 1995. In order to avoid the fragmentation of national regimes - which was one of the main complaints from all stakeholders previously - the Commission proposed to replace the Directive with a Regulation that would provide a much high level of harmonisation across the EU.[343] The Regulation is in its final stages of legislative scrutiny.

Privacy advocates saw the initial draft Regulation as a very positive step. In contrast many industry groups reacted with alarm and set in motion one of the largest lobbying operations ever seen in Brussels. This resulted in more than 3,000 amendments[344] being presented by Member of the European Parliament (MEPs).

Despite the lobbying, many MEPs understood the need to protect fundamental rights and voted for an amended version of the Regulation that overall maintained a good level of protections. The next step in the legislative process was for the European Council, representing the governments of member states, to prepare their own amended version. Unfortunately, European governments have used the Council to try to carve out special dispensations, which has led to a hollowed out version of the Regulation. Civil society organisations wrote to President Juncker in April 2015, concerned that the new Regulation might well go below the current levels of protection afforded by the Directive.[345]

EU protocol then requires that the three main bodies: Parliament, Commission and the Council - represented by the country holding the Presidency at the time - sit down for tripartite negotiations

to agree a final version. These negotiations started in June 2015 and are expected to last at least until early 2016, with the Regulation is expected to come into force in 2018.[346]

5.2.5 E-privacy

The processing of data in electronic communications is regulated in the directive on Privacy and Electronic Communications, also known as the E-privacy Directive. This directive complements the 1995 Data Protection Directive, which at the time could not foresee in detail the specific risks brought by the convergence of electronic communications.

The E-privacy Directive is very important for the regulation of digital identities in Europe because it sets clear limits on what companies can do with personal data. Developers of digital participation platforms in the EU must take this directive into account.

Here we look at some of the most relevant aspects of the directive without attempting to provide a comprehensive overview. The transposition of the directive to each country will involve modifications to the rules beyond the scope of this short report. The main areas we consider here are confidentiality of communications, cookies and marketing.

There are some inconsistencies in the scope of organisations that have to comply with some of its sections. The main parts of the directive apply to "publicly available electronic communications services in public communications networks". This definition covers broadband, telephony and mobile operators. But it does not cover so-called "information society services" provided over the internet, from search engines to social networks, etc. This means that a telephony provider such as Vodafone has to comply with the main provisions of the directive, but voice-over-ip (VoIP) providers such as Skype do not.

The recent Digital Single Market Strategy we describe includes plans for the review of the E-privacy Directive, which may provide more consistent protections to EU citizens.

Confidentiality of Electronic Communications Data

The directive defines quite narrowly the purposes for which providers of services can use their customers' data. These roughly relate to delivering the service and being able to bill for it. Once these purposes have been achieved, any data associated with the provision of a communication should be destroyed. Alternatively, providers can ask for consent to use the data for further reuse - e.g. to provide value-added services - or render the data not personal through anonymisation. There are escape clauses for security services to be able to access data if required.

The content of the communications must always be strictly confidential, but also the associated data must be protected. The main types of data covered in the directive are: traffic data - the data associated with delivering the communication, including whom and when; subscriber data - information required for billing, etc.; and importantly also **location** data, for example from mobile phone masts.

Communications providers, including mobile phone companies,[347] increasingly try to monetise their customers' data by developing ancillary big data analytics services. But many of these services could well be in breach of the directive if they fail to either fully anonymise the data or obtain consent from their customers.

Cookies

The regulation was amended in 2009 to force internet companies to obtain consent when storing permanent information in their users' devices unless this is needed for the provision of the service.[348] Currently, this mainly affects so-called cookies in web browsers, with thousands of websites now asking users to click to agree for cookies to be installed. This has been one of the most controversial and misunderstood pieces of legislation affecting digital identities.

Cookies are small files that are placed in the users' computers to uniquely identify that machine, in order to provide some form of continuity over time. This could be simply a so-called "session cookie" that keeps a shopping basket consistent and disappears after the user closes the browser. But other permanent cookies allow third party marketing companies to track internet users' web browsing details on an on-going basis. The intrusive capacity of cookies led legislators to regulate their use, but unfortunately this has not worked as expected. With some notable exceptions, such as BT.com, websites do not offer a real choice to visitors and consistently fail to explain what kinds of cookies are used. Users are simply offered a choice to either click through on the basis of vague information, or abandon the website.

Concerns that forcing websites to obtain consent would make the internet unusable[349] have not materialised, but the situation is not satisfactory. Much of this hinges on what constitutes "freely given, specific and informed" consent. This is now under review in the new Data Protection regulation. This data collection is a key element of the Online Behavioural Advertising that fuels much the Internet and is central to the concept of Digital Identities.

Online Marketing

The directive also sets out clear obligations on online marketers to obtain consent in a move designed to stem the tsunami of spam that already in 2002 was clogging internet users' inboxes. Users must "opt-in" to marketing and must also be offered "opt-out" at any time. The 2009 amendments brought stronger obligations of transparency and a right to take action against spammers, while extending the scope to other messages such as SMS. These provisions have met with more success than the cookies, although loopholes are sometimes exploited, including pre-ticking consent boxes. But overall, there is widespread awareness and most legitimate marketers take some steps to obtain consent. Nevertheless, these regulations have not stopped the accumulation of large marketing databases.

5.2.6 Digital Identities and the EU Digital Single Market

The future European regulatory landscape for personal information beyond the GDPR is set out in the Digital Single Market (DSM) Strategy, presented by the European Commission in May 2015. The Strategy aims to move "from 28 national markets to a single one[350]", hoping that this will contribute €415 billion per year to the EU economy and create 3.8 million jobs. In order to achieve this, Europe will embrace big data, cloud services and the Internet of Things, as productivity enablers. The "free flow of data" we mentioned in the previous section in relation to free trade agreements is also a key plank of this strategy.

A full analysis of the DSM Strategy - which e.g. includes major proposals for the reform of copyright - is beyond the scope of this report. The main proposals in relation to digital identities and personal information are a mix of very concrete interventions and vaguely defined ideas.

Assessment of the role of online platforms

This will include the sharing economy and online intermediaries - and will be mainly focused on market and competition issues. The DSM Strategy acknowledges that platforms generate, accumulate and control an enormous amount of data about their customers and will also look into platforms' usage of the information they collect.

E-Government Action Plan 2016-20

The Commission will present a new e-Government Action Plan 2016-2020 which will include the interconnection of some public registers and an initiative with the Member States to pilot the 'Once-Only' principle. These proposals can be positive but carry privacy risks, as they require extensive data sharing. Calls to integrate European and national portals towards a 'Single Digital Gateway' appear unwarranted given the low volume of cross border e-government engagement, and are quite problematic in that they could centralise identity data on most EU citizens. Proposals for interoperable e-signatures may have important implications for digital identity management in Europe.

Integrated standardisation plan and review of the European Interoperability Framework

Interoperability of systems is one of the foundations of the internet but carrying digital identities across systems increases the privacy risks for individuals and opens up the question of what kind of identities will be used and who will control them. The Commission wants to focus on some specific technologies that show innovative potential - such as data driven services, cloud services, cyber security, e-health, e-transport and mobile payments - all of which require careful consideration to ensure that open standards that enable privacy are used.

Initiatives on data ownership, free flow of data and EU cloud

These proposals from the Commission - which we copy verbatim below - have huge potential implications for the workings of digital identity services.

"The Commission will propose in 2016 a European 'Free flow of data' initiative that tackles restrictions on the free movement of data for reasons other than the protection of personal data within the EU and unjustified restrictions on the location of data for storage or processing purposes. It will address the emerging issues of ownership, interoperability, usability and access to data in situations such as business-to-business, business to consumer, machine generated and machine-to-machine data. It will encourage access to public data to help drive innovation. The Commission will launch a European Cloud initiative including cloud services certification, contracts, switching of cloud services providers and a research open science cloud."[351]

Review of the Privacy of Electronic Communications (E-Privacy) Directive

After the GDPR is approved the Commission will review the ePrivacy Directive, which we described above, with discussions about extending its scope from telecoms such as Vodafone to information society services such as Skype.

5.3 Competition Law

Competition law is a central plank of the European Single Market and it is chiefly concerned with creating efficient markets that give consumers choice. The scope of competition law includes

controlling excessive market power and monitoring corporate mergers. It also has a mandate to promote trade across the EU and liberalise the public sector.

In the view of the EDPS, competition law could go beyond its traditional focus on corporate entities and choice to ensure the internal market benefits consumers through competition, "including not only the wish for competitive prices but also the wish for variety, innovation, quality and other non-price benefits, including privacy protection".[352]

The TFEU contains several articles covering many aspects of competition including cartel behaviour, market domination and discrimination against foreign EU companies. These principles are developed into a series of directives and regulations[353]. Here we will give a very simplified overview of a very complex area of legislation, with a focus on those aspects more relevant to the regulation and creation of value around digital identities.

5.3.1 Market dominance

Market dominance is important in the digital sector because, as we discussed in the previous sections, network effects tend to concentrate markets. Generally, a player is said to dominate a market when it can set prices and control production, which is normally shorthanded to having a market share of 40%, although it can be less[354] if other circumstances apply. In digital markets this is more complicated to establish. For example Microsoft and Google dominate respectively the markets for PC operating systems and search. But at the same time they are also competitors in each other's main market through the Microsoft owned Bing search engine and Chrome OS developed by Google.

Under EU law dominance is not a problem in itself, and only becomes troublesome when it is abused to unfairly exclude competitors, or exploited in a way that harms consumers. Establishing the abusive exclusion of competitors - anti-competitive foreclosure - can be quite complicated, as companies that do better because they have built objective advantages through innovation in principle should not be penalised for their success. The ultimate criterion is that there must be "no net harm to consumers"[355].There are many mechanisms a dominant firm can use to abuse its position: predatory undercutting that sacrifices sustained loses to destroy competition; refusing to supply necessary downstream products to competitors or unfairly squeezing their profit margins; discriminating unfairly in prices or charging excessive patent fees.[356] As an illustration of the wide range of activities that can fall in this category, the EC carried out an investigation that found that pharmaceutical companies were using a variety of tactics to delay the introduction of generic medicines into the market[357]. These included: patent clusters, litigation and regulatory obstruction.

In the digital sector, tying and bundling diverse products are some of the main activities that can lead to market abuse. The case of Microsoft and Windows Media Player is one of the best-known examples, where the Commission was found that the company's tying behaviour harmed competition in the market for streaming media players[358].

The same case also raises another important aspect for competition in digital markets, where computer systems are increasingly connected: interoperability. The EC found against Microsoft's refusal to share interoperability information "indispensable for competitors to be able to viably compete in the work group server operating system market"[359]. Eventually Microsoft was fined €860 million. Despite the focus on consumer harms, there is surprisingly little consensus on what constitutes the excessive pricing that may eventually result from the abuses described above. Some case law and academics have proposed criteria.[360] This is an important issue in digital markets, as it

can be very hard to establish the exact costs of digital services. Similarly, the unfairness of *low prices* involved in predatory undercutting can be hard to establish in digital markets, where free products and services are widespread.

These pricing problems relate to the underlying difficulty to measure digital market power. This is true for any kind of digital goods or services, as evidenced by the continuous disputes over the music streaming market, including the recent investigation of Apple's new venture by the Commission[361]. But it is particularly difficult in relation to personal information, as the power of an intangible asset such as data can bear little relation to actual sales volume. The EDPS[362] has proposed that competition, consumer protection and data protection authorities should "collaborate in identifying scenarios and in developing a standard for measurement of market power in this area. This standard could then be used to assess suspected infringements in the three areas."

5.3.2 Mergers

Another aspect of competition law that affects digital identities and their value is the control over mergers and acquisitions in order to avoid concentrations of corporate power that would distort effective competition. The EU regulates operations that have a "Community dimension" beyond individual countries, based on turnover, through the Merger Regulation 139/2004.

In principle merger regulators could look into whether personal data gives a company excessive market power, but this is not very common. As an exception, the German Monopolies Commission has recently published a report on digital markets which recommends[363] personal data to be considered, particularly in relation "new internet service providers, characterised by low turnover, but potentially highly valuable data inventories". The Monopolies Commission looked at search, online advertising and social networks, with concerns about the latter's tendency towards network effects and lack of interoperability.

US consumer and privacy organisations have called on the Federal Trade Commission to launch an investigation into the impact of the concentrations of data and digital markets.[364] The call was triggered by the acquisition of data broker Datalogix by the Oracle Corporation, which would give the company the ability to consolidate "a consumer's various identities across all devices, screens and channels." The perceive needed to track internet users in a much more complex environment where people access the net via phones and smart TVs has led many companies to follow a similar strategy. For example Twitter acquired the marketing technology company TellApart for its "unique cross-device retargeting capabilities".[365]

The most significant digital merger operation examined in the EU has been the acquisition of advertising company DoubleClick by Google[366], which eventually received approval from the Commission to go ahead in 2008. The Commission, applying the threshold calculation criteria, initially determined that the merger lacked a Community dimension, but due to numerous complaints it had to be considered. The Commission found that the companies were not direct competitors but part of a vertical integration strategy that was becoming common in the sector. The Commission separated Google's search activities, where the company dominated the market in the EU, and concentrated on the advertising side. Here it concluded that there were enough competitors with access to web browsing data that could also serve targeted adverts.

The Commission focused exclusively on the market aspects of the operation and made clear that their decision was without prejudice to any data and privacy considerations[367] about the merger of two large databases of internet users' behaviour.

The EDPS has been highly critical of this approach in their more recent report on big data regulation:

> *"With such a purely economic approach to the case, the Commission did not consider how the merger could have affected the users whose data would be further processed by merging the two companies' datasets, conceivably to provide services, perhaps bundled or even tied to the simple search service, that were not envisaged when the data were originally submitted. The decision did not refer to consumer welfare nor to the users of Google's search engines, even though this potentially implicated every Internet user in the EU. It therefore neglected the longer term impact on the welfare of millions of users in the event that the combined undertaking's information generated by search (Google) and browsing (DoubleClick) were later processed for incompatible purposes."*[368]

The Commission is currently examining Google's potential abuse of its search monopoly to promote its own commercial services.[369] But how the tracking and accumulation of personal information enables an unmatched search accuracy, is not been taken into account by the Commission, who is "missing the larger point" according to the Guardian newspaper.[370]

This case illustrates the problems that regulators have in understanding digital multi-sided markets with personal information.

5.4 Consumer protection

A "high level of consumer protection" is enshrined in article 38 of the EU Charter of Fundamental Rights,[371] while the Treaty on the Functioning of the European Union (TFEU) provides further details, including a right to information and to form consumer organisations.[372]

These protections are justified on the basis that promoting consumers' welfare - transparency, choice, fairness, quality, safety, etc. - is necessary to maintain confidence in the markets and helps promote competition.[373] In addition there is an imperative to protect consumers from risks. These risks are traditionally seen in relation to physical health and safety, but this protection is also extended to potential harms caused by abuses of personal information. As explained by the FRA "the concern for product safety, meanwhile, complements both the concept of the exploitation in competition law and the stress in the proposed General Data Protection Regulation on impact assessment, and subsequent discussions on a progressive risk-based approach and on the principle of accountability."[374]

Another area of overlap with Data Protection is the obligations of fairness and provision of accurate information in consumer contracts. Choice and transparency are fundamental tenets of consumer protection, and also rights under the Data Protection Directive.

Terms and conditions (T&Cs) for digital services and goods are widely seen as problematic[375], particularly in relation to the use of Technological Protection Measures to control intellectual property. But most T&Cs will in many cases also contain the privacy policies and form the basis for consent to the use of data. This increasingly includes agreeing to the monitoring of consumption habits. The Electronic Frontier Foundation has showed the extent of this consumer monitoring in the e-book market, but other media has similar issues.[376]

As we discuss elsewhere in this report, there are growing concerns about the viability of the consent model in this context of data protection. In most cases users of digital services are not able to negotiate contracts or receive alternative services. And this lack of transparency and choice also

clashes with consumer protections. One particularly thorny issue from the point of view of consumer rights is the definition of "free" digital services. In many cases, the service requires the user to provide information with Facebook and Gmail being some of the best-known examples. Until now these services have not been challenged to clarify the quid pro quo, but there are several regulations that could potentially make them do so.

The 1993 Directive on Unfair Contract Terms[377] provides some limited protection and expects terms to be drafted in plain language, with any doubt about the meaning of a term to be interpreted in favour of the consumer. The Unfair Commercial Practices Directive[378] tackles misleading descriptions, including describing a product as 'free' or 'without charge' when the consumer has to pay anything other than delivery or other basic costs. The Consumer Rights Directive[379] defines a new digital content category that is distinct from other goods and services and includes some new obligations. These include informing customers of hidden costs, but also of any incompatibility, including technical protections. Unfortunately the directive did not update unfairness and contracts to the digital age.

In summary, consumer laws could play an important role in the regulation of personal data. But these laws need to be updated to the digital age and get stronger enforcement mechanisms. There is a clear need for more clarity in contracts for online services, and consumer legislation could spearhead this change.

5.5 Other Regulations affecting Digital Identities

5.5.1 Public Sector Information

The European Commission has a large program to promote open data in Europe.[380] This will have an impact on digital identities in several ways. For example, more public registers containing personal information may become open data with fewer restrictions on reuse. This could increase the ability of organisations to build profiles of EU citizens. The main piece of legislation in this package is European Directive on the Reuse of Public Sector Information (PSI Directive).[381]

5.5.2 Open standards

Although there is no binding European legislation on open standards, the EU has pushed for it as early as 1999, arguing that interoperability was a key requirement for the implementation of eGovernment across the European Union.[382] The Digital Agenda, which is the Commission's plan for the next years in order to create a Digital Single Market, includes a guide that calls for "the use of same standards and technical specifications".[383]

5.5.3 Intellectual Property and the Database directive

Copyright and intellectual property are important in anything to do with the digital world, as copying is involved at every stage. Everything we put online, our tweets, Facebook status, blogs, etc. is subjected to copyright, as is also is any material we may incorporate into our own. Copyright in the EU is mainly government by the Directive on copyright and related rights in the information society from 2001,[384] which is currently under review.

Copyright does not protect simple "facts" anywhere in the world, so for example much of the data produced by sensors would not be copyrighted. But in the EU, databases of materials that in

themselves would not be protected by copyright are regulated by a special "database right" under the Database Directive.[385]

5.5.4 E-signatures directive

A 1999 European Union directive[386] gave electronic signature the same legal weight as the hand-written one; provided that they can give enough evidence that they indeed belong to the persons that claim to use them. In order to avoid fragmentation of the common market, later communications[387] by the European Commission encouraged the Member States to implement mutually recognised and interoperable electronic signatures.

5.5.5 E-identity

Government electronic identification systems have been developed in several European countries such as Italy, Germany or the Netherlands.The European directive on electronic communication[388] of July 2014 didn't aim at making eID mandatory (as the issue of having even a paper ID remains controversial in some Member States), but rather wants to greatly increase the mutual recognition of eID between countries, in order to facilitate cross-border business as well as international administrative tasks for citizens. Though harmonisation is the goal, it is not equivalent to a European eID, or to a European centralising of Member States' eID information.

5.5.6 Security, surveillance and data retention

After the London bombings of 2005, the idea of a unified framework for data retention in the European Union became reality after several years of being pushed for by various countries. The controversial directive[389] was chaotically transposed into local laws, as several countries failed to transpose it before the deadline, and its implementation was annulled in other countries, such as Germany, on privacy considerations.

The directive was declared invalid[390] in its entirety by the European Court of Justice in May 2014, which made a sharp criticism of bulk data retention as not complying to the principle of proportionality regarding its aim (national security), and for the insufficiency of the safeguards. European countries are now far from harmonized, with countries that have seen their data retention law declared void, countries trying to take into account the judgment while keeping their laws, like Luxembourg, and countries going arguably even further in data retention, for instance the United Kingdom[391] or France.[392]

5.6.7 Financial information

Financial data are not considered particular or sensitive data in the Data Protection Directive,[393] and as such only the general rules apply. However, the European Union specifically addresses certain issues such as cross-border payments via baking transfers,[394] which leads to data flows of personal information. The European Data Protection Supervisor has issued in 2015 Guidelines on data protection in EU financial services regulation,[395] to ensure that the right to privacy and protection is well implemented with regards to these sensitive and valuable data.

The European Union is also part of the United States' Terrorist Finance Tracking Programme[396], which allows United States' authority access to a database of financial information or the prevention,

investigation, detection, and prosecution of conduct pertaining to terrorism or terrorist financing. The European Parliament called for the end of this agreement after it has been revealed that the United States, and notably the National Security Agency, collected millions of citizens' personal data in this database, bypassing the safeguards.

5.6 Some key issues with EU Data Protection

The following are some of the areas where the new Regulation may introduce changes. Given the current state of the legislative process we must make it clear that the outcomes are not decided. For each topic, we just present the issue, the proposed changes in the draft Regulation, and any amendments introduced that may limit the effect of the original proposals.

The Regulation is a very large piece of legislation, with currently three versions under dispute totalling 630 pages[397], and it would be impossible to provide a comprehensive summary in this report. The process for approval may take at least until 2016, but it could be longer, or it may even fail at some stage.[398]

For the sake of brevity, in these sections we will refer to the Commission as EC, the European Council as Council and the Parliament as EP.

5.6.1 Personal Data, anonymity and pseudonymous data

The very definition of personal data would appear to be a simple matter, but instead has become one of the most hotly disputed issues in this field,[399] with critical implications for emerging areas such as online behavioural advertising and big data. Much of the debate hinges on the effectiveness of techniques to de-identify personal data, and thus the likelihood that it can be linked to individuals.

In the current Directive, defining de-identified data is left to Recital 26, which states that data shall be considered anonymous if the person to whom the original data referred to cannot be identified by the controller, or by any other person, by any means reasonably likely to be used. But this has not been implemented in the national legislation of many member states, including the UK, leading to a very confusing landscape.[400]

An example of the confusion is whether Internet Protocol (IP) addresses used to identify a device connected to the internet at a particular time are personal data or not. Internet Service Providers (ISPs) can link an IP to a customer, and some countries such as Germany consider IPs personal information. But in other countries such as the UK, it is assumed that other people would not have the "means" to link the IP to a real person, so IPs are seen not fully as personal information, but at best as "pseudonymous".

The EP introduced the definition of pseudonymous data in the Regulation[401], as '*personal data that cannot be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution.*' This would create a third category of data that would fall under the Regulation, but with lesser protections. Privacy groups and watchdogs, also concerned about proposals to allow the creation of pseudonymous profiling as a legitimate interest, have opposed this lower protection.[402]

According to privacy expert Caspar Bowden, pseudonymous data *"is a bit like saying your car plate number is not fully personal data because you need to ask the authorities to do the matching. So tracking your car's movements is ok until I find out your name".*[403] The reference to tracking is very apt, as for example IPs are used by internet marketers to track online behaviour.

Advances in computer science have thrown into question even processes where the data is meant to be completely anonymised, and whether such data should fall out of scope of data protection[404] without other considerations. There is still hope the techniques may work in most cases, but many regulators acknowledge the bar is a lot higher than expected[405] and absolute certainty of anonymisation may be impossible to guarantee

This is huge for big data, which relies on massive amounts of data that supposedly has been anonymised, or at least pseudonymised.

5.6.2 Consent

Consent is one of the main lawful avenues in the processing of data. In the current Directive and all version of the Regulation, consent must be "freely given, specific and informed", but the EC and EP want the Regulation to be stronger and ask for "explicit consent"[406], evidenced by "a statement or by a clear affirmative action". Explicit consent currently applies to the processing of sensitive data, such as sexual orientation[407]. The EP would also want to see consent tied to a specific purpose and nothing else. After pressure from national governments[408] such as the UK, the Council instead has settled on proposing the keep a weaker definition of "unambiguous consent"[409]. This would allow technology companies, for example, to consider default settings in programmes as a valid form of consent, while "explicit consent" would force them to present users with a choice. Removing implied consent has been criticised for likely leading to an endless and onerous questioning of consumers. Any changes to the definition of consent will need to be carried on to other legislation that relies on consent, such as the E-Privacy Directive we discussed in a previous section.

In addition, there are some disagreements over the role of the Regulation to balance the power of users and companies. The EC has proposed that consent is declared invalid if there is a 'significant imbalance' between the data subject and the data controller, and the EP wants to make invalid contract terms requiring agreeing with uses of data which are unnecessary for supplying a service.[410] But the Council rejects entirely the idea that the Regulation should be so clearly sided with citizens against businesses.

5.6.3 Legitimate interests

Under the current Directive, companies can process data without consent. In many cases this will allowed if it is necessary to perform a specific function or contract, or in an emergency, but there is also a very pragmatic provision that allows the "legitimate interest" of the organisation processing the data (or third parties) to override the privacy of individuals. This is a very confusing part of the legislation, as it appears to contradict the very idea of data protection, but there are some limits to what companies can do.

The purposes for which the information is used must be clearly defined (so called "purpose limitation") and there should be a balancing exercise that ensures there is not an excessive intrusion on individuals' rights and freedoms.[411] As explained by civil rights group EDRI,[412] this means, for example, that if you give your data to a supermarket for your loyalty card, they can use this information for relevant and related purposes. But they cannot sell your data to a health insurance company that, again as an example, will profile you as potentially unhealthy based on your food-buying habits. In short, data may only be processed when it is not excessive and is done for explicit and legitimate purposes.

The EP has proposed to further narrow down the legitimate interests to those matching the "reasonable expectations" of the persons whose data is processed. In contrast, the Council has proposed to weaken the purpose limitation to allow for new purposes, and for the data to be passed on to third parties who could then use it for their legitimate purposes. This would severely weaken the Regulation. Again in EDRI's words:[413] *If a company you have never heard of can process your data for reasons you've never heard of, what is the point in having data protection legislation?*

5.6.4 Transparency and Privacy Policies

As we saw in the section 5.2 transparency is one of the most important aspects of data protection in its current form. Privacy policies are generally long and hard to understand, and in some case information ostensibly given in one context can end up being used for very different things that can only be found in very small print. To give a scale of the problem, researchers have found over 60 independent projects attempting to simplify policies, terms and conditions in order to improve privacy protections.[414]

The Regulation would strengthen the rights of citizens by forcing companies to disclose more information on how their data are processed or if the provider has transferred data to public authorities or intelligence services. Data controllers will have to explain which user data they process for what purpose.[415] There are additional requirements for language to be simplified, and the EP has even proposed that standardised icons should replace long pages of legalistic language in privacy policies.

5.6.5 Rectification, portability and erasure

The current Directive gives citizens certain rights to control the information held on them. This includes a right to obtain a copy and to ensure that the information is relevant, correct and up to date. The Regulation introduces stronger provisions in these areas. There is a right to rectification, in article 16 which the Council wants to water down by allowing supplementary notices instead of corrections.[416] Accessing your own data will no longer incur a deterrent fee.

Most controversially a new right to erasure[417] and "to be forgotten" - although this part was removed in the EP proposal - has been introduced that would allow citizens to demand the deletion of data when it is no longer necessary for the original purpose, or under certain circumstances. These include withdrawal of consent and objecting to processing under different provisions. This has caused consternation among industry organisations, among other reasons because it brings an obligation on the controller who has made data available to the public or under licence to chase third parties so they also delete it. But this right is seen as fundamental to ensure people can control the trail of data they leave behind in the digital world. As we saw in the famous case of Google,[418] a right to erasure already existed under the current legislation, albeit not in such explicit terms. The Regulation also makes explicit the need to balance this right with freedom of expression, the public interest and historical, social and scientific research.

While most attention has focused on the potential of the right to erasure to raise such cases of conflicting public information, the impact could be felt mainly by companies who keep private data and profiles of customers long after they leave their services.

But for citizens to have control over their data, deletion and access are not enough, so the EC introduced a new right to portability[419] that allows citizens to obtain copies of data held on them in

electronic format. The EP removed this right, but the Council has introduced a modified version. This right has been opposed by many national governments concerned about the impacts[420] on businesses. Countries such as the UK promote a very limited version of this "right" under consumer initiatives, such as Midata,[421] but these are framed under consumer choice, and tend to be limited to some sets of data, such as consumption records, bank statements, etc. So it would be unfair to compare them to a general right to portability. Authorities for a variety of reasons can restrict these rights[422], but the Council has introduced some fairly broad clauses relating to public registers, social protection and public health; and some very narrow exceptions for archives of former totalitarian states to keep records of "political behaviour".

5.6.6 Profiling

Most modern organisations strive to use data to better tailor their services. From health and security to financial credit and advertising individual pictures of users are created through the collection and analysis of their behavioural data. This profiling can have positive or negative consequences, possibly at the same time, e.g. targeted adverts may be more relevant but also creepy. The negative consequences of profiling can be life changing, including the denial of medical care or credit for a home.

The current directive does not refer directly to profiling, but instead refers to 'automated individual decisions'. Article 15 says that individuals have a general right "not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him".

There are some exceptions to this, such as "performance or entering a contract", but generally there is a right to ask for a human review of any decision made by a computer on things serious enough to be considered legal effects (e.g. employment or credit).

The new Regulation[423] makes this more explicit by calling it profiling, and broadens the scope from "legal effects" to decisions that "significally affect" those profiled by businesses.[424] In addition, profiling based on sensitive data - sexual orientation, trade union membership, etc. - is now prohibited, while marketers must give people an opt out at any time. Privacy experts Hunton & Williams advise firms that require profiling to start thinking how they will manage proper consent.[425] Other analysts in contrast believe that the requirements to show profiling has significantly affected people will such a high bar than in practice they may amount to a continuation of business as usual.[426]

Civil rights groups are concerned about the weakening of the obligations on data controllers to provide meaningful information about the logic behind the profiling in the versions forth Council and EP. These groups are also worried about the re-insertion by the Council (after being deleted in the EP version) of profiling as a possible exception to the rules that could be implemented in Member State law. Governments can claim national security, defence, public security and even "other important objectives of general public interest" to profile citizens.[427]

5.6.7 Data Protection by design and the Risk based approach

Privacy by Design is one of the most important concepts developed in the past two decades in the field of privacy. We discuss it in more detail in section 6. The draft regulation introduces the principle of Data Protection by design and default,[428] adapting the above concept to mean that companies should be adopting a proactive approach where technical and organisational measures are

taken to ensure compliance with the requirements of the Regulation. This includes designing services to require a minimum of data for providing a service, and with privacy friendly default settings. But this positive principle is being undermined by the Council, which has introduced severe qualification on the principle, following a so-called "risk based approach" that allows companies to take "into account of the nature, scope, context and purposes of the processing as well as the likelihood and severity of the risk for rights and freedoms of individuals posed by the processing".[429] According to civil rights groups, this undermines the essence of data protection by default, and gives the companies an unchecked right to decide whether or not to comply with obligations that would provide citizens with high standards of data protection.[430]

5.6.8 International data flows

European data protection laws restrict the transfer of data to countries without adequate protections. This is a major issue for international trade and subjected to much discussion, mainly with the US government and American businesses. But it is also important for organisations building decentralised social platforms. If there are no national protections, organisations must rely on other legal binding agreements to transfer data, such as: binding corporate rules ("BCRs"); standard data protection clauses adopted by the Commission; standard data protection clauses adopted by a regulator; and contractual clauses authorised by a regulator.[431]

It is beyond the scope of this paper to provide more details of this highly technical area, but it would be advisable for any implementation of the DCENT platform that involves sending personal data outside the EU to seek specialist advice once the Regulation is approved. The EP is trying to also enforce controls over the handling of data to third party governments, including for security and surveillance. An article with such controls was already contained in a first draft of the Commission's proposal, but deleted after intensive lobbying of the American government. It was put back by the Parliament after the Snowden revelations.[432] The Council wants to remove this article.

5.6.9 Jurisdiction and one stop shop

Some of the most far-reaching proposals in the new Regulation relate to the procedures and formal structures of data protection, as part of the harmonisation drive. Key aspects here are the introduction of one stop shop concept and a beefed up role for a European Data Protection Board.

The 'one-stop-shop' approach means that citizens can go to their national data protection authority for complaints that cover data abuse anywhere in the EU. Conversely companies will only have to deal with the authority in the country of their main establishment. There is basic agreement among all parties on the principle, but the Council would like this system to only apply in "important cross-border cases"[433]. The Council has also introduced very complex bureaucratic procedures that may completely undermine the concept.[434] As part of this harmonisation, a European Data Protection Board, composed of national data protection authorities, would be the arbiter with the capacity to make binding decisions for cases of Europe-wide relevance.[435]

According to Jan Albrecht, the MEP who led the drafting of the EP version, these changes will stop the "'race to the bottom' in EU member states with weak law enforcement".[436]

5.6.10 Enforcement

Adequate enforcement makes all the difference for citizens, as many countries around the world have decent data protection on paper but no effective ways to make organisations comply. The current enforcement regime is fairly weak, but the Regulation could make a big difference, although this is one of the areas that have suffered the heaviest corporate counter lobbying. Strong proportionate sanctions are needed to focus corporate priorities. The EC and the Council propose fines of up to 2% of global turnover for severe cases of illegal data processing, but the EP has raised this to 5% up to €100M. This seems a lot, but for example in the case of anti-competitive practices, those guilty may be liable to a fine of up to 10% of their total group turnover in the preceding business year.[437]

Financial compensation to those directly affected is still secondary to fines. But the proposed draft introduced stronger options for the collective defence of citizens' rights. There is a new right for public interest organisations to act on behalf of citizens and consumers by means of collective complaint actions.[438] But the Council wants to preclude these class action suits, so organisations could no longer be mandated by more than one citizen to complain on their behalf, or possibly take collective complaints in their own name. In addition the Council wants to restrict any action by public interest groups only to Data Protection Authorities, not courts[439].

5.6.11 Research and science

The scientific research community and related industries, particularly in the life sciences[440], have been some of the most active groups lobbying around the Regulation. Researchers believe that some of the provisions in the new law would make their job unviable, going as far as claiming that the Regulation "could make cancer research impossible'.[441] The problem centralises on the stronger requirements for consent to be related to a particular purpose, which in their view would not allow the reuse of databases for many different research queries.[442]The EP introduced exemptions to consent when "the processing of medical data is exclusively intended for public health purposes of scientific "[…] research (that) serves a high public interest, if that research cannot possibly be carried out otherwise". There are also requirements to apply de-identification techniques. This has been rejected by the sector, which got the Council to introduce stronger exemptions[443], including a consideration of broad consent in a recital.

Privacy organisations and also groups concerned about corporate power in the health sector have strongly opposed weakening the exemptions[444], which they see as going back on current protections, for example in allowing the sharing of pseudonymised health data with any company, including Google, without consent. In their view, the issue is not about data for saving humanity or curing cancer, but simply about the corporate exploitation of sensitive personal information by big businesses, including some that happen to make their money by selling medicines. Their opponents claim that nowadays vital research is carried out everywhere and it is impossible to separate big business from public interest. This is a really difficult issue as clearly both sides have a point, and it is quite unfortunate that public interest organisations working on different aspects - privacy and health - have ended up in such an entrenched conflict.

6. Economic, policy, and technical alternatives for identity

6.1 Economic strategies

6.1.2 The economics and ethics of technology

The economics of technology is a broad field that can include studies on how firms use existing and future technologies, the consequences of government intervention and regulation in technological change and technological proliferation, the implications of technological innovation for the welfare (economic and other) of different social groups or the efficiency of government subsidies to promote technological innovation, among others.

However, when dealing with the study of the economic impact of technology, it is common to find a discourse that describes technological developments in terms of novelty and progress. 'It is common for new technologies to be hailed as signalling a fundamental change in the way we live and communicate, and as having the ability to efficiently solve problems that up until the technology's arrival had not yet been identified as problems'[445] This uncritical belief in the abilities of engineering or technological solutions to solve social problems is referred to as the **'technological fix'**, and is exposed every time the solution to a social problem is limited to the possibility of buying or developing some technical solution.[446]

A few recent examples are a case in point here. In the case of **body scanners** in airports, a development that caused significant amounts of controversy after their introduction in 2012, specifically due to their perceived intrusiveness and impact on people's dignity and privacy, Hallinan and Friedewald[447] review existing figures and conclude that at the EU level the evaluation of the costs of such systems only include 'direct and identifiable costs of deployment'.[448] In the US, the Transport Security Administration (TSA) 'has not conducted a cost analysis at all, despite specific observation from the Government Accountability Office'.[449] The authors mention an independent study that, taking into account the indirect costs of body-scanner deployment, the economic implications of the perception and feeling toward body scanners, the potential economic impact of a terrorist attack and the reduction in risk due to the application of body scanners as a security measure, concludes that body scanners would need to disrupt at least one US-originating attack every two years to justify their cost.[450]

Smart metering systems currently being deployed in the EU are another example that has caused controversy around privacy issues, so much so that the European Data Protection Supervisor (EDPS) warned that while 'smart metering systems may bring significant benefits, it will also enable massive collection of personal data which can track what members of a household do within the privacy of their own homes', and urged the CE to 'prepare a template for a data protection impact assessment' and 'assess whether further legislative action is necessary at EU level to ensure adequate protection of personal data for the roll-out of smart metering systems'.[451] While the EC's Directorate-General for Energy (DG ENER) has announced that a template for a Data Protection Impact Assessment (DPIA) and guidelines on Cost-Benefit Analysis methodology, benefits and costs will be developed, these have not yet been made public.

As these examples show, there is an emerging consensus around the need to develop and improve the methodologies used to assess the costs (economic, ethical and social) of new technological

developments, especially when they collect personal data and impact on privacy and fundamental values. However, an 'economics of data-intensive technologies' is difficult to develop independently of the goals and context of each specific project or initiative, and an understanding of the economic impact of the identity market will necessarily have to learn from methodologies and approaches developed for other related fields. The lack of specific methodologies and guidelines, as well as the difficulties intrinsic to the evaluation of the monetary cost of intangible goods, are now delaying the process of impact assessment becoming a necessary step to assess whether a new technology is indeed useful, necessary and socially desirable.

Therefore, current assessments of the economic impact of large-scale technological projects not only tend to use very abstract figures and methodologies, but also fail to take into account the opportunity costs of technological investment –as Graham emphasizes, investment in technological solutions is unquestionable, even when it is done at the expense of social investment.[452] This broad-spread technological determinism is probably one of the reasons why promoters of technological expenditure have been able to justify large and costly projects without providing investors with detailed analysis of the costs, benefits, impact and alternatives. As the above-mentioned case of the US Transport Security Administration (TSA) shows, the belief that technology is a superior solution to any problem translates into large investments being made without the necessary precaution in the management of financial resources.

In 'The Limits of a Technological Fix to Knowledge Management', for instance, Currie and Kerri tell the case of the CEO of a pharmaceutical company who decided to invest in knowledge management software. They quote one of the employees saying 'He intuitively believes there is value in it. Reflecting this, unlike most other things, he hasn't asked for its value to be proved'.[453] Any investment in technology is thus seen as a good investment, regardless of its cost or impact, which are never evaluated as value is taken from granted. This has a deep effect on the economics of technology, as rational assessments and decision-making processes are clouded by assumptions and beliefs that broaden the gap between problems and solutions.

6.1.3 Ways to measure the value of privacy

The value of personal data continues to be an under-researched field. Most current insights come from economists and scholars working on the economics of privacy and studying the economic cost-benefit trade-offs individuals undertake when disclosing personal data in economic transactions and the competitive implications of the protection of personal data for service providers. Most of the existing literature is based on surveys that explore the social exchange aspect, economic experiments implementing real transactions are still scarce.[454] Moreover, most existing research uses behaviour-based pricing and product personalisation as a way to determine what value can be assigned to the customer's privacy concern.

In order to assign a monetary value to the right to privacy and its infringement, most studies use two alternative methodologies –the *willingness to pay/accept* and the *cost of corrective measures*.

Willingness to pay and willingness to accept

According to economic theory, in a perfectly competitive market the price of a commodity or good reflects the value that consumers are willing to pay for it. In the case of privacy, which is a good that has no market price, the willingness to pay can be a good way to monetise its value. In order to calculate this willingness to pay, several methodologies can be used, such as *contingent valuation*,

based on distributing surveys among users and exploring different alternatives.[455] These surveys can attempt to find out how much users/consumers are *willing to accept* as compensation for their loss of quality of life –in this case, loss of privacy.

There is an extensive literature on calculations made using surveys in matters related to environmental damage. Valuation methodologies of this kind were used in the case against Exxon, to calculate how much the corporation should compensate those affected by the Exxon Valdez spill in Alaska in 1989.[456] However, there are no similar precedents in the field of privacy, and so alternatives have to be found. A good option is the use of controlled experiments, a tool that is being increasingly used in the valuation of intangible goods. When these are carried out in adequate conditions, the results are consistent.[457] Two of such experiments are worth mentioning, as they provide some useful reference values. One involves giving away a gift card to use in a specific shop. The card has a specific value if the carrier choses to use it anonymously, but if he or she accepts to provide their personal data, this value is increased.[458] With this controlled experiment, it is possible to arrive to a figure representing the percentage of people that prefer to remain anonymous and how much they are willing to pay for their anonymity. A 2009 experiment along these lines found that the value of privacy represents between 1.3 and 8.7% of the value of the product that is obtained in exchange for one's personal data, depending on what alternatives are provided. Overall, the authors decided to set that value of privacy at 5.8% of the product price.

The same authors studied the relationship between 'willingness to pay' and 'willingness to accept', While the willingness to pay is the maximum amount a person is willing to pay, lose or exchange in order to receive a good, the willingness to accept represents the minimum amount an individual is willing to receive to give up a good (in this case, privacy). They conclude, as most of the literature shows,[459] that in the 'willingness to accept' scenario the value of privacy increases by 70%, and therefore the value of privacy is set at 10% of a product's price. Another useful study is 'Data Users versus Data Subjects. Are Consumers Willing to Pay for Property Rights to Personal Information', Rose explores different price ranges and different options to explore how much would people be willing to pay for privacy-enhancing systems and concludes by setting the price of privacy at 12% of the product's retail value.[460]

Cost of corrective measures

An alternative to the methodology just described is to start not from the consumer's willingness to pay or accept, but from the principle of interchangeability. This method calculates the costs of the measures that need to be implemented in order to reduce the impact of a negative externality. In the case of a specific service such as data brokerage, for instance, the cost of corrective measures approach would take into account the cost of developing and implementing software to anonymise personal identities.

The bright side is that this method is relatively easy to implement, as these costs are usually easy to calculate. On the dark side, this methodology does not address the origin of the problem, and it is not always guaranteed that these corrective measures will guarantee the privacy of the citizen/user, as only the most problematic aspects are usually addressed using corrective measures. In the experiments carried out suing this calculation, the cost of privacy is considerably lower than the figures found using the previous alternative.

Another methodology worth mentioning, which is less centred on the monetary value of the fundamental right to privacy but broader in scope are **Privacy Impact Assessments (PIAs),** which have proliferated in the last few years in countries such as the US, the UK and Australia.

PIAs[461] assess whether and to what extent privacy is affected by a specific initiative, and identifies whether the necessary steps have been taken in order to comply with the legal framework in terms of gaining consent from the data subject, establishing who will collect the data, for what purpose and who will have access to it. They are 'a way to detect potential privacy problems, take precautions and build tailored safeguards before, not after, the organisation makes heavy investments. The costs of fixing a project (using the term in its widest sense) at the planning stage will be a fraction of those incurred later on. If the privacy impacts are unacceptable, and corrective measures are not developed or fail to do away with the privacy infringement, the project or initiative may even have to be cancelled altogether. Thus, a PIA helps reduce costs in management time, legal expenses and potential media or public concern by considering privacy issues early. It helps an organisation to avoid costly or embarrassing privacy mistakes.'[462]

PIAs are therefore, an "early warning system",[463] useful mainly in terms of ensuring legal compliance. While costs are mentioned in some PIA methodologies, the possibility of assigning a monetary value to a privacy infringement is not explored, and the theoretical savings in efficiency and early detection of potential failures are a common-sense assumption and not the product of a privacy calculus. PIAs may thus contribute to a better understanding of the privacy implications of a technology or surveillance initiative, but they are not *directly* useful for an economic assessment.

More recently some authors have explored a broader version of Privacy Impact Assessments – Surveillance Impact Assessments (SIAs). The main differences between the two are mainly that SIAs have a wider focus (they address the impact of surveillance systems and projects not only on privacy but also on other social, economic, financial, political, legal, ethical and psychological issues), are principally focused on groups or society as a whole and engage a wider range of stakeholders.[464]

PIAs and SIAs are assessment methodologies that take into account costs and benefits, but are broader than a financial assessment or a cost-benefit analysis. They should however include specific methodologies to assign monetary value to the infringement of fundamental rights and values in order to become a useful tool to assess the economics of technology and the cost of personal data and identities.

Structural transaction costs

In the specific context of the US, where constitutional protections are a balancing act and judges need to assess whether (privacy) rights have been infringed upon and make adjustments to ensure that people continue to enjoy them, some authors rely on Surden's 'structural privacy rights' proposal.[465] His contention is that the costs of data-intensive technologies (physical, technological and otherwise) act as non-legal structural regulators of technology proliferation. However, if these costs are lowered, this can impact on the structural regulators and quickly alter the playing field in terms of incentives. According to this theory, the relevant actors should recognise this and make the necessary adjustments by developing new legal protections, creating a situation in which new legal costs compensate for the diminishing financial costs.

On the basis of Surden's theory, Bankston and Soltani[466] develop a case study on the diminishing economic costs of technologies to collect information on citizens (GPS, mobile tracking, IMSI catchers) to contribute to better decision-making. Their premise is that on the face of ever-decreasing economic costs, and therefore lower structural costs, there is a need for increased legal protection of privacy rights.

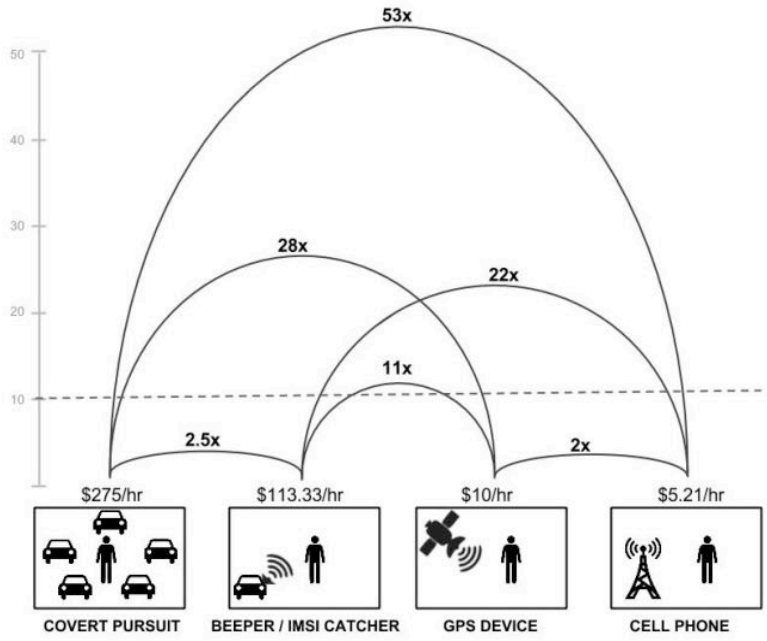

Figure 13. Hourly costs of different location tracking techniques. Source: Bankston and Soltani.

6.1.4 A New Deal on Data

MIT professor Sandy Pentland, co-founder of the ID^3 project, has proposed that we need a complete rethink of how personal data is used, what he terms a 'New Deal on Data". Pentland has been working for over a decade on what he labels "reality mining": how to use the data we generate though mobile phones to better understand human behaviour and interactions. In a pioneering paper from 2004 his team mapped the social networks and behaviours of 100 mobile users using call logs and bluetooth proximity data.[467]

The growing ability to monitor and predict human behaviour from mobile data has the potential to transform public policy, governance and management. But there is one big stumbling block around the privacy of end users and data ownership. Pentland proposes a new deal that will help "create value for the producers and owners of the data while at the same time protecting the public good".[468]

His concern is avoiding a situation where either the data is held by companies and not used for what he calls the "Common Good', or governments develop big brother monitoring systems inaccessible to the public. He calls for "workable guarantees that the data needed for public goods are readily available".[469]

Pentland's new deal would be based on regulation supporting an open information market approach to data that would allow people to give their data in exchange for monetary rewards. But as a first step, this requires that people own their data. The other tenet of the new deal would be to implement default data aggregation or anonymisation, as this would be enough to gain valuable insights. Transparency is a prerequisite, important to understand who holds the whole picture of the individual.[470]

These proposals have been widely discussed in business circles, as Peatland has a very high profile and is connected to institutions such as the World Economic Forum. Some data brokers such as

Acxiom support Pentland's market approach, where consumers can decide, as they believe that thesis more effective.[471]

Giving people ownership of their data is more complicated than it appears. Pentland's idea of what ownership would mean in practice - possess, control, dispose - looks quite close to some of the proposals in the new EU Data Protection Regulation: a right to portability and erasure. He also proposes an opt-in by default approach that chimes with the E-Privacy Directive.

Anonymisation is also a problematic issue, with raging debates as to whether it can be fully effective at the individual level. After a string of high profile cases where individuals were identified in anonymised datasets, there has been a move to acknowledge the limits of these techniques.[472] But in response, advocates of anonymisation argue that those cases are exceptions built into a "myth", the techniques work well on an everyday basis, and we would be hampering the potential of big data if we stopped their use.[473] Still experts in the field caution that anonymisation is not a "silver bullet".[474]

In any case, it is unclear how to establish the ownership of anonymised data, as in most cases it is not recognised as personal information and it would not be possible to link it to an individual.

More generally, it is also unclear how a market approach would deliver the public good, when markets consistently fail at this in other domains. The imperative to make data available would appear some form of controlled expropriation. In the case of built infrastructure the state has a clear role in taking over land, but with data would private companies also have a right to demand our data for the public good?

6.1.5 Democratising Monetisation

A quick search on Google shows that most of the discussions about the monetisation of personal data centre on developing the ability of firms to generate income from the data they hold on their customers or data acquired by other means. The ability of individuals to generate income from their own data is a concern for a minority of organisations that are part of the movement towards user-centric personal data ecosystems.

There is no fully tried and tested model for the monetisation of personal data by individuals but many different experiments are currently taking place. Some projects, including some personal data stores and data cooperatives,[475] focus on controlling access to personal data by third party companies, with the benefit for individuals coming in mainly through enhanced privacy and to a lesser extent money generation.

Other projects try to enhance the capacity of individuals as consumers to get better pries or quality. Collaborative shopping or crowdfunding platforms and so-called "intentcasting" tools[476] - which allow prospective buyers call for offers in their own terms - in principle could reverse the situation where commerce platforms learn everything about their customers for better or worse.

The straightforward sale of data is common with companies and even public bodies. The UK driving license authority made £25m in five years by selling personal details of millions of motorists to parking enforcement firms.[477] But it is a lot less clear for individuals. Despite some high profile stunts, such as Dutch student Shawn Buckles selling his data soul for 350 €,[478] this area remain highly speculative. Projects such as Datacoup[479] and Citizenme[480] have different models with one getting buyers to pay and the other one taking a cut from sellers. Both have started to build sophisticated technological platforms, but apparently there are no actual buyers for data so far.

The company Handshake takes a different approach, focusing on market research[481], promising to "turn what has previously been stolen into a currency which can be traded"[482]. Data is already used as a form of currency by companies. For example traffic app Waze expanded into Latin American swapping data generated by its customers while using the service in exchange for high quality maps.[483]

6.1.6 Consent and licensing

We have discussed strategies based on the premise that individuals should be able to own or control their data and extract value from it. But as we saw in the previous section, selling personal data is not simply selling the information itself. The control required for the monetisation of data implies giving access to the data while establishing certain conditions for what can and cannot be done with it, who else may have access, etc. These conditions are normally expressed in consent agreements and privacy policies, which are established by the *data controller* with little room for debate or input from the *data subject*.

There is plenty of room for improvement at the policy level, increasing transparency and choice for individuals, and we discuss this in section 6.2.5. But if we are to reverse the data ecosystem to out individuals at the centre, turning it into a sellers' market, consent would also need to change to become part of the supply side of data. This is particularly the case when the data is directly generated through wearable sensors or other systems that do not require a platform whose owners could claim a stake in the data.

If the current system were to be turned upside down and individuals truly owned and controlled their data, the conditions for the processing of data set out by the individual would constitute the basis for an organisation to accept to engage and use the data. This would not be exactly a privacy policy or a consent form, but a different kind of contract or license to use the data. As Mireille HIndebrandt has put it, we need to move "from consent to legality and mutual empowerment".[484] Importantly, these arrangements would need to apply to data that has been de-identified and may not be covered by privacy policies.

From the individual's point of view, making information publicly available tends to be perceived as giving up any claims to further control, even if this is not strictly true at least in Europe.[485] Control would mean being able to define specific purposes and uses of data, e.g. banning military uses. As we saw in section 5.5, in the EU there is a drive towards explicit and specific consent, where organisations must clearly explain the purposes for which each kind of information will be used. The challenges of being able to define flexible and broad purposes of data uses while being specific enough are a major concern for researchers. But this is an issue for any innovative use of data, even in a model where data subjects can define the purposes. In relation to monetisation, the conditions may need to be more specific about the processing of the data than purely from a privacy point of view.

Given the complexity of data ecosystems, user defined conditions would need to control further transfers of data to third parties. In order to be effective these should also apply to re-users of data. Currently this does not involve the individual originating the data, although in the EU there should be some continuity in the purposes for which the new organisation will reuse the data, which should be consistent with the original privacy policy.[486] But what about inferred data? Should individuals retain some control?

Pentland and his colleagues working on the New Deal on Data acknowledge that in the complex data processing we see today, granular user control downstream is hard to achieve. They propose a combination of business practices, legal rules and technical solutions around the concept of "living informed consent"[487] This starts with transparency over what data is held by whom, and the ability to authorise any sharing while understanding the implications. Pentland is also involved in developing the Open Mustard Seed (OMS) trust framework, which we discuss in section 6.2.6. OMS manages user preferences for data sharing was part of *manifests* that captures the "operating rules" for a community of trust.[488]

Changes in circumstances present another challenge to models of control and consent. As we saw in section 5, this is a serious problem in relation to research, as scientists wish to repurpose the data they hold for different projects. The need for more dynamic consent models has been explored by various projects, such as the EU funded EnCoRE collaboration between industry and academia. The project built tools to allow individuals to change their consent preferences over time. This included sophisticated technological measures such as cryptographic 'sticky policies' that helped ensure that these consent preferences remained associated with the data they referred to.[489]

As discussed elsewhere, portable consent is an impotent issue in health and research, and here is where we have seen the most innovative developments. John Wilbanks from Creative Commons advocates a "portable consent". This is based on the "*open consent"* approach that prioritises data sharing over control in order to build a data commons for research, and first developed by the Personal Genome Project at Harvard Medical School.[490] Wilbanks - through Sage Bionetworks - is developing a tool called Portable Legal Consent (PLC) for anyone who would like to donate their health data for research purposes with little restrictions; and who are prepared to take some risks, such as being re-identified. PLC provides a web based consent form that allows the data to be shared with various organisations, and for more than one research project.[491]

The current working model of this system is an improvement over traditional consent forms, but it does not yet allow for consent to be carried forward to third parties.[492] Apple has developed a tool called ResearchKit which allows for the gathering of consent in an easy manner[493], and Sage is using it for some research studies.[494]

6.1.7 Self-management models: Data cooperatives

The arguments for data cooperatives are fairly simple and as in the wider cooperative movement they are centred on the pooling of resources for the direct benefit of the collective and other social values. Cooperatives have a long tradition and are well established in many other areas. They operate under shared ownership and direct democracy - one member one vote - instead of share voting blocs.

Discussions about data cooperatives[495] have looked at issues of transparency and governance that are shared by many other cooperative organisations. But in the case of data there are some added complications. Personal data is critical to self-representation and autonomy, and transferring data give the organisation power over the individual. Issues around informed consent and wider data protection remain even in a member-led organisation.

In cooperatives, trading in established economic sectors, such as industry and agriculture, the processes for adding value are widely understood but in the data economy it is less clear how and when value is added to data and how much of that value should be returned back to the individual. Transferring the data outside of the organisation for aggregation is a particular problem. These

questions are broader than cooperatives, but these are forced to confront them head on. Traditional cooperatives — with notable exceptions such as the UK Cooperative Bank and associated organisations and the Spanish Mondragon Coooperative Conglomerate — tend to gather individuals working on a focused project. But more recently the idea of cooperative organisation has extended to other kinds of multi-stakeholder projects for social objectives such as health or education. Given that the direct economic benefits of trading small scale raw data are unclear, data coops may benefit from a wider constituency and a clear public benefit approach.

It is important to distinguish between cooperative developments — such as open source software and crowdsourcing projects like Openstreetmap — and cooperatives proper. A lot of cooperative development is not matched by a real Cooperative organisation behind. Even democratically governed collaborative non-profits projects such as Wikipedia wouldn't fit the criteria.

The Good Data Coop

The UK based Good Data Cooperative[496] allows its members to make some money by becoming active players in the online behavioural advertising ecosystem. The coop provides an anti-tracking browser extension to stop third parties from collecting any data, while collecting search queries from its members. There are restrictions on sensitive data such as health, sexuality, etc. and the data is not linked to personal information on file. The organisation then sells that data to re-targeting networks which specialise in the tailoring of online advertising. The monies paid to tracking companies is then paid to the coop, which splits the profits between social lending - through the microcredit platform Zidisha - and technical development of the platform. The coop is at an early stage of development and it is difficult to predict the viability of the model. Given the very small amounts paid per advert it would require a very high volume to generate substantial sums. Their public statement shows that they have made just under $300 from 329 monthly active users.[497] Partnerships with consumer organisations could give this approach the numbers needed.

Datacommons Cooperative

US based Datacommons[498] cooperative takes a completely different approach. Conceived as a "movement-building organization", it is owned and controlled by other organisations of the social and solidarity economy movement, such as cooperative development centres, and ethical consumer groups. The coop is a platform for sharing information to amplify the scope and impact of their members" activities. The view of data coops as a model for collaborative governance has been explored elsewhere[499].

Health Data Cooperatives

The health sector is one of the areas where cooperative data sharing has generated a lot of interest. As we saw in the section on the new EU data regulation, health sciences research is a highly contested area with an ongoing battle between privacy and health advocates about informed consent and the role of corporations in delivering public benefits. Building a cooperative data pool has been touted by many as the solution to these problems.

A cooperative approach to health data raises some additional issues, such as the security of the data and the need for independent oversight and regulation.[500] The health sector will also need to consider the public interest and balance it with any desire of coop members to monetise their health data. Restrictions of research based on privacy could be replaced by restrictions based on funding to

pay for the data, to the detriment of society at large. These balances are quite delicate and difficult to communicate. For example, the UK Care.data project to share national health data with commercial companies has generated huge negative reactions, despite its public benefit rhetoric.

Cooperative health insurance services are common in many places without a proper national health service[501], while cooperatives specialising on alternative medicines[502] or areas neglected by national services, such as mental health are growing elsewhere. But these have not generally developed fully fledged data services.

Organisations such as PatientsLikeMe[503] advocate the open sharing of medical data to empower users of medical services, who can then compare the effectiveness of treatments, etc. The Data For Good initiative allows patients to donate their data for research. But these organisations go further towards "participant-led research", enabling wider access and analysis of health data[504]. PatientsLikeMe have created the Open Research Exchange (ORE), an "open platform for developing, validating and sharing health outcome measures that better reflect patients' experiences with a disease".[505]

PatientsLikeMe empowers people with health issues butultimately it is a for-profit company that makes money from selling data to pharmaceutical companies. It has a strong social mission, but this does not provide the same assurances for members as a cooperative. Even if the "members" do not want to receive a share of the profits, this arrangement gives them less control over their data. For example, their privacy policy makes clear that in the event PatientsLikeMe goes through a business transition, such as a merger, acquisition, or sale of its assets, personal information might be among the assets transferred.[506]

Many other organisations are trying to collect and aggregate health and genetic data. This includes DNA screening company 23andMe, who in addition to providing paid analytical services on health and ancestry also have a research branch.

There are not many examples of existing health data coops but some new projects show some promise. In the US, Our Health Data Coop (OHDC) is currently creating a platform for users of health services *to share anonymously their health records so a valid comprehensive evidence-based clinical research database is created to answer: "What is the best treatment for my disease?"*[507] OHDC is an actual cooperative registered under Minnesota's coop legal system and have considered many of the key issues around governance and security.[508]

Another example comes from Switzerland, where many global pharmaceutical and chemical companies have their headquarters. HealthBank is a Swiss *société cooperative* that aims to become the VISA for healthcare as *"an intermediary to provide a single point of transaction for health information for individuals, health care providers and researchers."*[509]

The Healthbank coop is not fully functional yet but it has a highly professional team with expertise in finance and research and appears to be a very serious endeavour. One possible obstacle to their growth may be their 100CHF joining fee — some 95EUR at the time of writing. This raises a fundamental issue for any economic alternative approach to data and identity: the requirements to raise capital cripple most alternative economic projects and the use if data as currency may not completely overcome this hurdle.

6.1.8 Data as commons

Describing personal data as oil may be useful from the point of view of an individual firm, but it is not very helpful to understand the wider implications and how it should be governed in accordance with rights and freedoms. There are other potentially more useful analogies in looking at data as a natural resource. In the discussions about property we saw how data could be understood in analogous terms as property crossed by a river, giving some rights to the land owner but not allowing exclusive control, with other users of the river also having some strong rights. Certain resources such as rivers, certain fisheries or rainforests are described as part of the "commons": resources accessible to all members of a society, including natural materials such as air, water, and a habitable earth.[510]

The application of the commons model to intangible resources was pioneered by Free Software pioneers such as Richard Stallman, who created the open first viral licenses that perpetuated free sharing of code. The Creative Commons project has successfully built an ecosystem of legal tools that allow creators to share all kind of copyrightable works. Several projects have taken this approach to data, such as the Open Data Commons' Public Domain Dedication and Licence (PDDL). Most of the discussions around data commons centre on public sector data, maps or other data by organisations. The basic idea is that such kind of data can be seen as an economic public good,[511] meaning that is both non-excludable and non-rival in that individuals cannot be effectively excluded from use and where use by one individual does not reduce availability to others. This public good approach has been extended to discussions of personal information, which "is currently treated as a pure public good and that data users are the primary beneficiaries of collective economic rights to personal information due to the presence of asymmetric information and transaction costs."[512]

The commons could provide a model for the governance of data contributed by individuals, particularly data that has had any personal identifiers removed, and where the value of the data resides in aggregation or big data analytics. When personal data that belongs is taken out of data protection or privacy regulations by removing any identifiers individuals may lose any rights over it. Rather than seeing an organisation capture the exclusive benefits of the data, a commons model would ensure these benefits are available to everyone. Many individuals would want to contribute their personal data - for example data from personal sensors in fitness bands - to a common pool that would allow them and other people to benefit but would not be hoarded by a single organisation. There are examples where this is happening in the field of health sciences, e.g. The University of Chicago's Genomic Data Commons,[513] but still mainly driven by organisations rather than the originators of the data themselves.

6.2 Policy strategies

In section 5 we examined the complex legal regulation of identities in the EU but despite that plethora of laws and regulations, people working in the field and public interest groups watching over the uses of data find that conflicts continue to appear. In some contexts, focusing on strict legal compliance, even in the most stringent form, is not enough.

This is a particular problem when we are confronted with a potential social good that could come from the use or release of personal information. We saw an example of this in the case of researchers, but participatory platforms and other similar public interest projects can also face a similar tension between the need to protect the data of those involved and the need to make innovative uses of such data.

In this section we look at some of the existing approaches that have been used to deal with this kind of situation where we want to build trust and engagement beyond legal compliance.

6.2.1 Privacy and data protection by design

Privacy by Design (PbD) is a set of fairly common sense practices developed by the former Information and Privacy Commissioner of Ontario, Canada, Dr. Ann Cavoukian in the 1990s. The approach is based on the idea that regulation is not enough and we must design technology and organisational practices with privacy as a key driver.

PbD centres around a set of 7 Foundational Principles which we reproduce verbatim:[514]

1. *Proactive not Reactive; Preventative not Remedial*
2. *Privacy as the Default Setting*
3. *Privacy Embedded into Design*
4. *Full Functionality – Positive-Sum, not Zero-Sum*

dichotomies, such as privacy vs. security, demonstrating that it is possible to have both.

5. *End-to-End Security – Full Lifecycle Protection*
6. *Visibility and Transparency – Keep it Open*
7. *Respect for User Privacy – Keep it User-Centric*

The principles have been applied to a very broad range of applications from biometrics in casinos to remote home care.[515] PbD has also been developed into a brand - even with its own logo - which has allowed the idea to be marketed quite successfully. PbD has become widespread beyond Canada, with many organisations worldwide integrating it into their practices. The global trade body for the mobile industry, GSMA, has issued PbD guidelines for the development of mobile applications.[516] The UK Information Commissioner recommends PbD, among other reasons because potential problems are identified at an early stage.[517] The British government has embedded this approach in their impact assessment for the new smart meters currently being deployed[518].

The US Federal Trade Commission has repeatedly endorsed PbD in its proposed privacy frameworks. Their 2012 report on protecting consumer privacy included PbD as one of the central pillars, together with more transparency and consumer choice.[519]

There have been many criticisms levelled at PdB, despite its popularity. There is no denying that PbD has managed to get privacy taken more surely in many organisations, but some companies struggle to convert the top level principles into practices without help.[520] It is also useful to systematise analytics and the design of technology and organisational processes, although at certain levels more sophisticated privacy engineering tools may be required.[521]

But like any popular system trying to provide simple heuristics or a set of general principles PbD can be a victim to its own success. The principles are too vague and general and can be used to justify practices that would fall short of adequate protections. For example, the Open Rights Group has been critical of some of the data practices of UK mobile companies, which are selling big data analytics for third parties.[522] Yet most of the companies challenged claimed that their practices are all driven by Privacy by Design. Similar concerns have also been raised by privacy researchers.[523]

When privacy is in conflict with the fundamental business models of organisations general principles may not be sufficient. Although some companies may take them seriously, in most cases there is no legal compulsion to follow the principles and PbD remains a justification for existing practices without transformative effects. But this may soon change in EU. As we saw in section 5.5, Article 23 of the new General Data Protection Regulation is titled "Data protection by design and by default", taking its lead from PbD.

6.2.2 Information Accountability

Information accountability (IA) is not strictly a policy approach, as it entails technological development. IA as promoted by Tim Berners-Lee, Daniel Weitzner and others[524] starts from the premise that current approaches to privacy - and wider information policy for that matter - are excessively focused on access controls. The result in their view is that once information has been released it is completely uncontrolled. As we saw in the previous sections, this is not completely accurate for the EU but appears to be the case in the US and other countries. But in any case they make a compelling argument that information that has been publicly released should still be used appropriately.

Another premise of this approach is that information cannot easily be constrained in practice, and once out it can be combined with other information to generate inferences that are not explicitly revealed. The information accountability approach aims to build Web technologies — called Policy Awareness — that support transparency and accountability by adding machine-readable labelling information about data provenance and appropriate use.

Embedding technical controls over the wider uses of data in the Web is undoubtedly very useful, but it should not be at the expense of efforts to control access in the first place. This argument has been played out in many contexts, such as whether we should protect the right of LGBT people to remain private — in the so-called closet — or protect them once their out. It is not a binary choice.

The information accountability approach follows from previous attempts to use web technologies to support privacy, most famously the Platform for Privacy Preferences (P3P), created in the 1990s to convert privacy policies into code. P3P has been very influential in the development of privacy technologies, but it has failed to gain traction for various reasons. For consumers it lacked enforcement and for industry it meant too much transparency. The underlying analysis about P3P remains applicable to newer attempts to implement information controls in web technologies: these "need to be used in concert with effective legislation, policy oversight and other privacy enhancing tools".[525]

6.2.3 Contextual Integrity

Contextual integrity (CI) is a philosophical approach to privacy develop by Helen Nissenbaum that has become very influential in recent policy developments, such as the US consumer data bill. The basic tenet of CI is that a "right to privacy is neither a right to secrecy nor a right to control, but a right to *appropriate* (N.B. Italics in original) flow of personal information".[526] Like information accountability, this approach that disclosure is not the end of privacy as some expectations will still apply about how the information is used.

The method of CI consists of breaking down the context of an information flow into components: roles and power structures, typical activities, associated norms and values. The latter are important to guide any decision on the appropriateness of a particular flow. Contextual integrity is defined in terms of informational norms: it is preserved when informational norms are respected and violated when informational norms are breached. Another important concept in CI is the *transmission principle, a* constraint in the distribution of information such as confidentiality, reciprocity or entitlement.

Nissenbaum answers some of the potential criticisms that can be levelled to CI: that it can be inherently conservative and support the status quo, e.g. "the tyranny of the normal".[527] She presents a complex argument for why this may not always be the case and why sometimes rules must be broken. But the nuances in this argument may be lost to some as the CI approach becomes popular with industry and some policy makers. Despite the caution applied by Nissenbaum herself, the idea that people have no right to control their information can easily be translated into anti-privacy practices and calls for weaker regulations.

One fundamental issue is who can decide a CI violations occurred. Creating a framework for enquiry without changing the fundamental power imbalances may not be sufficient. Nissembaum sees the drivers for decision making in terms of balancing social norms with general values, ends and purposes. But there are many social conflicts over norms and values that cannot be reduce to a balancing exercise. In addition, many new technological interactions have no clear precedent in the offline world, and may rely on metaphors that carry implied norms that may not be correct.

In addition CI requires utmost transparency, but some informational contexts such as online advertising may be too complex for simple heuristics. Nissembaum calls this the "transparency paradox". The basic ideas around contextual integrity are entering mainstream privacy policy and CI is used to analyse privacy issues by academics. But the detailed implementation of CI into logical systems to guide privacy decision making has not really happened at scale.[528]

The overall approach can be useful, but "norms" can be hard to map the real world, as people constantly change their decisions on privacy.

6.2.4 Social Acceptability

One of the problems with the way privacy regulation has developed is the focus on compliance, which sometimes leads to accusations of excessive red tape. Up to a point this is unavoidable, as the very process of handling data involves developing technical systems and data flows. Checks and controls should be built at every stage, and approaches such as Privacy by Design can greatly help here.

But unfortunately sometimes this is all that organisations believe it is required, and do not look at the wider issues. When a scandal hits the news they are surprised at the outcry.

This was the case with the Renew rubbish bins with advertising screens installed int he City of London, which sniffed the unique identifier (MAC address) of mobile phones in the vicinity that had WIFI enabled. The company had used this as part of a wider market research strategy, but found themselves forced to retire the bins after public outcry[529] and went into administration soon after.

The company saw nothing wrong, while admitting to operating at the boundaries of regulation,[530] and had even held hackathons on their tech.[531] They complained the facts had been blown of all proportion, and that they were not identifying individual users, and media reports looked at capacities that had not been built. In their view they were simply applying internet tracking techniques to the physical space, like "cookies for the street[532]".

WIFI sniffing techniques are indeed used to profile individual customers, including by retailers such as Nordstrom in the US[533], and this would have been questionable in the EU. But it appears that the UK company did nothing illegal. More transparency, possibly a big sign on the bins, would have helped improve their accountability, but they just miscalculated the acceptability of their project. The company providing the underlying WIFI tracking technology to Renew, Presence Orb,[534] has survived the scandal but now wants to listen to end users and privacy organisations.[535]

The difficulties in assessing privacy attitudes

Seeing privacy as a social good that requires consensus beyond the narrow technicalities of compliance is a good approach, but in practice there are problems with measuring privacy attitudes.

The US privacy group EPIC maintains a list of privacy related published surveys[536] that consistently reports high levels of concern. The UK ICO has published a report on what the public wants from data protection laws, including control, transparency, etc.[537]

Yet in practice individuals appear not to follow on those concerns and continue to share their data. This conundrum has been investigated by researchers such as Alessandro Acquisti, who has concluded that most people apply all forms of cognitive biases to privacy decisions: *"There are reasons to believe that consumers act myopically when trading off the short term benefits and long term costs of information revelation and privacy invasions."*[538]

In this context it is understandable that most companies may believe that the public will not care. Perceptions of privacy are context dependent. We must be careful not to assume that a willingness to share personal details in social media automatically translates into lower concerns about sharing of data on tax, health, education or social security. Privacy is also heavily dependent on exposure and direct experiences, such as media scandals or a close relative suffering identity theft. So what appears to be ok today may cause outrage tomorrow. There is an element of unpredictability on what is going to generate a reaction, but striving to build a broad social consensus is important.

6.2.5 Terms and contracts

Transparency is the basic foundation for any strategy to improve user led control over digital identities. Any technical or organisational innovation will need to be explained and accepted by users. As we saw in section 5, there are many problems with the way privacy policies and terms and conditions currently work. They are too long and complex and few people actually read them. In many cases they just present long lists of types of data that may be collected and long lists of potential activities to be performed on the data, which leave the end user none the wiser about what is actually happening.

Marketing driven data platforms such as Google and Facebook are particularly difficult as they combine data from various sources and track internet users in and out their applications. Terms and policies are also a particularly thorny issue with mobile apps, where the amount of data collected can be a lot more intrusive and in many cases there is no policy document at all.[539]

Many projects are trying to help solve these problems. For example, mobile app developers Docracy have released a simple open source generic privacy policy for mobile apps with variants for collecting location data and advertising funded models.[540]

Many projects attempt to communicate policies with visual icons.[541] But given the complexity of most policies this may be hard to implement without simplifying the policy itself, and could potentially mislead users. There are many examples where this is being put into practiced. The EU funded PrimeLife project developed two sets of icons: for website's data handling and for data disclosure in social network sites.[542] Like most other similar projects the icons have not gained widespread traction. One issue with icons is that in order to be quickly understood they would need to be consistent across many websites and platforms.

But even in the best possible scenario where policies are simple and easy to understand there are limitations to what can be achieved. The policy is meant to inform the user about what is going to happen so s/he can make a decision. But in most cases, these are spurious choices. When it comes to using dominant internet services the choice can be social participation or self-ostracism.

6.2.6 Trust frameworks

Trust frameworks are one of the alternative solutions to the concentration of power on corporate identity providers. These frameworks consist on a combination of legal and social arrangements that allow individuals to have more control over their data, and organisations to collaborate in a less centralised manner than if one single dominating company were to provide a platform, as in Facebook's case. They normally rely on a shared technology platform, such as a personal data store or personal cloud where the data is primarily stored.[543]

The Open Identity Exchange supports two trust frameworks from Mydex and Respect, which consist of agreed principles such as *"we will respect each other's digital boundaries".*[544] While these are positive developments it is unclear to what extent they can be enforced, and importantly, whether they would provide any more or less protections to end users than simply having strong data protection and consumer regulation. Most of these frameworks still require individuals to trust an organisation to behave ethically. Trust frameworks have also been proposed in relation to opening e- government services,[545] and are in place in academic institutions with the Eduroam system which allows students, researchers and staff from participating institutions to obtain Internet connectivity across campus and when visiting other participating institutions.[546] The UK government has developed an identity assurance scheme based on similar principles, where prospective users of e-government services can register and be authorised via a network of external identity providers.[547]

A slightly different approach has been taken by the Institute for Data Driven Design (ID^3), created by digital identity pioneers Sandy Pentland, John Clippinger and David Bollier, who have collaborated with the WEF. ID^3 aims to make trust frameworks available to common internet users. Their Open Mustard Seed (OMS) *"is an open-source framework for developing and deploying secure and trusted cloud-based and mobile applications. OMS integrates a stack of technologies including hardware-based trusted execution environments, blockchain 2.0, machine learning, and secure mobile and cloud based computing. This platform enables the managed exchange of digital assets, cryptocurrency, and personal information."*[548]

The ID^3 has gained traction with Bitcoin companies.[549] There is a fundamental connection between digital currencies and identity. The idea of data as currency we discussed in the previous section has more profound implications for digital identities and society at large. Digital money expert David Birch proposes that digital identities can lead to a cashless society where currencies like Bitcoin become the norm.[550] For Birch the main reason we need money is to give us enough trust to trade. In the past we could have achieve this with face to face personal trust or letters of credit, and now we start to use non monetary mechanisms such as cards and mobile payment systems.

New digital identity systems can build the trust that until recently required a national bank and cash. The technology behind Bitcoin, the blockchain - a distributed edger protected by cryptography - has proved a very versatile technology for authentication without the need to rely on a central authority. There is even a project called Ethereum to develop a complete internet-like computing environment based on blockchain technologies to allow applications to communicate without any central trust service.

6.3 Ethical frameworks

6.3.1 Ethics

There is a growing interest in the ethics of information, particularly in the context of Big Data, with whole books dedicated to the subject. Ethical approaches to data processing will also go beyond what is legally acceptable to ask "are we doing the right thing?". It is fundamentally a process of enquiry that may not give easy answers without effort. This enquiry can be broad and include questions on issues that are outside strict privacy laws - such as creating monetary value from users' data.

It is important to stress that Ethics are not a substitute for compliance for proper legislation and respecting human rights and should build atop these. But this is not always the case in some of the proposals in circulation. For example, writers such as Kord Davis remove fundamental rights from the equation and leave individuals at the mercy of companies making the right decision.[551] This is unfortunate, as some of his proposals are quite sound, although very centred in US business culture. His proposed process would start with an organisation articulating its fundamental values, then translating these into concrete actions. Davis rightly stresses the importance of internal processes and getting buy in from difference parts of the organisation. This is very important to avoid delegating privacy to a specialist officer instead of embedding it in the organisation. He also provides useful questions to ask - such as what rights users have - although many of these are answered under EU law. He also includes some useful tools for internal analysis. But care should be applied, as these ready made toolkits while appealing to business culture, can easily slide into compliance checklists. Where his proposals fall short is in failing to engage external stakeholders.

Asking people is important, if time consuming. The BBC carried out a very interesting programme of research to help guide their approach to personal data in the Internet of Things revolution, that included many interviews with people from outside the organisation.[552] The kind of questions here are different from the surveys we mentioned in the previous section. People need the space to discuss the issues in more detail.

But the Ethics of personal information did not start with Big Data. There is a wealth of expertise in the area of academic, social and health research, where ethical boards normally have to approve the

use of personal data in any research project. There are well established ethical guidelines for approaching the use of personal data in the design of research proposals[553].

As research processes become more complex, so do the requirements for ethics compliance. The UK Data Archive, which houses the largest collection of research data from the social sciences and humanities, require adherence to a set of ethical principles in order to use the service[554]. These include *"a duty to treat participants as intelligent beings, able to make their own decisions on how the information they provide can be used, shared and made public (through informed consent)"* but also *"a duty to wider society to make available resources produced by researchers with public funds (data sharing required by research funders)"*.

A very valuable contribution to the ethics of data comes from the UK Nuffield Council on Bioethics, who have published a report on "The collection, linking and use of data in biomedical research and health care: ethical issues"[555]. The report proposes to define a set of morally reasonable expectations about how data will be used in the data initiative, to be done in a participatory manner and with proper consideration of the public interest. Importantly, they frame the ethical approach and governance of data under the rule of law and the respect for human rights.

6.3.2 Responsible Innovation Frameworks

The use of ethical enquiry and approvals is not without its limitations. Obtaining ethical approval in research can become a one off hurdle that becomes the exclusive responsibility of an ethics committee without any external input. The long term impacts sometimes fall out of the scope of consideration, or the expected impacts are defined narrowly.

These limitations are felt more acutely in areas of innovation with new technologies of high uncertainty, and the response has been to create a broader model of enquiry called Responsible (Research and) Innovation, which is now part of most EU funded science projects.[556] Responsible research and innovation is described as "making science with society and for society"[557], and it involves broader participation in the discussions and looking at long term effects of new technologies.

Developments in the use of personal information could benefit from this approach, which until now has been mainly restricted to other areas, such as nanotechnology and synthetic biology, despite some tentative research on its application to information technologies through the ETICA[558] project which looked much broader than privacy issues.[559]

The EU RRI model[560] focuses on education, governance, ethics and open access to results. This means "democratic governance of the purposes of research and innovation, steering these towards the 'right impacts' that are anchored in societal values"[561]. But importantly, the model tries to deal with the unpredictability of outcomes, something that is very relevant in the context of innovative uses of data that may have been de-identified. Anticipation, reflection and deliberation should inform action with a broader collective responsibility through engagement. The following paragraph gives an idea of what a RRI exercise would look like?

> *"To give an example, imagine a collaborative research project on a mobile biometric security device for online banking applications. Actors with responsibility for privacy in such a project might include the policy-makers who approved a call, funders who administer the budget, researchers who adhere to professional standards or end user organisations which represent user interests. These subjects of responsibility could discharge their responsibilities by including technology foresight, implementing*

value-sensitive design or privacy by design, or using methodologies from constructive TA (Note: technology assessment). Their shared normative commitment could refer to specific legal requirements, such as the European data protection framework, but also to a broader goal of improving the greater good of society or minimising the potentially negative impact of end user perception on the acceptance of the technology.[562]"

Like in any other approach there are opportunities for abuse, and the idea of making data subjects co-responsible would smack of opportunism unless there are very clear benefits for them even if channelled through society a large. Ultimately, those handling the data have a higher responsibility. In addition, RI may create unrealistic expectations and incur unacceptable overheads for small organisations. Yet, overall it is a positive development.

6.4 Technical Strategies

In terms of realizing fundamental rights to self-determination of personal data and the right of free expression, a number of open standards and corresponding code has been created that allow both control over data and encryption. These standards were earlier overviewed in D4.1 in 2014, but they were not given a detailed privacy and security analysis. After reviewing the basic concepts and available technology, current cryptographic measures to preserve privacy and private communication will be explained, as well as their limits. Next, we'll overview identity frameworks based on open standards as currently implemented by current large providers such as Facebook, Google, and Twitter - but also easily implemented by open-source frameworks. This framework is currently based primarily on the use of OAuth (Web Authorization), an IETF standard[563]. We'll look at common criticisms of OAuth and alternatives such as User Managed Access[564] as well as WebID+TLS (Story et al., 2014), both of which fail to implement elementary security and privacy considerations. Then we'll revisit data portability and the Activity-Streams based standards of the W3C Social Web Working Group allow a measure of data-portability. Lastly, we'll provide a number of basic recommendations and security guidelines to improve the use of these standards in D-CENT, as well as future directions in decentralization using blockchains.

6.4.1 Identity and Anonymity

Identity

Identity frameworks ultimately have the goal of verifying that some digital information about an entity - be it an individual human or some collectivity such as a formal organisation - holds about the entity itself. Traditionally, this is considered to be some of spanning of the gap between the digital and the analogue world, where the digital information stored in multiple and often fragmented databases can be "attached" to the individual "in of themselves." However, this neo-Kantian division between the digital data and the real "flesh and blood" human may no longer hold true in a world where digital information play and important and increasingly seamless role in our daily lives: It becomes increasingly difficult to thread apart the 'real life' of an individual and their Facebook profile, given the repercussions of a profile on everything from social life to employment prospects.

The consumer self is ontologically not distinct from its representation in the electronic market-space. Thus, in general we find that identity frameworks are essentially now part of a larger movement of the digitization of everyday life. Historically, identity was controlled by the state (for example, via centralised identity registries). Today, Internet-based identity systems are controlled mostly by a few large Internet companies, each with its own custom applications that are often incompatible with other companies - and so earning the moniker "silo." As explored previously, as no single company has dominance over the entire market, an "identity eco-system' has to be created to allow them to exchange data about particular users across their silos and identify users across silos.

Anonymity

Identity is the opposite of anonymity, where both personal data and possibly unintentional "digital traces" are unlinked from the user. Thus, unlike identity systems that attempt to "link" attributes to a user across systems, anonymity systems aim for a property of "unlinkability," namely that the same user can use multiple services without their identity (and behavior) being linked. The process of discovering an identity of a user is as such then de-anonymization. Anonymizing technologies have been studied for decades since the first seminal papers by Chaum.[565] Anonymity has classically been defined as "the state of not being identifiable within a set of subjects," where that set of subjects is called the *anonymity set* .[566] Note that an anonymity set of possible subjects has been refined in terms of information-theoretic measures that look at anonymity on a much-more fine-grained level, such as the probability that a given identity is attached to a given transaction.[567]

Anonymity is usually defined not as either "yes" or "no," but in terms of the anonymity set – but also given the capabilities of a particular attacker that is attempting to identify a particular user or users. This attacker usually has the ability to make observations of a given system. In particular, anonymity is usually defined in terms of two kinds of attackers, an *active attacker* that is attempting to actively determine the identity of a user by attacks on some part of the flows of data (usually in order to decrypt messages), and a *passive attacker* that monitors meta-data and then attempts to use the collected flows to de-anonymize the user. These attackers can observe either the entire system (a global attacker) or only a local portion of it (a local attacker). For example, a local active attacker would be an attacker that actively scans wifi for passwords being sent in the clear (over HTTP rather than HTTPS for example) and then steals those passwords to gain access to a user's account, and so illicitly retrieve attributes such as home addresses or even credit card numbers. In terms of de-anonymizing, the attacks on the Firefox browser to de-anonymize Tor users would count as active attacks. A passive attacker could simply monitor all the e-mail messages sent, and the use those to de-anonymize users by discovering their social graph via the messages sent, even if the content of the messages were encrypted. Although not known to be used by the NSA, viewing all the entry and exit node traffic over the Tor network and then using that to statistically de-anonymize users would be an example of a global passive attacker.[568]. In general, mass surveillance by a powerful agency such as the NSA would be global passive attacker, while targeted surveillance would be local active attacker. One can also consider local passive attackers that can only observe the identity system partially. In general building anonymous systems is difficult. First, holistically almost any system tends to "leak" metadata (timing and other sources of side-channel attacks) that can be used to de-anonymize users even if messages are encrypted. Also, once data has been disclosed on the Internet, it tends to remain in the public domain, and so preventing disclosure is difficult. Lastly, studies have shown that a very small amount of data even in "anonymized data-sets" where personally identifiable attributes

have been deleted can lead to the data-set being de-anonymized. Although difficult, one goal of an identity eco-system is to maintain best level of anonymity for its users.

Privacy Technologies

Rather than aiming for anonymity, many systems aim for a less difficult goal of privacy-preserving technologies where the goal is to provide the user the most privacy possible even if anonymity itself is impossible against realistic adversaries. Traditionally, Data Protection in Europe aims only at personal data and aims at legal frameworks for enforcing privacy. However, as shown by the collection of metadata, particularly by global passive attackers, and due to cross-jurisdictional issues, legal frameworks are insufficient without some technical grounding as non-personal data can be used to identify a user and requests for redress can be legally ignored. As mentioned earlier, in 2009 Ann Cavoukian, the information and privacy commissioner of Ontario Canada, aimed for "privacy-by-design," where privacy-enhancing technologies deploying encryption and anonymizing techniques are used throughout the entire engineering lifecycle in addition to legal constraints.[569] Ideally, the legal definition of privacy and data protections would be enforced through the engineering process. Privacy itself is often left undefined, but in general can be thought of in two different ways: minimizing the amount of data disclosed to the be only that data necessary for the transaction (and so maximizing anonymity) or as giving the user the most control possible over their data. The first kind of privacy-enhanced technologies aims to use cryptography to hide message contents and along with anonymizing techniques to prevent metadata analysis. The second kind is focused more on giving the user control over their own data, as explored in more detail in the concepts of *personal data stores* where essentially each person is given control over their own personal data, which is designed to be stored independently and shared by the identified user themselves with their full knowledge rather than created and shared without their knowledge by a data broker.[570]. These sorts of personal data stores are thus dependent heavily on policy can be used in conjunction with formalized policy languages, such as the W3C P3P language[571]or general purpose languages such as AIR.[572] However, often policy languages can be used without any technical actual enforcement mechanism and so their claim to be privacy-enhanced technologies per se is difficult to maintain in of themselves, as they focus rather on auditing existing systems for violations of whatever legal rules the system is supposed to uphold. Yet as part of larger system based on technical enforcement using cryptographic primitives and anonymizing techniques, policy languages could be useful. In summary, privacy-enhancing technologies and user-centric identity management systems based on policy are not necessarily contradictory. The goal of this section is to explore the privacy and anonymity properties of user-centric identity management systems after reviewing the current landscape of cryptographic tools.

6.4.2 Cryptographic Tools

Strangely enough, cryptography is a necessary building block for both identity and anonymity, as using cryptographic primitives such as digital signatures one can authenticate an identity and attach it to a person, while in terms of anonymity and privacy cryptography is necessary to both hide the identity of an entity and to prevent unwanted third-party access to the content and metadata of messages. For a thorough explanation of cryptography, please see D4.3.

Traditionally, the mental model used by people of encryption is that a single long-term public-private keypair is used to generate symmetric keys that then encrypt and decrypt messages, with signing

being done by a separate key. This is the model used by encrypted email such as OpenPGP.[573] However, the disadvantages is that if the key is compromised *all* prior and future email to be read. It can be argued that the use of a single long-term key is actually much more beneficial to law enforcement, since they can request access to the key or compromise the user device to gain access to the key. This is typically the strategy of actors like the NSA if done illegally, or the strategy of legalized "key escrow" pursued by the United States' failed "Clipper Chip" programme and new FBI efforts to legally mandate "backdoors" into cryptographic efforts, as well as parallel efforts in the UK. The newly reported MIT report "Keys Under the Doormat" provides an excellent overview of why such approaches, even if legal, damage security.[574]

A cutting-edge and usable cryptographic messaging tools that solve many of the problems of PGP is OTR. Unlike PGP, it features *perfect forward secrecy*. Given that "perfect forward secrecy" is a precise information-theoretic term defined by Claude Shannon,[575] we will continue to use the term "forward secrecy." Forward secrecy defines the property where for a given message, if the private key material is compromised, messages are not compromised as each message is encrypted with a new key. From the perspective of privacy this is vastly superior to traditional messages, and the compromise of a single long-term key does not allow past messages to be read, as the key is deleted. This key is generated per-message using a process called key-ratcheting. This was first demonstrated by "Off the Record" messaging for chat between two users (synchronous messaging),[576] and broadened into group-based asynchronous messaging by Silent Circle[577] and TextSecure.[578]

One of the few working decentralised anonymizing software solutions, Tor, focuses on the IP level and is aimed for anonymizing web-browsing.[579] Although in theory not resistant to a global passive attacker, it has proven difficult for even the NSA to tackle. In terms of usable software, mix-networking - which unlike the onion-routing used by Tor, is resistant to passive global attackers focusing on metadata analysis - has been mostly applied to email via software such as Mixmion, although the European Commission has recently funded a large-scale generic mix-networking platform called Panoramix.[580] In the mean-time, Riseup.net offers Off-the-Record messaging via Tor to enable decentralised, anonymized communications and Pond allows anonymous communications by mix-networking and decentralised hosting, although it is still very early in development. Thus, for the time being, there is little in usable decentralised privacy-enhanced messaging software.

Another hard requirement for D-CENT is decentralisation. However, most existing systems that use cryptography to hide the data of messages are not decentralised. In particular, most governments, such as the Finnish government, use as the backbone of their eID schemes centralised PKI registries that associate some key material to their citizens, and while that key material could be useful in authenticate citizens using digital signatures, these systems have no technical privacy-preserving characteristics. The European Commission-funded work on using zero-knowledge proofs to allow attribute-based credentials for identity solves many of these problems, allowing users to authenticate revealing only actual attributes needed (such as "I am over 18?" for voting rights) without revealing their identity. However, these techniques are not supported cross-platform, and the fast Montgomery Matrix operations needed to build them into the browser are not supported by the W3C Web Cryptography API. The best existing open-source library for enabling these attribute-based credentials, IDEMIX, does not yet have adequate speed (i.e. authentication often takes up to 10 seconds[581]) although smart-cards with attribute-based credentials can reach speeds of less than a second.[582]

In terms of messaging systems like e-mail, encrypted messaging based on S/MIME again suffers from centralised key management systems, and decentralised e-mail alternatives such as PGP present large usability problems by off-loading key management and identity management to the user. Current

best-of-breed Off-the-Record chat systems are centralised, including TextSecure and Signal by Open Whisper Systems.[583] The Jabber protocol itself that the original OTR ("Off the Record", as described earlier) protocol is implemented on it is decentralised and features end-to-end encryption, but Jabber is not well-supported in terms of the underlying codebase and there are very few Jabber serves in practice. In the future in terms of technically enforceable data protection measures, the European Commission should support increase research on end-to-end encryption both for chat and e-mail as well as anonymizing technologies such as mix networking. Without these building blocks properly constructed, technical enforcement of eID and Data Protection will be impossible.

6.4.3 Identity Ecosystems

In terms of identity, an *identity ecosystem* is a collection of services that wish to share data about an entity. In this work, we assume there is a user that is sending some kind of information to a *relying party*, a services that wish to access verified identity claims. The source of the identity claims is called an *identity provider*, a service that stores and can possibly verify identity claims on behalf of a user. The common example would be having a user send their username and password combination to Facebook via Facebook Connect, the identity provider, to sign-on to a third party service such as a newspaper like the Guardian, the relying party. The Guardian also may require some information from Facebook, such as the full name of the users and their interests in their Facebook profile, in order to customize their service. This information required by the relying party from the identity provider are considered *identity claims* or *assertions*.

Identity frameworks are socio-technical frameworks, with both a legal and technical component. The legal component may be legislated from government(s), created via industry self-regulation, or in some cases be non-existent as there may be no suitable legal framework or any existing framework is ignored or overridden due to terms-of-service agreements with the user. This latter case is the most common case. There have been attempts to self-regulate in the United States of America (with elements of a public-private partnership due the National Strategy for Trusted Identities in Cyberspace). The legal framework often includes auditing and certification requirements that must be fulfilled for one to participate in the identity eco-system. This may mean that a third-party has to inspect the (usually cryptography-based) security of the system or determine if the identity eco-system obeys certain laws, such as compliance with Data Protection (for example, not retaining or sharing data beyond what is necessary).

For example, Open Identity Trust Framework (OITF) "is, a set of technical, operational, and legal requirements and enforcement mechanisms for parties exchanging identity information" that assess whether or not identity providers and relying parties can be certified in following the OITF industry-self regulation in the United States in this space.[584] OITF includes the Open Identity Exchange (the first trust framework provider certified by the US Government. Booz Allen Hamilton, CA Technologies, Equifax, Google, PayPal, Verisign, and Verizon), the UK Identity Assurance Programme (IDAP)[585] and the Respect Network. This self-regulation includes terms of service between identity parties and relying parties being established. In terms of security and reliability, levels of assurance are provided. Policy-makers are addressed via a "Memorandum of Agreement" rather than binding regulations. Auditors may be called into check to see if the agreements are being followed. End-users are represented via a relatively weak mechanism known as an ombudsman whose job is to look "after the interests of individual users under their respective jurisdictions." Of course, the danger is that the identity provider itself controls the ombudsman, leaving the role to be nothing but

marketing where "Facebook represents its users." Although Europe does not yet have a large-scale private-sector identity framework, such a future framework could strengthen the agreements between users if unified rights and directives such as the Data Protection Directive were taken seriously.

The problem with the Open Identity Trust Framework is that it often disguises a total lack of privacy. In particular, Respect Network has been publicly shamed for claiming to use an approach based on "privacy-by-design" but having no real technology to back it up despite being certified by the Open Identity Trust Framework. If anything, this is proof that industry self-regulation without adequate technical grounding is harmless at best, but dangerous and disingenuous at worse. Note that the Respect Network was founded by the same Drummond Reed that attempted earlier to create his own patented proprietary identity system to replace the Web and falsely claimed to patent "push" technologies,[586] so it should be no surprise that similar bad business practices are being repeated in the identity space. In particular, it was noted that while the Respect Network claims that "We believe privacy, control, and portability are requirements, not features," their critics at security firm Sophos noted that "The highlighted words look as though they're links to further information, but they're not."[587] In fact, Sophos noted that the Respect Network was taking advantage of distrust in Facebook to have users that enter into an "on-line social contract without explaining who you are, what your intentions are, and what mechanisms you have in place - now and for the future - to protect that privacy." Of course, a rights-based approach that required disclosure requirements that was technically and legally backed would not let users be fooled by such "privacy snakeoil."

6.4.4 Security Analysis of Identity Protocols

Given that we cannot only rely on legal frameworks to defend the security of identity transactions and user privacy, a technical analysis is in order. A number of different protocols have been proposed for the authorization of the transfer of identity protocols. While many high-security and privacy-respecting protocols have been proposed relying on zero-knowledge proofs (also called "attribute-based credentials"), unfortunately these protocols have not achieved widespread usage.[588] This is in general due to the inability of the user's browser or client device to support the required cryptographic primitives for zero-knowledge proofs, as well as a lack of binding legislation that required them, such as in the recent European eID directive where a requirement for attribute-based credentials were removed. This is unfortunate as such protocols based on zero-knowledge proofs are technically the most privacy-preserving and secure technologies.

Thus, we will restrict our analysis to the more insecure and less private yet popular authorization protocols used in the wild, namely OAuth and its variant, OpenID Connect. For the last several years a number of alternative proposals based on public-key cryptography (which is supported by the browser and most client devices) have also been proposed such as BrowserID (also called Mozilla Personae)[589] and WebID. These latter alternatives have all also failed to have much uptake outside the developer community, while touting themselves as privacy-preserving and secure.

For each system, we will outline the system, provide a detailed step-by-step information flow, and then analyse the system for two threat models. The first threat model is an active attacker that actively is attempting to gain as much information about the user, including their credentials and personal data, as possible by either maliciously impersonating a relying party or an identity provider.

OAuth 2.0 and OpenID Connect

OAuth 2.0 is the standard protocol for authorizing the transfer of identity claims across identity provider to relying parties, and is used by most major sites such as Google and Twitter[590]. The history of OAuth is that it was originally designed as a way for users to host their attributes on an identity provider of their own chosing and to provide only a selected number of attributes to be shared with a relying party, rather than (as was typical in 2004-2005) allow a relying party to have control over the username-password authentication credentials of another site (i.e. a user simply handed their username and password at one site to another!) and then access all of a user's attributes at that site. Given that in this scenario there was no way to restrict what attributes a relying party could obtain or to prevent the compromise of a single server acting as a relying party to compromise many accounts at many other servers, OAuth 1.0 was a great improvement. OAuth 2.0 committed a number of large changes to OAuth 1.0 to make it more secure (such as enforcing TLS usage between the user and the sites as well as between the identity provider and relying party) while keeping the general information flow.

On a high-level, OAuth 2.0 is an authorization protocol that gives the user the ability to consent to the transfer of attributes via redirecting the user to the identity provider for authorizing the attribute transfer and then re-directing them back to the relying party. OAuth 2.0 does not specify any particular authentication protocol, and so is used typically with user-names and passwords. The transfer of attributes is then done between the identity provider and relying party server-side via the creation of short-lived shared secrets given by access tokens that confirm to the identity party and relying party that the user has authorized the attribute transaction.

OpenID Connect is for the most part simply a profile of OAuth 2.0 for exchanging attributes, but adds a number of string identifiers in the response between an identity provider and relying party for common kinds of attributes such as username and address.[591]OpenID Connect also specifies that JSON Web Tokens (JWT) can be used in various flows to provide the identity provider and relying party the ability to sign transactions and even encrypt them (Jones et al., 2014). This provides another layer of security. The flow of OpenID Connect and OAuth 2.0 is shown in Figure 14.

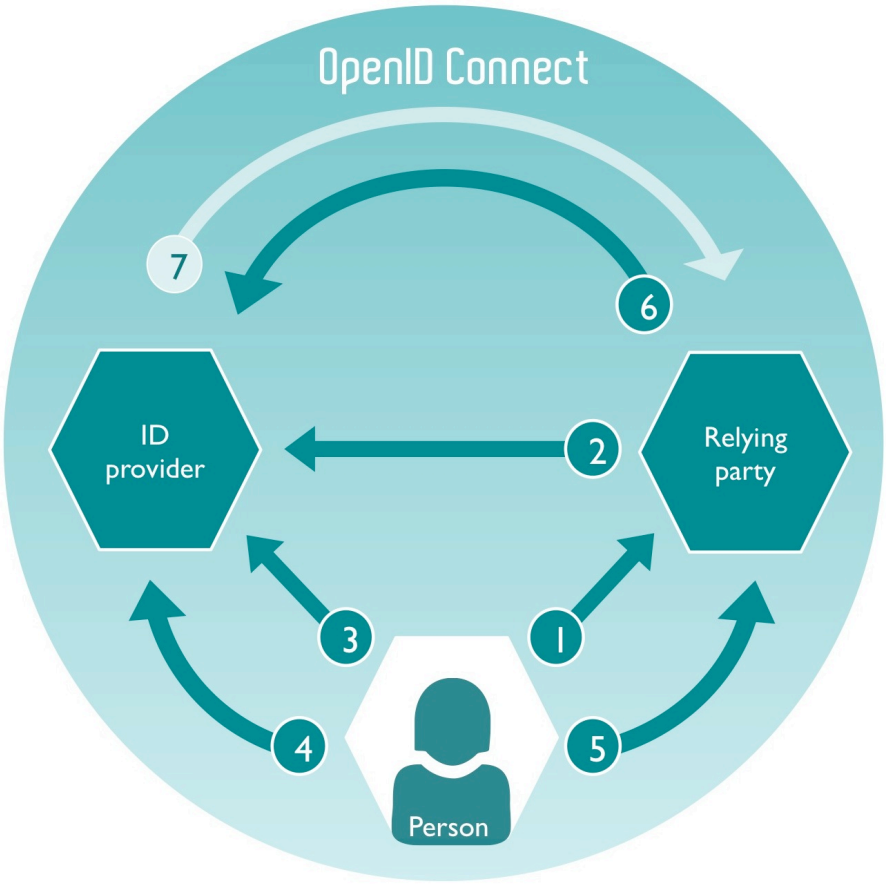

Fig. 14. OpenID Connect and OAuth 2.0 Information Flow

1. A user visits a relying party that needs attributes.
2. The relying party makes a request for attributes to the identity provider.
3. The user is redirected to the identity provider from the relying party.
4. The user authenticates to the identity provider (typically using a username-password combination) and is granted a bearer token.
5. User is redirected back to relying party and grants authorization token to relying party.
6. The relying party sends the authorization token to the identity provider and receives an access token (a bearer token with a scope and limited lifespan).
7. While the access token is valid, the identity provider sends attributes to the relying party.

One critique of OAuth 2.0 is that it does not allow a central point of enforcement for either a user's preferences. For example, a user may want to make sure multiple identity providers all maintain the same level of privacy protection. This is addressed in the User-Managed Access (UMA) specification by adding another party, called the authorization server, to the OAuth 2.0 authorization flow[592]. In particular, this authorization server simply sits between the identity provider and relying parties and makes sure the flows conform to the user's preferences in what is called a "three-legged" OAuth flow. This general schema can also be used to enforce not only user preferences but some kind of identity regulation, such as legal constraints, and so a model with a "hub" rather than an authorization provider has been adopted by the UK's identity service GOV.UK[593]

There are a considerable number of privacy problems with OAuth 2.0-based flows. First, we'll consider active attackers. If the relying party is malicious, it could redirect a user in Step 3 to a fake "identity provider site" that can then phish their credentials for a real identity provider. Although less likely, the same attack can be repeated by the identity provider itself in Step 5 by redirecting the user back to a fake "relying party" site and then, logging the user out, attempt to gain the credentials to the relying party site. These problems could be solved by better authentication technologies, such as those based on client-side key materials or zero-knowledge proofs. A related problem that is unsolvable via simply better authentication is that if the tokens are shared secrets rather than signed tokens, they can be stored and used for replay attacks if either the identity provider or relying party is compromised (or in the case of OAuth 1.0, if they are sent over HTTP rather than TLS, as demonstrated by the Firesheep browser plug-in[594]).

If the identity provider is compromised, they have all control over a user's attributes and can share them with any party they wish without explicit user permission. This is particularly dangerous is relying parties are colluding with an identity provider and there is no ability for a user to audit what attributes have been shared. Worse, there is an unsolvable overriding privacy problem with this information flow is that the identity provider can observe all transactions of a user to all relying parties that need information from that identity provider, and link these transactions to a particular user. As detailed by recent research[595] this problem is made even worse, not ameliorated, by a centralised "hub" as given by GOV.UK, and these sort of blatant privacy violations would also apply to UMA-based systems.

Since all information is sent in TLS, in terms of content OAuth 2.0 based flows are safe from passive attackers. Local passive attackers are capable of de-anonymizing based on timing observations, which would be difficult. However, any global passive observer that can observe the identity provider can also likely de-anonymize a user by simply observing the redirection flows between relying parties and one or more identity providers, even if the actual attributes are encrypted.

In general, it was viewed that one large weakness of OAuth was that authorization was out of the hands of the user, and that this was partially a side-effect of the user not having control over any key material. The very fact that this was assumed complicates the OAuth flow, leading to many redirections that are also the source of possible attacks. The WebID proposal attempts to provide a way for a user to achieve secret key material and then use this key material to share attributes (Story et al., 2014). In general, the main advantage of WebID is that, since the user can provide a request for attributes signed by their own private key, they do not need to be redirected to the authorization provider.

The WebID protocol is actually WebID+TLS, since it relies on TLS to assign the user a public-private keypair via the generation of a client certificate by the user. WebID+TLS states that a URI to the identity provider can then be provided by inserting the URI into the "Subject Alternative Name" field, and that the user is assumed to be able to post the public key associated with their client certificate to the identity provider. The main issue facing WebID is that most browsers do not support using self-signed certificates in client re-negotiation of a TLS connection without a confusing user-interface that invokes an error message. Therefore, there has been little to no adoption of WebID+TLS outside of a handful of users.

A variant of WebID+TLS has been proposed[596] that attempts to avoid this using the W3C WebCrypto API to send a signature from the private key material corresponding to the public key material published on the site. The general flow of WebID is shown in Figure 15below:

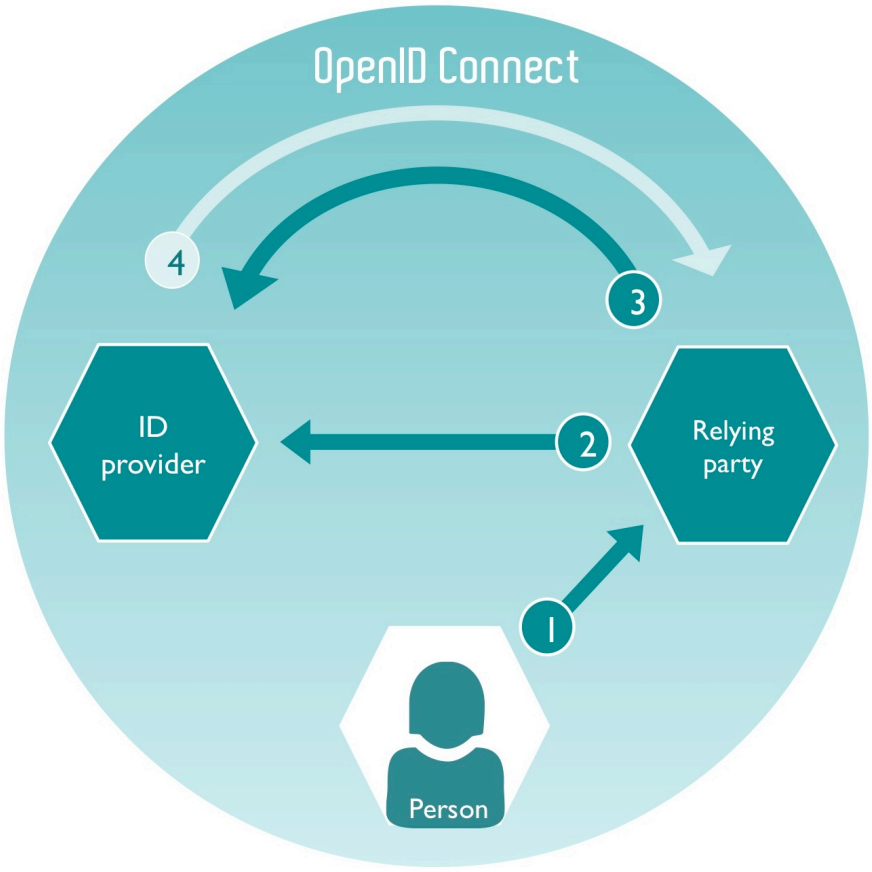

Fig. 15. WebID Information Flow

1. User presents a client certificate that includes the URI of the identity provider to the relying party.
2. Relying party extracts the URI of identity provider from client certificate and retrieves public key from identity provider.
3. If public key matches key in client certificate, authenticate user as the user can be proven to be in possession of a private key corresponding to the hosted public key.
4. Relying party retrieves identity claims from identity provider.

Unfortunately, WebID+TLS also suffers from a number of fatal security flaws. While it is superior to OAuth insofar as it avoids authentication to an identity provider via redirection, it assigns a user secret key material via TLS client negotiation. First, this violates the separation of the network level of TCP/IP from the protocol layer of an application, and thus is rightfully considered an error by browsers. Second, a critical security vulnerability has been discovered in TLS renegotiation – the infamous Triple Handshake Attack – and so TLS client renegotiation is being deprecated in TLS 1.3.[597] So WebID+TLS will not work in future versions of TLS and future browsers.

Simply moving out of TLS, as suggested by WebID+RSA, does not solve the problem. First, simply matching the public key pairs as given in Step 3 (or a variant that used only public-private key materials sent on the application-level) is not enough, as an active attacker could do a credential-forwarding attack by impersonating ("man in the middle") or otherwise successfully compromising the relying party. This malicious relying party would simply forward a request for the public key to

the user and then forward the response of the user with the key to the identity provider. The way to ameliorate this attack is to sign not only the user's identity URI but also a server's identity with a nonce.

One possible privacy advantage of using user-centric key material would be that an identity provider could encrypt the user's attributes in transit, or even encrypt the values for given attributes on the identity provider. This would be a large gain for privacy, as then the identity provider would not have full access to all user attributes like they do in OAuth 2.0 However, a larger architectural problem with WebID is that it suffers from a basic misunderstanding of cryptography of identifying a user with a single key, with the assumption that this key is used for both signing (as would be needed in WebID+RSA or any flow that used signatures to authenticate the user's permission for any attribute request) and encryption (as would be needed to defend the content of a user's attributes). Using the same key for encryption and signing opens one up to Bleichenbacher vulnerability that still occurs in the wild for RSA.[598]

In terms of passive attackers, WebID+TLS fares worse than OAuth 2.0 as the first client certificate is sent in the clear in Step 1 (before the TLS connection is established) and so the user data for the identity provider would be leaked to any passive observer monitoring the connection between the user and relying party, including only local attackers. In OAuth 2.0, this is addressed by using TLS over all connections so no information is leaked in the clear. This is addressed by WebID+RSA or other improved variants that do not use TLS client re-negotiation by keeping any authentication out of the network level. If authentication to the network level is needed, then it could be done using a ChannelID identifier in TLS. In detail, a WebID private key could be used to sign the TLS-level ChannelID that only identifies the current TLS connection, binding the user's authentication to a distinct TLS connection. A global passive attacker that was watching the information flow between a user and a relying party and the user and an identity provider would also be able to de-anonymize a user in the same manner as they could with OAuth 2.0.

6.4.5 Decentralisation and Blockchains

One possible technical alternative that has been provided is to try to make the contracts between the authorization provider and identity providers into a form of "machine-enforceable" smart contracts. However, it is difficult completely to make any contract actually enforceable completely by technical means except perhaps automated transfer of funds, as given by Bitcoin. For example, if a contract is broken, although the blockchain may have proof of the contract being signed, it would not have the ability (i.e. compulsion via the threat of state violence) to force the defecting party to the contract to comply. Yet even if blockchain technologies were not used, smart contracts could help enforce the dependence on policy and terms of service given in OAuth-style identity eco-systems by providing some kind of audit log. For example, smart contracts could be recorded for each of the components of an identity eco-system and an audit log could automatically check compliance for each transaction.

Regardless, there has been much excitement generated by the use of blockchain technology as a foundational technology in new kind of identity eco-system whose architecture would differ radically from OAuth and WebID systems. For example, one could imagine that transactions of identity attributes happen in a peer-to-peer manner similar to Bitcoin, without the entire server infrastructure of relying parties and identity providers. However, such a system could easily happen

without a blockchain infrastructure, which would be mostly useful for logging the transactions of identity attributes in some privacy-friendly fashion, such as by hashing them. Pentland and other researchers have begun to explore these kinds of techniques.[599]

Yet advocates of the blockchain forget that blockchain technologies are inherently public audit logs of transactions, and thus fail to have privacy and anonymity built-in, having at best only a weak form of pseudonymity. So, if a blockchain was used for the transfer of identity attributes, one would risk the fact that each of your relying parties would be revealed (similar to how the identity provider is a weak link in WebID and OAuth 2.0, but with blockchains every user would know the pattern of relying party access). So if a user was visiting anything from a shopping site to a politically sensitive site, all of these transactions for identity attributes could be revealed.

This does not mean that blockchain technology is inappropriate for identity attributes. Auditing logs based on Merkle trees could be exceedingly important in the future to guarantee that certain transactions took place, a role that is provided by "hubs" and UMA providers in a unsatisfactory manner. So a user, as well as a regulatory authority such as a Data Protection Authority, could then use this audit log to verify and track their identity attribute transactions. However, rather than push for the extreme distributed version of decentralization where every user has their own audit log on their own device and shares a single blockchain, which would suffer from the aforementioned privacy concerns, instead a user could opt for a number of small number of private server-side audit logs based on blockchain technology (i.e. Merkle trees in particular) with a variety of network perspectives and then use these to determine if their identity attribute transfer has been logged correctly in a privacy-preserving manner. A similar approach to this has already been proposed for public key logging in the CONIKS,[600] and a similar proposal could be an important missing technical piece of the identity eco-system, as it would allow both user-centric auditing of identity transactions and a trail of proof for regulatory powers who were trying to enforce conformance to whatever certifications ruled an identity eco-system or binding regulations such as the revised Data Protection Directive.

6.4.5 Conclusions and Next Steps

In terms of identity frameworks, there are several large issues. First, none of the existing identity frameworks provide anonymity from a global passive attacker, and so all of the frameworks can lead to information being leaked via traffic analysis. While it seems this could be theoretically addressed by mix networking-based approaches, practical mix networking libraries are still under development. However, virtually no system without a solid base in mix-networking can prevent some form of global passive adversary from linking attributes to identities, yet as global passive adversaries such as NSA bulk data collection monitoring should be countered, they are still nonetheless hard to practically defend against. What is more worrisome is that the primary mode of authorization, OAuth and variants like OpenID Connect, fail to address authentication and do not provide any user-centric security or privacy measures. In summary, if an identity provider acts maliciously to compromise user privacy and security, there is little a user can do. To make matters worse, "hubs" as put forward by GOV.UK and the UMA specification only make matters worse, as these hubs defeat reasonable attempts at security and privacy by monitoring all user transactions regardless of the identity provider and can often, at least in terms of implementation, also monitor the content. In effect, all of these systems act as "trusted third parties" despite the fact that it is often in their commercial best interest to monitor users behaviour. Inside of Europe only legal restrictions on the identity providers in terms of enforcement of Data Protection regulations may be necessary but are

far from sufficient, as authorization providers or other "hubs" have no technical oversight and human oversight in the form of certification or inspection is difficult given the vast number of identity transactions, although they could be focused on large national or corporate identity providers. In order to prevent any legal rules from being trivially undermined by malicious identity providers, establishing an audit-trail of identity transactions via blockchain systems such as CONIK may at least provide a way for auditors to determine if regulations have been followed.

Technically, the solution is to provide end-to-end encryption between a user's attribute on an identity provider and relying parties, where the user – not the identity provider – controls the key material needed to unlock the relying parties. However, currently no technical process exists to do this. WebID+TLS relies on a cryptographically flawed authentication mechanism with known attacks, although the future W3C work on Web Authentication based on the current work of the FIDO Alliance should allow key material to be used authenticate with a high-level of security. Using these kinds of techniques or alternative zero-knowledge proof techniques[601] such as Secure Remote Password,[602] the user may maintain control over some secret key material that can encrypt their data in transit and sign transactions from their client device without giving a third-party identity provider complete control over their data. All tokens should be signed and access to key material can be considered as a capability.[603] Still, even with these two security considerations in place, there is no way that current identity frameworks do not stop the movement of attributes between identity providers and relying parties without legal restrictions. In order to defend user privacy, basic procedures as outlined by Brandão et al. can be put into place to allow both identity providers, relying parties, and (if necessary) hubs to not be able to trivially link a user's identity to their attribute flow to relying parties and so allow a degree of anonymization in identity systems (2015). If these technical recommendations are followed by future developers of identity eco-systems and has a mutually beneficial relationship with a legal framework that upholds rights such as that of data protection, rather than be a honey-pot for surveillance, an identity eco-system that is truly user-centric and based on the fundamental right of autonomy of data can be established by Europe.

Endnotes

[2] http://www.w2spconf.com/2012/papers/w2sp12-final6.pdf
[3] https://idcubed.org/chapter-15-necessity-standards-open-social-web/
[4] http://www.w3.org/2011/08/webidentity-charter.html
[5] http://dcentproject.eu/wp-content/uploads/2014/01/D4.1-State-of-the-Art_new_2.pdf
[6] Online_Voting:_What_do_we_need_to_have_happen_in_"identity"_before_online_voting_happens%3F
[7] Sullivan, C.L., 2010. *Digital Identity, an Emergent Legal Concept*, University of Adelaide Press.
[8] http://www.binarytattoo.com/about-us/
[9] http://www.xperthr.co.uk/blogs/employment-tribunal-watch/2013/04/social-media-10-employment-cases-involving-facebook/
[10] https://medium.com/matter/actually-it-s-about-ethics-in-doxxing-1651b3deac77
[11] http://www.forbes.com/sites/deannazandt/2012/10/16/the-tyranny-of-anonymity-reddit-and-the-future-of-your-body-online/
[12] http://journals.uoc.edu/index.php/idp/article/viewFile/n17-hildebrandt/n17-hildebrandt-en
[13] DISCLAIMER: ORG is a partner on this project. http://www.horizon.ac.uk/My-Life-in-Data
[14] https://www.schneier.com/blog/archives/2006/01/kevin_kelly_on.html
[15] http://www.camera.it/application/xmanager/projects/leg17/attachments/upload_file/upload_files/000/000/194/Internet_Libe_inglese.pdf
[16] http://www.theguardian.com/uk-news/2015/jan/12/david-cameron-pledges-anti-terror-law-internet-paris-attacks-nick-clegg
[17] https://www.openrightsgroup.org/assets/files/pdfs/reports/digital-surveillance.pdf
[18] http://www.liberosapere.org/iie-2013.pdf
[19] https://deutschebank.nl/nl/docs/DB_Research_Big_data_the_untamed_force_May_2014.pdf
[20] http://www.oecd.org/sti/ieconomy/46968784.pdf
[21] https://en.wikipedia.org/?title=Machine_learning
[22] http://www.technologyreview.com/featuredstory/520426/the-real-privacy-problem/
[23] http://www.datajustice.org
[24] http://www.faz.net/aktuell/feuilleton/debatten/the-digital-debate/shoshan-zuboff-on-big-data-as-surveillance-capitalism-13152525.html
[25] Agar, J. (2003). The Government Machine. Mit Press.
[26] http://www.huffingtonpost.com/edwin-black/ibm-holocaust_b_1301691.html
[27] Pridmore, J., & Zwick, D. (2012). The Rise of the Customer Database, 1–11.
[28] http://www.experian.co.uk/marketing-services/products/mosaic/mosaic-in-detail.html
[29] http://faculty.uml.edu/sgallagher/harvard__law_review.htm
[30] Cohen, J.E., 2012. What privacy is for. *Harv L Rev*, (126), pp.1904–1933.
[31] ibid. p. 1905
[32] http://www.nist.gov/nstic/NSTIC-FIPPs.pdf
[33] Quoted in http://www.dataprotection.eu/pmwiki/pmwiki.php?n=Main.SecondGeneration
[34] http://www.liberosapere.org/iie-2013.pdf
[35] Bria et al. 2013
[36] Zuboff,S.,Big Other: Surveillance Capitalism and the Prospects of an Information Civilization, April 4, 2015 Journal of Information Technology (2015) 30, 75–89. doi:10.1057
[37] Bowcott, O., GCHQ spied on Amnesty International, tribunal tells group in email. Available at: http://www.theguardian.com/uk-news/2015/jul/01/gchq-spied-amnesty-international-tribunal-email [Accessed July 2, 2015].
[38] http://www.nytimes.com/2014/01/28/world/spy-agencies-scour-phone-apps-for-personal-data.html?_r=0
[39] Mayer-Schönberger, V. & Cukier, K., 2013. *Big Data*, Houghton Mifflin Harcourt.
[40] House, W., 2014. *Big Data: Seizing Opportunities, Preserving Values*, Exceutive Office of the President. http://www.whitehouse.gov/issues/technology/big-data-review
[41] Dhar, V., 2013. Data science and prediction. *Communications of the ACM*.
[42] Zaki, M.J. & Wagner Meira, J., 2014. *Data Mining and Analysis*, Cambridge University Press.
[43] Conway, D. & White, J., 2012. *Machine Learning for Hackers*, "O'Reilly Media, Inc.."
[44] Article 29 is a joint collaboration of EU privacy watchdogs to agree policy lines.
[45] Article 29 Working Party on Data Protection, Opinion 8/2014 on Recent Developments on the Internet of Things (Sept. 16, 2014) (p. 7)
[46] http://opendatahandbook.org/guide/en/what-is-open-data/
[47] http://opendatatoolkit.worldbank.org/en/
[48] https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex
[49] US Federal Trade Commission, 2014. *Data brokers: A call for transparency and accountability*,
[50] D Haggerty Richard V Ericson, K. (2000). The surveillant assemblage. *British Journal of Sociology*, *51*(4), 605–622. http://doi.org/10.1080/00071310020015280
[51] Bauman, Z., & Lyon, D. (2013). Liquid Surveillance. John Wiley & Sons.
[52] Amoore, L. (2013). The Politics of Possibility. Risk and Security Beyond Probability. Duke University Press
[53] ibid.
[54] For a good overview of anthropological concepts of distributed identity see: Budja, M. (2010). The archaeology of death: from'social personae'to'relational personhood. *Documenta Praehistorica*.
[55] https://medium.com/@AntonioCasilli/four-theses-on-digital-mass-surveillance-and-the-negotiation-of-privacy-7254cd3cdee6
[56] Abelson, H., & Lessig, L. (1998, December 10). Digitial Identity in Cyberspace. Retrieved May 1, 2015, from http://groups.csail.mit.edu/mac/classes/6.805/student-papers/fall98-papers/identity/linked-white-paper.html
[57] http://europe-v-facebook.org/EN/en.html
[58] https://epic.org/algorithmic-transparency/
[59] Hardt, M., How big data is unfair. Medium. Available at: https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de [Accessed June 24, 2015].
[60] ibid.
[61] Barocas, S. & Selbst, A.D., 2014. Big Data's Disparate Impact. *Available at SSRN 2477899*.
[62] ibid. (p. 22)

[63] Mayer-Schönberger, V. & Cukier, K., 2013. *Big Data*, Houghton Mifflin Harcourt.
[64] Harper, T., Thieves beware: police IT may predict crime | The Sunday Times. *thesundaytimes.co.uk*. Available at: http://www.thesundaytimes.co.uk/sto/news/uk_news/Society/article1565712.ece [Accessed June 24, 2015].
[65] http://balkin.blogspot.co.uk/2015/06/the-party-as-database.html
[66] http://www.behaviouraldesignlab.org
[67] http://quantifiedself.com
[68] Morozov, E., 2013. *To Save Everything, Click Here*, PublicAffairs.
[69] Barocas, S. & Nissenbaum, H., 2014. Big data's end run around procedural privacy protections. *Communications of the ACM*, 57(11), pp.31–33.
[70] http://quantifiedself.com/2013/02/how-to-download-fitbit-data-using-google-spreadsheets/
[71] http://mobihealthnews.com/34847/employer-gets-280k-insurance-discount-for-using-fitbits/
[72] Pasquale, F., The Algorithmic Self. *The Hedgehog Review*. Available at: http://www.iasc-culture.org/THR/THR_article_2015_Spring_Pasquale.php [Accessed June 24, 2015].
[73] Morozov 2013, To Save everything Click here: The Folly of Technological Solutionism. London, Allen Lane.
[74] Sunstein, Cass, and Richard Thaler. "Nudge." *New Haven, CT: Yale University Press."Ten Examples of Early Tortoise-Shell Inscriptions." Harvard Journal of Asiatic Studies* 11 (2008): 1-2.
[75] http://www.newscientist.com/article/mg21228376.500-nudge-policies-are-another-name-for-coercion.html
[76] OECD, 2013. T*he Digital Economy. Hearings on the Digital Economy held at the Competition Committee sessions of October 2011 and February 2012*. Retrieved from: www.oecd.org/daf/competition/The-Digital-Economy-2012.pdf
[77] Íbid.
[78] McKinsey Global Institute, 2011. *Internet matters: The Net's sweeping impact on growth, jobs, and prosperity*. Retrieved from: www.g8.utoronto.ca/summit/2011deauville/eg8/eg8.mckinsey.pdf
[79] http://fortune.com/global500/
[80] PwC, 2015. *The Risers and Fallers. Global Top 100 Companies*. Retrieved from: http://www.pwc.com/gx/en/audit-services/capital-market/publications/assets/image/global-top-100-infograph-2015.jpg
[81] Castells, M., 2004. "Informationalism, Networks, and the Network Society: A Theoretical Blueprint" in Castells, M. (Ed.), The network society: a cross-cultural perspective, Northampton, MA: Edward Elgar.
[82] Castells, M. 1996 (second ed. 2010). *The Rise of the Network Society, The Information Age: Economy, Society and Culture Vol. I.* Oxford, UK: Wiley-Blackwell. (P. 17)
[83] Carlsson, B. The Digital Economy: what is new and what is not?, *Structural Change and Economic Dynamics*, Volume 15, Issue 3, September 2004, Pages 245-264.
[84] OECD, 2013, *Op. Cit.*
[85] Lee, Chung-Shing, 2001. "An analytical framework for evaluating e-commerce business models and strategies", *Internet Research*, Vol. 11 Iss: 4, pp.349 - 359
[86] http://drpeering.net/FAQ/Who-are-the-Tier-1-ISPs.php
[87] Zimmermann, Hans-Dieter, 2000. Understanding the Digital Economy: Challenges for New Business Models (August 1, 2000). AMCIS 2000 Proceedings. Paper 402. Available at SSRN: http://ssrn.com/abstract=2566095 or http://dx.doi.org/10.2139/ssrn.2566095
[88] OECD, 2013, *Op. Cit.*
[89] Gawer, 2009
[90] Gawer ed. 2009
[91] Boudreau, K., A. Hagiu. 2009. Platform Rules: Multi-Sided Platforms as Regulators. A. Gawer, ed. Platforms, Markets and Innovation. Edward Elgar, London, UK
[92] Sundararajan, A. Economics of IT. [Personal website]. Available at: http://oz.stern.nyu.edu/io/network.html [Accessed July 11, 2015]
[93] Arthur, C. 2013, August 2. How low-paid workers at 'click farms' create appearance of online popularity. *The Guardian*. Available at: http://www.theguardian.com/technology/2013/aug/02/click-farms-appearance-online-popularity [Accessed: July 11, 2015]
[94] Arthur, C. 2013, August 2. Facebook Is Riddled With Click Farms Where Workers Sit In Dingy Rooms, Bars On The Windows, Generating 1,000 Likes For $1. *Business Insider*. Available at: http://www.businessinsider.com/how-low-paid-workers-at-click-farms-create-appearance-of-online-popularity-2013-8?IR=T#ixzz3WghzNNf4 [Accessed: July 11, 2015]
[95] EMC Corporation, 2011, June 28. Worl' Data More Than Doubling Every Two Years—Driving Big Data Opportunity, New IT Roles. Available at: http://www.emc.com/about/news/press/2011/20110628-01.htm [Accessed: July 11, 2015]
[96] Carson, C. 2014, November 18. How Much Data Does Google Store? Cirrus Insight. Available at: http://www.cirrusinsight.com/blog/how-much-data-does-google-store [Accessed: July 11, 2015]
[97] Sullivan, D**.** 2012. "Google: 100 Billion Searches per Month, Search to Integrate Gmail, Launching Enhanced Search App for iOS." Search Engine Land. http://searchengineland.com/google-search-press-129925
[98] Vagata, P.; Wilfong, K. 2014, April 10. Scaling the Facebook data warehouse to 300 PB. [Code Facebook blog]: https://code.facebook.com/posts/229861827208629/scaling-the-facebook-data-warehouse-to-300-pb/
[99] Tay, L. 2013, May 10. Inside eBay's 90PB data warehouse. *Itnews for Australian Business*. http://www.itnews.com.au/News/342615,inside-ebay8217s-90pb-data-warehouse.aspx
[100] European Commission, 2010. *The State of the Electronic Identity Market: technologies, stakeholders infrastructure, services and policies*. Available at: http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=3739
[101] Íbid., page 11
[102] van Dijck, J. 2014. Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197-208. http://www.surveillance-and-society.org | ISSN: 1477-7487
[103] International Institute for Analytics, 2014. *Advanced Analytics & Big Data Adoption Report.* Available at: http://iianalytics.com/analytics-resources/market-research/advanced-analytics-big-data-adoption-report
[104] Zittrain, J. 2012, March 21. Meme Patrol: "When something online is free, you're not the consumer, you're the product". *The Future of the Internet.* [blog] http://blogs.law.harvard.edu/futureoftheinternet/2012/03/21/meme-patrol-when-something-online-is-free-youre-not-the-customer-youre-the-product/

[105] Davidson, J., 2014 September 4. Forget Facebook, Meet The Company That Will Pay You For Your Personal Data. *Time*. Available at: http://time.com/money/3001361/datacoup-facebook-personal-data-privacy/ http://www.technologyreview.com/news/524621/sell-your-personal-data-for-8-a-month/
[106] Iyer, G.; Soberman, D.; Villas-Boas, J. M., 2005. The Targeting of Advertising. *Marketing Science*. Vol. 24, No. 3, Summer 2005, pp. 461–476. http://faculty.haas.berkeley.edu/giyer/index_files/tgtadv.pdf

[107] Farahat, A. "How effective is targeted advertising?," American Control Conference (ACC), 2013 , vol., no., pp.6014,6021, 17-19 June 2013 doi: 10.1109/ACC.2013.6580780. P. 111
http://ieeexplore.ieee.org/xpl/login.jsp?reload=true&tp=&arnumber=6580780&url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel7%2F6573242%2F6579790%2F06580780.pdf%3Farnumber%3D6580780
[108] Beales, H. 2010. *The Value of Behavioral Targeting*. Available at: www.networkadvertising.org/pdfs/Beales_NAI_Study.pdf, page 1
[109] ACLU, August 2004. *Combating the Surveillance-Industrial Complex* (Report). https://www.aclu.org/report/combatting-surveillance-industrial-complex
[110] Rifkin, J. 2014. *The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism*. New York: Palgrave/McMillian
[111] Simon, H. A. 1971, "Designing Organizations for an Information-Rich World", in Martin Greenberger, M., *Computers, Communication, and the Public Interest*, Baltimore, MD: The Johns Hopkins Press.
Goldhaber, M. H. 1997, "The Attention Economy and the Net", *First Monday* 2 (4)
Richard A. Lanham, R. A. 2006. *The Economics of Attention: Style and Substance in the Age of Information*. Chicago: University of Chicago Press.
[112] Mayer-Schoenberger, V. and K. Cukier, 2013. *Big Data. A Revolution that will transform how we live, work, and think*. London: John Murray Publishers
[113] Laney, D., 2001. '3D Data Management: Controlling Data Volume, Velocity, and Variety'. *Technical report, META Group*, 6 February. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf
[114] Clarke, R., 1988. 'Information technology and dataveillance'. *Communications of the ACM* 31(5): 498-512.
[115] Degli Esposti, S., 2014. 'When big data meets dataveillance: The hidden side of analytics'. *Surveillance & Society* 12(2): 209-225. http://www.surveillance-and-society.org
[116] Hemp, P., 2009. Death by information overload. *Harvard Business Review*, 87(9), 83-89.
[117] Bott, E. 2014, February 6. Apple, Google, Microsoft: Where does the money come from? *ZDNet* http://www.zdnet.com/article/apple-google-microsoft-where-does-the-money-come-from/
[118] McKinsey Global Institute, 2011, Op. Cit.;
Deloitte, 2013. Data as the New Currency. Deloitte Review. Issue 13. http://d2mtr37y39tpbu.cloudfront.net/wp-content/uploads/2013/07/DR13_data_as_the_new_currency2.pdf
[119] Dryburgh, A. (2014, December 17). Memo To The CEO: Beware The Big Data Trap. *Forbes*. http://www.forbes.com/sites/alastairdryburgh/2014/12/17/memo-to-the-ceo-beware-the-big-data-trap/
[120] McFarlane, G. (2012, November 5). How Does Google Make Its Money? Investopedia. Available at: http://www.investopedia.com/stock-analysis/2012/what-does-google-actually-make-money-from-goog1121.aspx;
Mohan, M. (2014, December 1). How Does Google Make Money? *Minterest*. Available at: http://www.minterest.org/how-does-google-make-money/
Channel 4 (2012, November 27) If Google is free, how does it make so much money? *Channel 4 News* Available at: http://www.channel4.com/news/if-google-is-free-how-does-it-make-so-much-money
[121] http://www.business-management-degree.net/facebook/
[122] Bott, E. 2014. Op. Cit.
[123] Deloitte, 2013, Op. Cit.
[124] Milian, M. (2012, November 15). Data Bartering Is Everywhere- *Bloomberg Business*. Available at: http://www.bloomberg.com/bw/articles/2012-11-15/data-bartering-is-everywhere
[125] Clippinger, J. H. and Bollier, D. (2012). The Social Stack for New Social Contracts. *IDCUBED* [blog] Available at: https://idcubed.org/digital-law/socialstack/
[126] Birch, D., 2014. *Identity is the New Money*, London: London Publishing Partnership
[127] Aspan, M. 2014, March 20. Amazon Becomes Retail Bank Role Model. *American Banker*.
Available at: http://www.americanbanker.com/people/amazon-becomes-retail-bank-role-model-1066419-1.html
[128] Morgan, E. V. (1965). *A history of money* (Vol. 699). London: Penguin books.
[129] Coeckelbergh, M. (2015). *Money Machines. Electronic Financial Technologies, Distancing, and Responsibility in Global Finance*, Surrey/Burlington: Ashgate. p. 21
[130] Morgan, E. V. (1965). Op. Cit.
[131] Coeckelbergh, M. (2015), Op. Cit. p. 24
[132] Coeckelbergh, M. (2015), Op. Cit.
[133] Morgan, E. V. (1965). Op. Cit.
[134] Maurer, B. (2006). *The anthropology of money*. Annu. Rev. Anthropol., 35, p. 20
[135] Bedford, M. (2013). Bitcoin and the age of bespoke silicon. In *Proceedings of the 2013 International Conference on Compilers, Architectures and Synthesis for Embedded Systems (CASES '13)*. IEEE Press, Piscataway, NJ, USA, Article 16. Available at. http://cseweb.ucsd.edu/~mbtaylor/papers/bitcoin_taylor_cases_2013.pdf
[136] Grinberg, R. (2011), Bitcoin: An Innovative Alternative Digital Currency, *Hastings Science & Technology Law Journal*, Vol. 4, p.160. Available at SSRN: http://ssrn.com/abstract=1817857 , p. 160
[137] Bedford, M. (2013), Op. Cit.
[138] McCoy, K. (2015, June 1). Silk Road founder hit with life imprisonment. *Usa Today*. Available at: http://www.usatoday.com/story/money/2015/05/29/ulbricht-silk-road-sentencing/28072247/
[139] Roio et al. (2015). *Design of social digital currency*. Available at: http://www.nesta.org.uk/publications/d-cent-design-social-digital-currency
[140] Rivard, C. L., Rossi, M. A. (2001). *Is Computer Data "Tangible Property" or Subject to "Physical Loss or Damage"?*. Available at: http://www.irmi.com/expert/articles/2001/rossi08.aspx
[141] Cohn, C. (2012, October 31). Megaupload and the Government's Attack on Cloud Computing. *EFF* Available at: https://www.eff.org/en-gb/deeplinks/2012/10/governments-attack-cloud-computing
[142] Lewis, M. (2014, June 25). Lien out: electronic data is not tangible property so no lien arises. *Penningtons Manches*. Available at: http://www.penningtons.co.uk/news-publications/latest-news/lien-out-electronic-data-is-not-tangible-property-so-no-lien-arises/
[143] Tripp, C. (2014, July 18). So you think you own your electronic data, do you? Think again. *Lexology*. Available at: http://www.lexology.com/library/detail.aspx?g=94d3eb7a-f71f-4955-a956-b344b804a6d0
[144] Commission of the European Communities (2005, December 12): *First evaluation of Directive 96/9/EC on the legal protection of databases. Dg internal market and services working paper*. Retrieved from: http://ec.europa.eu/internal_market/copyright/docs/databases/evaluation_report_en.pdf
[145] Evans, Barbara J. (2011) "Much Ado About Data Ownership." *Harv. JL & Tech.* 25: 69.
[146] Hamlin, K. (2011). *Presentation at the NSTIC Privacy Workshop*. Available at: www.nist.gov/nstic/presentations/nstic-privacy_kaliya.pdf

[147] e.g. https://www.personal.com/owner-data-agreement/
[148] http://cyber.law.harvard.edu/projectvrm/VRM_Development_Work
[149] Protalinski, E. (2012, May 3). 13 million US Facebook users don't change privacy settings. *ZDNet* Available at: http://www.zdnet.com/article/13-million-us-facebook-users-dont-change-privacy-settings/
[150] Waugh, R. (2011, November 3). Half of Facebook users 'can't keep up' with site's snooping policies as privacy rules change EIGHT times in two years. *Daily Mail.* Available at: http://www.dailymail.co.uk/sciencetech/article-2057000/Half-Facebook-users-sites-snooping-policies-site-changes-privacy-rules-EIGHT-times-years.html
[151] http://www.nhs.uk/NHSEngland/thenhs/records/healthrecords/Pages/what_to_do.aspx
[152] http://www.nhs.uk/NHSEngland/thenhs/records/healthrecords/Documents/2015/po-patient-faqs.pdf
[153] World Economic Forum 2011 *Personal Data: The Emergence of a New Asset Class*. World Economic Forum.
[154] http://infonomics.ltd.uk
http://api.ning.com/files/S62Cl6*7k9DQgTv8NpuGsHEEo4V0csxmLblZW*TpR*7oCwWrMOCjbEeLKwIfA0qvCn-PcN1bd4rV7SyPGEojneA*Wln2dZ2m/ICforum20110912InfonomicsLaney.pdf
[155] https://en.wikipedia.org/wiki/Asset_classes
[156] http://www.ft.com/cms/s/0/9d2a73fe-a54a-11e3-8070-00144feab7de.html
[157] http://www.theguardian.com/technology/2013/aug/23/tech-giants-data
[158] http://www.consumerfocus.org.uk/blog/personal-data-the-new-oil
[159] http://www.internetsociety.org/blog/tech-matters/2014/10/they-say-"personal-data-new-oil"-thats-good-thing
[160] Machlup, F. (1955) *Characteristics and Types of Price Discrimination, in Business Concentration and Price Policy*, Princeton University Press. Pg. 397.

[161] Shiller, B.R. (2014) *First-Degree Price Discrimination Using Big Data.* Available at http://benjaminshiller.com/images/First_Degree_PD_Using_Big_Data_Apr_8,_2014.pdf
[162] Hannak, A. et al. (2014) *Measuring Price Discrimination and Steering on E-commerce Web Sites*. Available at http://www.ccs.neu.edu/home/cbw/pdf/imc151-hannak.pdf
[163] Whitehouse: Executive Office of the President (2014, May). *Big Data: Seizing Opportunities, Preserving Values*. Available at https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf
[164] World Economic Forum 2011 *Personal Data: The Emergence of a New Asset Class*. World Economic Forum.
[165] D-CENT, D4.1 - State of the Art of social networking systems, identity ecosystem and social data stores, accessed vial http://dcentproject.eu/wp-content/uploads/2014/01/D4.1-State-of-the-Art_new.pdf
[166] Birch, D., 2014. *Identity is the New Money*, London: London Publishing Partnership
[167] http://www.theguardian.com/money/blog/2014/nov/14/airbnb-wont-let-book-room-facebook-friends
[168] https://www.airbnb.co.uk/support/article/450?cref=d311806ec
[169] Pew Research Center, November 2014. "What Internet Users Know About Technology and the Web". Available at: http://www.pewinternet.org/files/2014/11/PI_Web-IQ_112514_PDF.pdf
[170] Castells, M. (1996 [2010]), Op. Cit.
[171] WHO (undated). *Public-Private Partnerships for Health*. Available at: http://www.who.int/trade/glossary/story077/en/
[172] Privacy Rights Clearinghouse (2014). *Fact Sheet 41: Data Brokers and Your Privacy.* Available at: https://www.privacyrights.org/content/data-brokers-and-your-privacy
[173] US Senate - Committee On Commerce, Science, and Transportation (2013). *A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes*. Available at: http://educationnewyork.com/files/rockefeller_databroker.pdf
[174]Crain, M. (2013).*The Revolution Will Be Commercialized: Finance, Public Policy, and the Construction of Internet Advertising*. [Doctoral dissertation] "University of Illinois, Urbana-Champaign"
[175]http://culturedigitally.org/2013/10/reveal-a-little-to-hide-a-lot-acxioms-consumer-data-portal/#sthash.FC0BwqxK.dpuf
[176] Crain, M. (2013, October 8). Reveal a little to hide a lot: Acxiom's consumer data portal. *Culture Digitally*. Available at: https://www.privacyrights.org/content/data-brokers-and-your-privacy
[177] Gary Wolf and Kevin Kelly founded the Blog QuantifiedSelf.com in 2007, giving birth to this term: http://www.quantifiedself.com/
[178] FTC (May 2014). *Data Brokers. A Call for Transparency and Accountability*.
[179] Rud, O. (2009). *Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy*. Hoboken, N.J: Wiley & Sons.
[180] The World Privacy Forum and Privacy Rights Clearinghouse, among others.
[181] Clifford, S., Hardy, Q. (2013, July 14). Attention, Shoppers: Store Is Tracking Your Cell. *The New York Times*. Available at: http://www.nytimes.com/2013/07/15/business/attention-shopper-stores-are-tracking-your-cell.html?_r=0
[182] Hill, K. (2012, February 16). How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did. *Forbes*. http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/
[183] Fung, B. (2013, October 19). How stores use your phone's WiFi to track your shopping habits *The Washington Post*. http://www.washingtonpost.com/blogs/the-switch/wp/2013/10/19/how-stores-use-your-phones-wifi-to-track-your-shopping-habits/
[184] Warman, M. (2013, August 12). Bins that track mobiles banned by City of London Corporation. *The Telegraph*. http://www.telegraph.co.uk/technology/news/10237811/Bins-that-track-mobiles-banned-by-City-of-London-Corporation.html
[185] ACXIOM (Undated). *Response to the European Commission consultation on European Data Protection Legal Framework*. Available at: http://ec.europa.eu/justice/news/consulting_public/0003/contributions/organisations/acxiom_en.pdf
[186] FEDMA (undated). *Ethical Personal Data Management Charter*. http://www.fedma.org/fileadmin/documents/Legal_A_Eth_C/FEDMA_Charter_Ethical_Personal_Data_Management.pdf
[187] FTC (2012, December 18). *FTC to Study Data Broker Industry's Collection and Use of Consumer Data*. https://www.ftc.gov/news-events/press-releases/2012/12/ftc-study-data-broker-industrys-collection-use-consumer-data
[188] Crain, M. (2013, October 8). Op. Cit.
[189] US Senate - Committee On Commerce, Science, and Transportation (2013). *Op. Cit.*
[190] www.acxiom.com
[191] http://www.fedma.org/index.php?id=34&L=2%27%20and%20char%28124%29%20user%20char%28124%29%3D0%20and
[192]www.datalogix.com
[193]www.intelius.com
[194] Some of these items may fall in both categories, depending on the structure and outsourcing level of the companies
[195] https://www.i-behavior.com
[196] Privacy Rights Clearinghouse (2014). Op. Cit.

[197] Van Laeke, M. (2015, May 22). How big data can drive competitive intelligence. *Thoughs on Cloud.* Available at: http://www.thoughtsoncloud.com/2015/05/how-big-data-can-drive-competitive-intelligence/
[198] Ericson, J. (2012, August 12). Big Data As You Go. *Information Management.* Available at: http://www.information-management.com/blogs/Big-Data-As-You-Go-10023096-1.html
[199] WEF (2013). Unlocking the Value of Personal Data:
From Collection to Usage. Available at: http://www.weforum.org/reports/unlocking-value-personal-data-collection-usage
[200] WEF (2011). *Personal Data: The Emergence of a New Asset Class*. p.16 Available at:
http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf
[201] Bachman, K. (2015, March 25). Confessions of a Data Broker Acxiom's CEO Scott Howe explains how self-regulation can work. *AdWeek*. Available at: http://www.adweek.com/news/technology/confessions-data-broker-156437
[202] https://datacoup.com/
[203] Klaassen, A. (2014, March 26,). Acxiom CEO: Rockefeller Data Bill Worse Than Worst Part of Obamacare. *Advertising Age.*
http://adage.com/article/dataworks/acxiom-ceo-rockefeller-bill-worse-worst-part-aca/292333/
[204] WEF (2011). Op. Cit.
[205] Barnes, B. (2013, January 13). At Disney Parks, a Bracelet Meant to Build Loyalty (and Sales). *The New York Times.* Available at:
http://www.nytimes.com/2013/01/07/business/media/at-disney-parks-a-bracelet-meant-to-build-loyalty-and-sales.html?pagewanted=all&_r=0
[206] Ramesh, R. (2015, January 22). NHS disregards patient requests to opt out of sharing medical records. *The Guardian.*
http://www.theguardian.com/society/2015/jan/22/nhs-disregards-patients-requests-sharing-medical-records
[207] L'Hoiry X. & Norris, C. (undated). *Exercising democratic rights under surveillance regimes* (Report Draft). Available at:
http://irissproject.eu/wp-content/uploads/2014/06/IRISS-WP5-Executive-Summary-for-Press-Release.pdf
[208] https://www.similartech.com/technologies/facebook-connect
[209] Article 29 Working Party (undated). *Article 29 Cookie Sweep Results*. Available at: https://ico.org.uk/media/about-the-ico/documents/1043274/a29-cookie-sweep-combined-analysis-report.pdf
[210] Information Management (undated). *Gartner's 19 In-memory Databases for Big Data Analytics* http://www.information-management.com/gallery/In-memory-database-list-gartner-big-data-analytics-10027047-1.html
[211] https://hadoop.apache.org/
[212] International Institute for Analytics (undated). *The Data Lake Debate.* http://iianalytics.com/research/the-data-lake-debate;
PwC (undated). *Data lakes and the promise of unsiloed data* http://www.pwc.com/us/en/technology-forecast/2014/cloud-computing/features/data-lakes.jhtml
[213] Sweeney, L. (2000). *Uniqueness of Simple Demographics in the U.S. Population*. LIDAP-WP4 Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh.
[214] World Privacy Forum (2013, December 18). *Congressional Testimony: What Information Do Data Brokers Have on Consumers?* Available at:
https://www.worldprivacyforum.org/2013/12/testimony-what-information-do-data-brokers-have-on-consumers/
[215] http://lists.nextmark.com/
[216] http://www.listgiant.com/lists
[217] Kunderu, S. (2013, November 25). Demystifying machine learning techniques in forecasting. *Mu Sigma*. Available at: http://www.mu-sigma.com/analytics/thought_leadership/decision-sciences-Demystifying-Machine-learning-in-forecasting.html
[218] Felten, E. (2012, September 12). Accountable Algorithms. *Freedom to Tinker*. Available at: https://freedom-to-tinker.com/blog/felten/accountable-algorithms/
[219] http://www.toptal.com/machine-learning
[220] Hill, K. (2012, February 16). Op. Cit.
[221] Ad Age (2015, May 27). *Even Small Businesses Are Ready for Marketing Automation*. Available at:
http://adage.com/lookbook/article/marketing-automation/small-businesses-ready-marketing-automation/298780/
[222] FTC (May 2014). Op. Cit.
[223] Facebook (2013, April 10). Advertising and our Third-Party Partners. *Facebook and Privacy notes.*
https://www.facebook.com/notes/facebook-and-privacy/advertising-and-our-third-party-partners/532721576777729
[224] https://www.facebook.com/help/133547810119620/
[225] EPIC (undated). *40 Websites Offering Telephone Calling Records and*
Other Confidential Information. Available at: https://epic.org/privacy/iei/attachment_a.pdf
[226] World Privacy Forum (2013, December 18). Op. Cit.
[227] Angwin, J. (2014, January 30). Privacy Tools: Opting Out from Data Brokers. *Propublica.* Available at:
http://www.propublica.org/article/privacy-tools-opting-out-from-data-brokers
[228] World Privacy Forum (2014, January 10). *Data Brokers Opt Out List.* https://www.worldprivacyforum.org/2013/12/data-brokers-opt-out/
[229] Stop Data Mining Opt-out list: http://www.stopdatamining.me/opt-out-list/
[230] Privacy Rights Clearinghouse (2013, March). *Online Data Vendors: How Consumers Can Opt Out of Directory Services and Other Information Brokers*. Available at: https://www.privacyrights.org/online-information-brokers-list
[231] Fraunhofer IAO (2014, September 29). *Sharing Economy in Urban Environments* [Presentation]. Available at:
http://www.iis.fraunhofer.de/content/dam/iis/tr/Session%204_3_Sharing%20Economy%20Urban%20Environments_Fraunhofer%20IAO_Spindler.pdf
[232] Toffler, A. (1980): *The Third Wave*. New York: William Morrow.
[233] http://www.ft.com/cms/s/0/fd3a6246-f46c-11e4-bd16-00144feab7de.html#axzz3eJAcBXzj
[234] For instance, *Kickstarter*
[235] NESTA (April 2015). *Towards an Index of the Collaborative Economy*. Available at:
http://www.nesta.org.uk/sites/default/files/towards_an_index_of_the_collaborative_economy.pdf
[236] Botsman, R. & Rogers, R. (October 2010). Beyond Zipcar: Collaborative Consumption. *Harvard Business Review.* Available at:
https://hbr.org/2010/10/beyond-zipcar-collaborative-consumption/
[237] Leonard, A. (2014, March 14). "Sharing economy" shams: Deception at the core of the Internet's hottest businesses. *SALON*. Available at: http://www.salon.com/2014/03/14/sharing_economy_shams_deception_at_the_core_of_the_internets_hottest_businesses/
[238] Gozzer, S. (2015, May 28). Uber se desmarca de la economía colaborativa en su primer juicio. *El País.* Available at:
http://ccaa.elpais.com/ccaa/2015/05/28/catalunya/1432822456_478741.html
[239] Shontell, A. (2014, November 15). "Uber Is Generating A Staggering Amount Of Revenue". *Business Insider*. Retrieved 25 May 2015.
[240] Walsh, B. (2015, Janyary 26). Here's Why Uber Is Tripling Prices During A State Of Emergency. *Huffington Post.* Available at:
http://www.huffingtonpost.com/2015/01/26/uber-price-surge-blizzard_n_6548626.html

[241] Fiegerman, S. (2014, August 26). Uber Allegedly Engages in Broad 'Sabotage Campaign' Against Lyft. *Mashable*. Available at: http://mashable.com/2014/08/26/uber-competitors-marketing/
[242] Kastrenakes, J. (2014, November 18). Uber Executive Casually Threatens Journalist with Smear Campaign. *The Verge.* Available at: http://www.theverge.com/2014/11/18/7240215/uber-exec-casually-threatens-sarah-lacy-with-smear-campaign
[243] Bhuiyan, J. and Warzel, C. (2014, November 18). "God View": Uber Investigates Its Top New York Executive For Privacy Violations". *BuzzFeed*. Retrieved December 4, 2014.
[244] Halleck, T. (2015, February 27). Uber Confirms Database Breach, Personal Data Of 50,000 Drivers Revealed. *International Business Times*. Available at: http://www.ibtimes.com/uber-confirms-database-breach-personal-data-50000-drivers-revealed-1831576
[245] Blain, L. (2014, November 26). Uber's Android app caught reporting data back without permission. *Gizmag*. Available at: http://www.gizmag.com/uber-app-malware-android/34962/
[246] GironSec (2014, November 25). *What the hell Uber? Uncool bro.* [blog] Available at: http://www.gironsec.com/blog/2014/11/what-the-hell-uber-uncool-bro/
[247] Smith, B. (2014, November 18). Uber Executive Suggests Digging Up Dirt On Journalists. *Buzzfeed News.* Available at: http://www.buzzfeed.com/bensmith/uber-executive-suggests-digging-up-dirt-on-journalists#.fc7baX1ng
[248] Biggs, J. (2014, November 19,). Senator Al Franken Asks Uber's CEO Tough Questions On User Privacy. *TechCrunch*. Retrieved November 20, 2014.
[249] Roderick, L. (2014) Discipline and Power in the Digital Age: The Case of the US Consumer Data Broker Industry. *Critical Sociology*, first published on January 6, doi:10.1177/0896920513501350: 730.
[250] Taylor, A. and Sadowski, J. (2015, May 27). How Companies Turn Your Facebook Activity Into a Credit Score. *The Nation*. Available at: http://www.thenation.com/article/208441/how-companies-turn-your-facebook-activity-credit-score#
[251] Barrett J, Hendricks E, Singleton S, et al. (2001). Panel II: the conflict between commercial speech and legislation governing the commercialization of private sector data. Fordham Intellectual Property, *Media and Entertainment Law Journal* 11(1): 58–95
[252] Federal Reserve Bank of New York, 2012. *Federal Reserve Bank of New York (2012) Household Debt and Credit Report*. Available (consulted 22 June 2015) at: http://www.newyorkfed.org/householdcredit
[253] Manzerolle V. and Smeltzer S. (2011). Consumer databases and commercial mediation of identity: a medium theory analysis. *Surveillance and Society*. 8(3): 323–337. P. 335.
[254] Mierzwinski, E. and Chester, J. (2012, Octobe). Selling Consumers, Not Lists: The New World of Digital Decision-Making and the Role of the Fair Credit Reporting Act. *Suffolk University Law Review*, Forthcoming. Available at SSRN: http://ssrn.com/abstract=2188560
[255] Taylor, A. and Sadowski, J. (2015, May 27). Op. Cit.
[256] Roderick, 2014. Op. Cit.:738.
[257] Singer, N. (2012, August 18). Secret E-Scores Chart Consumers' Buying Power. *The New York Times.* Available at: http://www.nytimes.com/2012/08/19/business/electronic-scores-rank-consumers-by-potential-value.html?_r=0
[258] Ibid.
[259] http://www.ebureau.com/sites/all/files/file/datasheets/ebureau_escore_datasheet.pdf
[260] Singer, N. (2012, August 18). Op. Cit.
[261] http://www.ebureau.com/b2c/credit-risk-assessment
[262] Mierzwinski, E. and Chester, J. (2012, October 1): Op. Cit.
[263] Source: Bluekai http://www.bluekai.com/newsandmedia_pressreleases_20110629.php
[264] Source: Bloomberg Business http://www.bloomberg.com/research/stocks/private/snapshot.asp?privcapId=27128128
[265] Singer, N. (2012, August 18). Op. Cit.
[266] http://www.ebureau.com/privacy-center
[267] Worldwide Public Dataset Catalogs: http://datos.fundacionctic.org/sandbox/catalog/faceted/
[268] European Commission (2014). Factsheet on the "Right to be Forgotten Ruling" (C-131/12). Available at: http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf
[269] Source: GOV.UK. https://www.gov.uk/transformation
[270] Arthur, C. (2014, November 6). Gov.uk quietly disrupts the problem of online identity login. *The Guardian.* http://www.theguardian.com/technology/2014/nov/06/govuk-quietly-disrupts-the-problem-of-online-identity-login
[271] Ibid.
[272] http://oixuk.org/
[273] Hall, K. (2015, March 25). More suppliers join flagging GOV.UK Verify ID assurance scheme. *The Register.* Available at: http://www.theregister.co.uk/2015/03/25/more_suppliers_join_flaggin_govuk_verify/
[274] Government Digital Service (2015, May 14). *The next 6 months: services that plan to start using GOV.UK Verify.* [blog] https://identityassurance.blog.gov.uk/2015/05/14/the-next-6-months-services-that-plan-to-start-using-gov-uk-verify-2/
[275] National Technical Authority on Information Assurance and Cabinet Office (July 2014). *Good Practice Guide No. 45. Identity Proofing and Verification of an Individual.* Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/370033/GPG_45_identity_proofing_v2_3_July_2014.pdf
[276] Cabinet Office (2013, June 17). *Consultation outcome. Privacy and Consumer Advisory Group: Draft Identity Assurance Principles.* Available at: https://www.gov.uk/government/consultations/draft-identity-assurance-principles/privacy-and-consumer-advisory-group-draft-identity-assurance-principles
[277] Hall, K. (2015, June 22). Oi, UK.gov, your Verify system looks like a MASS SPY NETWORK. *The Register*. http://www.theregister.co.uk/2015/06/22/severe_security_flaws_in_verify_mean_backdoor_to_mass_surveillance/?mt=1435687884190
[278] Brandão L. T. A. N. et al., (2015): Toward Mending Two Nation-Scale Brokered Identification Systems, *Proceedings on Privacy Enhancing Technologies*, (2):1–22.
[279] Arthur, C. (2014, November 6). Op. Cit.
[280] European Parliament and the Council, Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 24.10.1995, article 8(1).
[281] Norris, P. (2004): *Building Knowledge Societies: The Renewal of Democratic Practices in Knowledge Societies*, UNESCO Report.
[282] Smith, S. M., Fabrigar, L. R., & Norris, M. E. (2008): "Reflecting on six decades of selective exposure research: Progress, challenges, and opportunities" in *Social and Personality Psychology Compass*, 2, 464-493.
Stroud, N. (2008). "Media use and political predispositions: Revisiting the concept of selective exposure" in *Political Behavior*, 30(3), 341–366.
[283] Howard, P. N. (2005): Deep Democracy, Thin Citizenship: The Impact of Digital Media in Political Campaign Strategy. *The ANNALS of the American Academy of Political and Social Science*, January 2005, 597: 153-170, doi:10.1177/0002716204270139
[284] Rubinstein, I. S. (2014): 'Voter Privacy in the Age of Big Data', *Winsconsin Law Review*, 86: 861-936.

[285] Nickerson, D. W. and Rogers, T. (2014): 'Political Campaigns and Big Data', *The Journal of Economic Perspectives*. Vol. 28, No. 2 (Spring 2014), pp. 51-73
[286] Source: EPIC. https://epic.org/privacy/publicrecords/
[287] http://www.aristotle.com/
[288] Sey, A. & Castells, M. (2004). From Media Politics to Networked Politics: The Internet and the Political Process. In M. Castells (Ed.), The network society: A cross-cultural perspective. Cheltenham, UK; Northampton, MA: Edward Elgar Pub
[289] Castells, M. (2009). *Communication & Power*. Oxford: Oxford University Press.
[290] Ibid.
[291] Reynolds. C. (2012, November 15). Obama and Social Media – 4 Years of Progress. *My Social Agency*. Available at: http://www.mysocialagency.com/obama-social-media-campaign-then-and-now/3759/
[292] Castells (2009). Op. Cit.
[293] Rutledge, P. (2013, January 25). How Obama Won the Social Media Battle in the 2012 Presidential Campaign. *The Media Psychology Blog*. [blog]. Available at: http://mprcenter.org/blog/2013/01/how-obama-won-the-social-media-battle-in-the-2012-presidential-campaign/
[294] Ibid.
[295] Eden, S. (2004, August 24). Democrats Unleash "Demzilla" on the GOP. *Plus Three*. Available at: http://plusthree.com/blog/news/20040824_demzilla/
[296] Evans, L. (2015, February 2). Ben Williamson – Programmable Schools? Governing education through code in the smart city. *The Programmable City Project*. [blog] Available at: http://www.maynoothuniversity.ie/progcity/2015/02/ben-williamson-programmable-schools-governing-education-through-code-in-the-smart-city/
[297] Williamson, B. (2015) Governing software: networks, databases and algorithmic power in the digital governance of public education, *Learning, Media and Technology*, 40:1, 83-105, DOI: 10.1080/17439884.2014.924527
[298] Briggs, S. (2014, January 29). Big data in education: Big potential or big mistake? *Innovation Excellence*. http://www.innovationexcellence.com/blog/2014/01/29/big-data-in-education-big-potential-or-big-mistake/

[299] Gartner, S. (2014, October 6). Moodle will always be an open source project. *Opensource.com* https://opensource.com/education/14/10/open-access-learning-moodle
[300] Slade, S. and Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas, *American Behavioral Scientist* 57, no. 10: 1509–1528.
[301] Simon, S. (2013, March 3). K-12 student database jazzes tech startups, spooks parents. *Reuters*. Available at: http://www.reuters.com/article/2013/03/03/us-education-database-idUSBRE92204W20130303

[302] Bracy, J. (2014, September 26). "Not just for educators: Lessons from InBloom's demise." *The Privacy Advisor*. Available at: https://privacyassociation.org/news/a/not-just-for-educators-lessons-from-inblooms-demise/

[303] Strauss, V. (2013, June 12). An exchange on controversial $100 million student database. *The Washington Post*. Available at: http://www.washingtonpost.com/blogs/answer-sheet/wp/2013/06/12/an-exchange-on-controversial-100-million-student-database

[304] Ibid.
[305] Lecker, W. (2013, May 31). Private data on children must stay that way. *Stamfordadvocate.com*. Available at: http://www.stamfordadvocate.com/news/article/Wendy-Lecker-Private-data-on-children-must-stay-4566834.php

[306] Simon, S. (2013, March 3). Op. Cit.
[307] SchoolBook. (2013, July 23). NYC parent sounds alarm on student privacy. *SchoolBook*. Available at: http://www.wnyc.org/story/307074-what-you-need-know-about-inbloom-student-database/

[308] http://oecdprivacy.org
[309] https://www.huntonprivacyblog.com/2012/03/22/philippines-passes-omnibus-data-protection-law/
[310] http://www.justice.gov/opcl/privacy-act-1974
[311] http://www.hhs.gov/ocr/privacy/
[312] https://epic.org/privacy/vppa/
[313] http://arnesvenson.com/neighborsbook.html
[314] http://hyperallergic.com/200601/artist-who-furtively-photographed-his-neighbors-wins-in-court-again/
[315] http://cyber.law.harvard.edu/archived_content/people/reagle/privacy-selfreg.html
[316] https://cdt.org/blog/ftc-once-again-says-privacy-self-regulation-isnt-enough/
[317] https://www.whitehouse.gov/the-press-office/2012/02/23/we-can-t-wait-obama-administration-unveils-blueprint-privacy-bill-rights
[318] http://www.pcworld.com/article/2892472/privacy-advocates-find-obama-proposal-lacking.html
[319] http://www.nationaljournal.com/tech/obama-s-privacy-bill-of-rights-gets-bashed-from-all-sides-20150227
[320] https://www.nymity.com/products/data-privacy-laws.aspx
[321] http://www.oas.org/dil/data_protection_privacy_habeas_data.htm
[322] http://www.whitecase.com/articles-09302011/
[323] http://www.mckinsey.com/insights/globalization/global_flows_in_a_digital_age
[324] http://www.apec.org/Groups/Committee-on-Trade-and-Investment/~/media/Files/Groups/ECSG/05_ecsg_privacyframewk.ashx
[325] See http://techliberation.com/2010/11/28/mueller's-networks-and-states-classical-liberalism-for-the-information-age/
[326] http://blogs.lse.ac.uk/mediapolicyproject/2015/01/15/multistakeholderism-unmasked-how-the-netmundial-initiative-shifts-battlegrounds-in-internet-governance/
[327] http://www.zdnet.com/article/u-n-s-itu-pursues-internet-control-again-this-week/
[328] http://www.ohchr.org/EN/Issues/DigitalAge/Pages/DigitalAgeIndex.aspx
[329] http://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=15763&LangID=E
[330] European Data Protection Supervisor. (2014). *Preliminary Opinion of the European Data Protection Supervisor Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*.
[331] http://www.ohchr.org/EN/ProfessionalInterest/Pages/CCPR.aspx
[332] http://www.equalityhumanrights.com/sites/default/files/documents/humanrights/hrr_article_8.pdf
[333] http://en.wikipedia.org/wiki/Member_states_of_the_Council_of_Europe
[334] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2010:083:0389:0403:en:PDF
[335] http://ec.europa.eu/justice/fundamental-rights/charter/index_en.htm

[336] http://conventions.coe.int/Treaty/en/Treaties/Html/108.htm
[337] European Union Agency for Fundamental Rights, **Handbook on European data protection law** (Jan. 2014) http://fra.europa.eu/sites/default/files/fra-2014-handbook-data-protection-law_en.pdf
[338] https://secure.edps.europa.eu/EDPSWEB/edps/EDPS/Dataprotection/QA/QA2
[339] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0010:FIN:EN:HTML
[340] https://www.justice.gov.uk/downloads/human-rights/human-rights-making-sense-human-rights.pdf
[341] https://en.necessaryandproportionate.org/#Footnote%202
[342] http://fra.europa.eu/sites/default/files/fra-2014-handbook-data-protection-law_en.pdf
[343] http://ec.europa.eu/justice/data-protection/
[344] http://www.lobbyplag.eu/lp
[345] https://edri.org/files/DP_letter_Juncker_20150421.pdf
[346] http://www.allenovery.com/publications/en-gb/data-protection/Pages/Timetable.aspx
[347] http://dynamicinsights.telefonica.com/652/telefonica-launches-telefonica-dynamic-insights-a-new-global-big-data-business-unit-2
[348] http://ec.europa.eu/ipg/basics/legal/cookies/index_en.htm
[349] http://www.fieldfisher.com/publications/2011/02/the-impact-of-the-amended-e-privacy-directive-on-e-mail-marketing
[350] http://ec.europa.eu/priorities/digital-single-market/
[351] http://ec.europa.eu/priorities/digital-single-market/docs/dsm-communication_en.pdf page. 15
[352] European Data Protection Supervisor. (2014). *Preliminary Opinion of the European Data Protection Supervisor Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*.
[353] A directive is a type of law that forces member states to create adapted national versions, while a regulation is law that applies straight away across the EU.
See the Modernisation Regulation 1/2003; Merger Regulation 139/2004; Implementing Regulation 1269/2013
[354] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/284422/oft402.pdf
[355] http://www.nabarro.com/downloads/abuse-of-dominance-in-the-eu.pdf
[356] ibid.
[357] http://ec.europa.eu/competition/sectors/pharmaceuticals/inquiry/communication_en.pdf
[358] https://fsfe.org/activities/ms-vs-eu/timeline.en.html
[359] ibid.
[360] http://competitionpolicy.ac.uk/documents/8158338/8256105/CCP+Working+Paper+10-4.pdf/da921167-c51a-451f-87ef-fe8f9b068df6
[361] http://www.ft.com/cms/s/0/79be3a06-d8fa-11e4-b907-00144feab7de.html
[362] European Data Protection Supervisor. (2014). *Preliminary Opinion of the European Data Protection Supervisor Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*.
[363] http://www.monopolkommission.de/images/PDF/SG/SG68/S68_summary.pdf
[364] https://www.democraticmedia.org/content/big-data-gets-bigger-consumer-and-privacy-groups-call-ftc-play-greater-role-data-mergers
[365] https://blog.twitter.com/2015/accelerating-direct-response-advertising-welcoming-tellapart-to-twitter
[366] http://ec.europa.eu/competition/mergers/cases/decisions/m4731_20080311_20682_en.pdf
[367] http://ec.europa.eu/competition/publications/cpn/2008_2_53.pdf
[368] European Data Protection Supervisor. (2014). *Preliminary Opinion of the European Data Protection Supervisor Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*.
[369] http://www.telegraph.co.uk/finance/newsbysector/mediatechnologyandtelecoms/digital-media/11537546/Google-charged-with-monopoly-abuse.html
[370] http://www.theguardian.com/technology/2015/apr/19/google-dominates-search-real-problem-monopoly-data
[371] European Union Agency for Fundamental Rights, **Handbook on European data protection law** (Jan. 2014) http://fra.europa.eu/en/charterpedia/article/38-consumer-protection
[372] TFEU 168
[373] European Data Protection Supervisor. (2014). *Preliminary Opinion of the European Data Protection Supervisor Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*.
[374] European Union Agency for Fundamental Rights, **Handbook on European data protection law** (Jan. 2014) http://fra.europa.eu/en/charterpedia/article/38-consumer-protection
[375] http://beuc.eu/publications/2012-00480-01-e.pdf
[376] https://www.eff.org/pages/reader-privacy-chart-2012
[377] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=URISERV:l32017
[378] http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:32005L0029
[379] http://ec.europa.eu/consumers/consumer_rights/rights-contracts/directive/index_en.htm
[380] http://ec.europa.eu/digital-agenda/en/open-data-0
[381] http://ec.europa.eu/digital-agenda/en/european-legislation-reuse-public-sector-information
[382] http://ec.europa.eu/idabc/en/document/2319/5938.html
[383] https://ec.europa.eu/digital-agenda/en/open-standards
[384] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32001L0029:EN:HTML
[385] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML
[386] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31999L0093:en:HTML
[387] http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52008DC0798
[388] http://eur-lex.europa.eu/legal-content/EN/TXT/uri=uriserv:OJ.L_.2014.257.01.0073.01.ENG
[389] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:105:0054:0063:EN:PDF
[390] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62012CJ0293
[391] http://www.legislation.gov.uk/ukpga/2014/27/contents/enacted
[392] http://www.assemblee-nationale.fr/14/ta/ta0511.asp
[393] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML
[394] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:319:0001:0036:en:PDF
[395] https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Thematic%20Guidelines/14-11-25_Financial_Guidelines_EN.pdf
[396] *http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2010.195.01.0003.01.ENG#L_2010195EN.01000501*
[397] http://amberhawk.typepad.com/files/eu-council-dp-reg-4column-2015_april.pdf
[398] http://privacylawblog.fieldfisher.com/2015/the-eu-dp-regulation-is-on-its-way-but-when

[399] http://eulawanalysis.blogspot.com/2014/10/the-proposed-general-data-protection.html
[400] http://personal-data.okfn.org/expert-workshop/anonymisation/
[401] http://www.mondaq.com/x/300206/Data+Protection+Privacy/Pseudonymisation+The+Benefits+And+Getting+Regulation+Right
[402] http://www.mondaq.com/x/300206/Data+Protection+Privacy/Pseudonymisation+The+Benefits+And+Getting+Regulation+Right
[403] ibid.
[404] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006
[405] http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf
[406] http://www.lexology.com/library/detail.aspx?g=444da6e8-8649-4579-8fe4-9b3a637015b4
[407] http://www.out-law.com/en/articles/2015/february/what-data-protection-reform-will-mean-for-obtaining-customer-consent/
[408] http://www.computing.co.uk/ctg/news/2388446/british-government-lobbies-to-water-down-consent-requirements-in-eu-data-protection-regulation
[409] http://www.janalbrecht.eu/fileadmin/material/Dokumente/Data_protection_state_of_play_10_points_010715.pdf
[410] http://eulawanalysis.blogspot.com/2015/03/basic-data-protection-principles-in.html
[411] https://www.huntonprivacyblog.com/files/2014/04/wp217_en.pdf
[412] https://edri.org/files/DP_BrokenBadly.pdf
[413] ibid.
[414] http://solon.barocas.org/?page_id=200
[415] http://privacylawblog.fieldfisher.com/2012/transparency-at-the-heart-of-the-new-eu-data-protection-regulation
[416] http://amberhawk.typepad.com/files/eu-council-dp-reg-4column-2015_april.pdf page 279
[417] art 17
[418] https://www.openrightsgroup.org/blog/2014/landmark-ruling-by-european-court-on-google-and-the-right-to-be-forgotten
[419] art 18
[420] http://amberhawk.typepad.com/files/eu-council-dp-reg-4column-2015_april.pdf page 292
[421] https://www.gov.uk/government/news/the-midata-vision-of-consumer-empowerment
[422] recitals 59 and 125
[423] Section 4 articles 19-20
[424] http://www.ceps.eu/sites/default/files/Summary%20-%20The%20General%20Data%20Protection%20Regulation.pdf
[425] https://www.hunton.com/files/Publication/e148d184-7b15-4e62-b295-0febc750f64d/Presentation/PublicationAttachment/a04eeb85-4b86-4034-a7ca-1ed5c3f50c56/Hunton_Williams_EU_Regulation_Guide_Overview.PDF
[426] http://www.out-law.com/en/articles/2015/february/eu-data-protection-reforms-the-implications-for-profiling/
[427] https://edri.org/files/DP_BrokenBadly.pdf
[428] Article 23
[429] http://www.statewatch.org/analyses/no-264-march-15-data-protection-reg.pdf p. 66
[430] https://edri.org/files/DP_BrokenBadly.pdf
[431] http://www.taylorwessing.com/globaldatahub/article_impact_draft_regulation_data_transfers.html
[432] http://www.janalbrecht.eu/fileadmin/material/Dokumente/Data_protection_state_of_play_10_points_010715.pdf
[433] http://www.consilium.europa.eu/en/press/press-releases/2015/03/13-data-protection-council-agrees-general-principles-and-one-stop-shop-mechanism/
[434] https://edri.org/files/DP_BrokenBadly.pdf
[435] http://eulawanalysis.blogspot.com/2015/03/when-super-regulators-fight-one-stop.html
[436] http://www.janalbrecht.eu/fileadmin/material/Dokumente/Data_protection_state_of_play_10_points_010715.pdf
[437] Article 23(2) of Council Regulation No 1/2003.
[438] Articles 73-76
[439] https://edri.org/files/DP_BrokenBadly.pdf
[440] http://www.euroscientist.com/blowback-spy-scandal-threatens-european-research/
[441] http://www.pharmatimes.com/article/14-07-27/EU_data_plans_could_make_cancer_research_impossible.aspx#ixzz3ZTNkjYrP
[442] http://www.biostorage.com/blog-posts/2015-eu-data-protection-regulation/
[443] http://www.datasaveslives.eu/news-resources/council-of-ministers-one-step-closer-to-agreeing-position-on-eu-data-protection-regulation/
[444] http://www.genewatch.org/uploads/f03c6d66a9b354535738483c1c3d49e4/Data_Protection_or_exploitation_fin.pdf
[445] Douglas, S. and Guback, T. (1984). Production and technology in the communication/information revolution. *Media, Culture and Society,* 6, pgs. 233-245..
[446] Ceyhan, A. (2006). 'Téchnologie et sécurité: une gouvernance libérale dans un context d'incertitudes'. *Cultures & Conflits*, 64 (hiver=, 2008, pgs. 11-32.
[447] Hallinan and Friedewald (2012) IRISS, Deliverable 1.1. Surveillance, fighting crime and violence. Pg. 237. Available at: www.irissproject.eu.
[448] Raab, C., Hallinan, D., Amicelle, A., Galdon-Clavell, G., De Hert, P., Galletta, A. and Jones, R. (2013). "Impacts of surveillance on civil liberties and fundamental rights". In *Surveillance, Fighting Crime and Violence.* Deliverable 1.1 of FP7 IRISS Project. Pgs. 254-302. Pg. 237. Available at www.irissproject.eu.
[449] Íbid.
[450] Stewart, M.G. and Mueller, J. (2011). *Risk and Cost-Benefit Analysis of Advanced Imaging Technology Full Body Scanners for Airline Passenger Security Screening*, Research Report 280.11.2101, Newcastle: The University of Newcastle (Australia),
[451] EDPS (2012). *Report on the Commission Recommendation on preparations for the roll-out of smart metering systems.*
[452] Graham, S. 'Homeland Insecurities? Katrina and the Politics of 'Security' in Metropolitan America'. *Space and Culture*, 9 (1). 2006, Pgs. 63-67.
[453] Currie, G. and Kerri, M. 'The Limits of a Technological Fix to Knowledge Management: Epistemological, Political and Cultural Issues in the Case of Intranet Implementation'. *Management Learning*, 35(1), 2004, pg. 18.
[454] ENISA (2012). *Study on monetising privacy*. Available at: http://www.enisa.europa.eu/activities/identity-and-trust/library/deliverables/monetising-privacy
[455] Hoehn, J.P. and Randall, A. 1987. A Satisfactory Benefit Cost Indicator from Contingent Valuation, *Journal of Environment, Economics and Management*, 14(3), 226-247
[456] Carson R.T., Mitchell, R.C., Hanemann, W.M., Kopp, R.J., Presser, S., Ruud, P.A., (1992). *A Contingent Valuation Study of Lost Passive Use Values Resulting From the Exxon Valdez Oil Spill, A Report to the Attorney General of the State of Alaska*.

[457] Brookshire, D.S., d'Arge, R.C., Schulze, W.D., and Thayer, M.A. (1981). "Experiments in valuing public goods," in Smith V.K. (Ed.) *Advances in Applied Microeconomics: Volume 1* .Greenwich CT: JAI Press
[458] Acquisti, A., John, L. Loewenstein, G. (2009). "What is privacy worth?", *Twenty First Workshop on Information Systems and Economics* (WISE) December 14-15, Arizona Biltmore Resort & Spa, Phoenix, AZ
[459] Plott, C.R. and Zeiler, K. (2005). "The Willingness to Pay/Willingness to Accept Gap, The 'Endowment Effect,' Subject Misconceptions and Experimental Procedures for Eliciting Valuations," *American Economic Review*, 95.
[460] Rose, E., (2005). "Data Users versus Data Subjects: Are Consumers Willing to Pay for Property Rights to Personal Information?" *Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS '05)*.
[461] Wright, D., and de Hert, P. (eds.) (2012). *Privacy Impact Assessment*, Springer, Dordrecht.
[462] Wright, D., K. Wadhwa, de Hert, P. and Kloza D. (2011). *A Privacy Impact Assessment Framework for data protection and privacy rights.* PIAF project.
[463] Íbid.
[464] Wright, D. and Raab C. (2012). *Conducting a surveillance impact assessment.* LiSS WG4 Workshop Proceedings.
[465] Surden, H. (2007) *Structural Rights in Privacy*, 60 SMU L. Rev. 1605.
[466] Bankston, K.S. and Soltani A. (2014) *Tiny Constables and the Cost of Surveillance: Making Cents Out of* United States v. Jones, 123 Yale L.J. Online 335. Available at http://yalelawjournal.org/forum/tiny-constables-and-the-cost-of-surveillance-making-cents-out-of-united-states-v-jones
[467] Eagle, N. & Sandy Pentland, A., 2006. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4), pp.255–268.
[468] Pentland, A., 2009. Reality mining of mobile communications: Toward a new deal on data. *The Global Information Technology Report 2008–2009*.
[469] Sandy Pentland, A., 2013. Why I'm calling for a New Deal on Data - Think Big - The Innovation Blog. Available at: http://en.blogthinkbig.com/2013/07/31/sandy-pentland-new-deal-on-data-mit/ [Accessed July 2, 2015].
[470] Sandy Pentland, A., 2014. With Big Data Comes Big Responsibility. Available at: https://hbr.org/2014/11/with-big-data-comes-big-responsibility [Accessed July 2, 2015].
[471] Wilhelm-Volpi, L., 2015. The New Deal on Data – A Great Deal to Think About - Blog. Available at: http://www.acxiom.co.uk/new-deal-data-great-deal-think/ [Accessed July 2, 2015].
[472] Ohm, P., 2009. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization.
[473] Cavoukian, A. & Castro, D., 2014. Big Data and Innovation, Setting the Record Straight: De-identification Does Work. *The Information Technology and Innovation* ….
[474] Narayanan, A. & Felten, E.W., No silver bullet: De-identification still doesn't work. Available at: http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf [Accessed July 2, 2015].
[475] http://datacommons.coop/vision/
[476] http://www.marksage.net/2012/10/intent-casting-shifts-balance-of-power.html
[477] http://www.dailymail.co.uk/news/article-2855464/DVLA-pockets-25m-five-years-selling-personal-details-millions-motorists-parking-enforcement-firms.html
[478] http://www.wired.co.uk/news/archive/2014-04/15/shawn-buckles-is-worth-350-euros
[479] https://datacoup.com/docs#payouts
[480] http://www.citizenme.com
[481] http://handshake.uk.com/hs/index.html
[482] http://techcrunch.com/2013/09/02/handshake/
[483] http://www.bloomberg.com/bw/articles/2012-11-15/data-bartering-is-everywhere
[484] http://www.europarl.europa.eu/stoa/webdav/site/cms/shared/2_events/workshops/2014/20141202/Hildebrandt.pdf
[485] For example, in the UK sensitive information can be fairly processed where the individual concerned has made it public. https://ico.org.uk/for-organisations/guide-to-data-protection/key-definitions/
[486] https://ico.org.uk/media/for-organisations/documents/1610/privacy_notices_cop.pdf
[487] Greenwood, D. et al., 2014. The New Deal on Data: A Framework for Institutional Controls. In J. Lane et al., eds. *Privacy, Big Data, and the Public Good*. Cambridge University Press, pp. 192–210.
[488] Hardjono, T., Deegan, P. & Clippinger, J.H., Social Use Cases for the ID3 Open Mustard Seed Platform. *Technology and Society Magazine, IEEE*, 33(3), pp.48–54.
[489] Whitley, E.A., 2013. Towards Effective, Consent Based Control of Personal Data. In M. Hildebrandt, K. O 'Hara, & M. Waidner, eds. *Digital Enlightenment Yearbook 2013*.
[490] John Wilbanks. (2014). Portable Approaches to Informed Consent and Open Data. In: Julia Lane et al. (eds.) Privacy, Big Data, and the Public Good. pp. 234-252. New York: Cambridge University Press.
[491] http://sagecongress.org/WP/wp-content/uploads/2012/04/PortableLegalConsentOverview.pdf
[492] http://sagebase.org/e-consent/
[493] http://arstechnica.com/science/2015/04/inside-apples-researchkit/
[494] http://sagebase.org/bridgeapps/
[495] Anon, 2014. Open : Data : Cooperatives - Synopsis. Available at: http://opendatamanchester.org.uk/2014/09/20/open-data-cooperatives-synopsis/ [Accessed July 1, 2015].
[496] https://thegooddata.org
[497] Anon, TheGoodData | Enjoy your data - Good Data. *thegooddata.org*. Available at: https://thegooddata.org/good-data [Accessed July 1, 2015].
[498] http://datacommons.coop/vision/
[499] Davies, T., Where co-operatives and open data meet… | Tim's Blog. timdavies.org.uk. Available at: http://www.timdavies.org.uk/2012/07/20/where-cooperatives-and-open-data-meet/ [Accessed July 1, 2015].
[500] Hafen, E., Kossmann, D. & Brand, A., 2014. Health data cooperatives - citizen empowerment. - PubMed - NCBI. *Methods of Information in Medicine*, 53(2), pp.82–86.
[501] http://ihco.coop
[502] http://www.munrohealthcentre.co.uk
[503] https://www.patientslikeme.com
[504] Wicks, P., 2014. Could digital patient communities be the launch pad for patient-centric trial design? - PubMed - NCBI. *Trials*, 15(1), p.172.

[505] Anon, Frequently asked questions - Open Research Exchange. *openresearchexchange.com*. Available at: https://www.openresearchexchange.com/faq [Accessed July 1, 2015].
[506] Anon, Privacy. *patientslikeme.com*. Available at: https://www.patientslikeme.com/about/privacy [Accessed July 1, 2015].
[507] Anon, ourhdc.com/mission.html. *ourhdc.com*. Available at: http://ourhdc.com/mission.html [Accessed July 1, 2015].
[508] Grant, P., Patient Ownership – creating the business environment to build the Learning Health System. *ourhdc.com*. Available at: http://ourhdc.com/files/77390820.pdf [Accessed July 1, 2015].
[509] Anon, Healthbank | About. *healthbank.coop*. Available at: https://www.healthbank.coop/about/#sthash.68DOmcaH.dpuf [Accessed July 1, 2015].
[510] https://en.wikipedia.org/wiki/Commons
[511] https://en.wikipedia.org/wiki/Public_good
[512] Rose, E., 2005. Data Users versus Data Subjects: Are Consumers Willing to Pay for Property Rights to Personal Information? *System Sciences*. (p. 1)
[513] http://news.uchicago.edu/article/2014/12/02/university-chicago-establish-genomic-data-commons
[514] https://www.privacybydesign.ca/index.php/about-pbd/7-foundational-principles/
[515] https://www.privacybydesign.ca/content/uploads/2013/01/operationalizing-pbd-guide.pdf
[516] http://www.gsma.com/publicpolicy/privacy-design-guidelines-for-mobile-application-development
[517] https://ico.org.uk/for-organisations/guide-to-data-protection/privacy-by-design/
[518] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/43044/7226-sm-privacy-ia.pdf
[519] https://www.ftc.gov/news-events/press-releases/2012/03/ftc-issues-final-commission-report-protecting-consumer-privacy
[520] http://www.futureofprivacy.org/privacy-by-design/
[521] Kavakli, E. et al., 2006. Incorporating privacy requirements into the system design process S. Gritzalis, ed. *Internet Research*, 16(2), pp.140–158.
[522] https://www.openrightsgroup.org/blog/2013/mobile-data-for-sale-meeting-with-ee-sheds-new-light
[523] Troncoso, C., Güerses, S. & Diaz, C., 2005. Engineering Privacy by Design. *K.U. LeuvenIBBT, ESATSCD-COSIC*. Available at: http://www.cosic.esat.kuleuven.be/publications/article-1542.pdf [Accessed June 4, 2015].
[524] Weitzner, D.J. et al., 2008. Information accountability. *Communications of the ACM*, 51(6), pp.82–87.
[525] https://cdt.org/files/pdfs/P3P_Retro_Final_0.pdf
[526] Nissenbaum, H., 2009. *Privacy in Context*, Stanford University Press.(p. 123)
[527] ibid. p. 160
[528] Barth, A. et al., 2006. Privacy and contextual integrity: framework and applications. *2006 IEEE Symposium on Security and Privacy (S&P'06)*, pp.15 pp.–198.
[529] http://www.theregister.co.uk/2013/08/12/spy_bins_scrapped_from_london_streets/
[530] http://www.wired.co.uk/news/archive/2013-08/09/recycling-bins-are-watching-you
[531] http://www.currybet.net/cbet_blog/2012/05/hacks-hackers-kaveh-memari.php
[532] http://news.nationalpost.com/news/londons-creepiest-startup-forced-to-pull-spy-trash-cans-that-could-track-london-pedestrians-via-smartphones#__federated=1
[533] https://nakedsecurity.sophos.com/2014/01/16/businesses-are-building-shopper-profiles-based-on-sniffing-phones-wifi/
[534] http://www.presenceorb.com/optout.aspx
[535] http://www.wired.co.uk/news/archive/2014-04/22/presence-orb
[536] https://www.epic.org/privacy/survey/
[537] https://ico.org.uk/media/about-the-ico/documents/1431717/data-protection-rights-what-the-public-want-and-what-the-public-want-from-data-protection-authorities.pdf
[538] Acquisti, a. (2010). *The economics of personal data and the economics of privacy*. Retrieved from http://www.heinz.cmu.edu/~acquisti/papers/acquisti-privacy-oecd-22-11-10.pdf
[539] http://memeburn.com/2013/08/most-top-mobile-apps-dont-have-privacy-policies/
[540] https://www.docracy.com/mobileprivacy/
[541] Pinnick, T., 2011. Privacy Short Notice Design. *Truste.com*. Available at: http://www.truste.com/blog/2011/02/17/privacy-short-notice-designpart-i-background/ [Accessed July 4, 2015].
[542] Final HCI Research Report *primelife.ercim.eu*. Available at: http://primelife.ercim.eu/images/stories/deliverables/d4.1.5-final_hci_research_report-public.pdf [Accessed June 13, 2015].
[543] http://security-architect.blogspot.co.uk/2013/12/trust-frameworks-are-second.html
[544] http://openidentityexchange.org/wp-content/uploads/2014/06/respect-trust-framework-v2.pdf
[545] https://openid.net/docs/Open_Trust_Frameworks_for_Govts.pdf
[546] https://www.eduroam.org
[547] https://gds.blog.gov.uk/2013/09/03/identity-assurance-first-delivery-contracts-signed/
[548] https://idcubed.org/open-platform/platform/
[549] https://idcubed.org/home_page_feature/21-top-bitcoin-digital-currency-companies-endorse-new-digital-framework-digital-identity-trust-open-data/
[550] Birch, D., 2014. Identity is the New Money, Do Sustainability
[551] Davis, K., 2012. *Ethics of Big Data: Balancing risk and innovation*, O'Reilly
[552] http://ethics.virt.ch.bbc.co.uk/index.htm
[553] http://www.ethicsguidebook.ac.uk/Key-ethics-principles-15
[554] http://www.data-archive.ac.uk/create-manage/consent-ethics/legal
[555] http://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf
[556] http://ec.europa.eu/research/swafs/index.cfm
[557] http://www.rri-tools.eu
[558] http://www.cse.dmu.ac.uk/~bstahl/index_html_files/2013_RRI_ICT_chapter.pdf
[559] http://ec.europa.eu/research/science-society/document_library/pdf_06/mep-rapport-2011_en.pdf
[560] http://ec.europa.eu/research/science-society/document_library/pdf_06/responsible-research-and-innovation-leaflet_en.pdf
[561] http://www.researchgate.net/profile/Phil_Macnaghten/publication/263662329_Responsible_research_and_innovation_From_science_in_society_to_science_for_society_with_society/links/00b4953ba045ae63e1000000.pdf
[562] http://spp.oxfordjournals.org/content/early/2013/09/19/scipol.sct067.full.pd#sec-9
[563] Hardt, M., How big data is unfair. Medium. Available at: https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de [Accessed June 24, 2015].

[564] Hardjono, T., Deegan, P. & Clippinger, J.H., Social Use Cases for the ID3 Open Mustard Seed Platform. *Technology and Society Magazine, IEEE*, 33(3), pp.48–54.
[565] Chaum 1981
[566] Pfitzmann & Köhntopp, 2001
[567] (Serjantov and Danezis, 2003).
[568] Murdoch and Danezis, 2005
[569] Cavoukian, 2009
[570] de Montjoye et el. 2012
[571] (W3C, 2002
[572] Kagal and Pato, 2010
[573] http://openpgp.org/
[574] http://dspace.mit.edu/bitstream/handle/1721.1/97690/MIT-CSAIL-TR-2015-026.pdf
[575] http://netlab.cs.ucla.edu/wiki/files/shannon1949.pdf
[576] https://otr.cypherpunks.ca/
[577] https://silentcircle.com/scimp-protocol/
[578] https://whispersystems.org/blog/advanced-ratcheting/
[579] https://www.torproject.org/
[580] http://mixminion.net/
[581] http://pomcor.com/2011/10/10/pros-and-cons-of-idemix-for-nstic/#7
[582] Camenisch and Van Herreweghen, 2002)..
[583] https://whispersystems.org/
[584] http://openidentityexchange.org/wp-content/uploads/the-open-identity-trust-framework-model-2010-03.pdf
[585] Note that while IDAP continues to work with Mydex, Mydex is no longer a certified provider. See more info at https://identityassurance.blog.gov.uk/2015/03/25/gov-uk-verify-and-mydex/
[586] http://www.wsj.com/articles/SB876251209242254500
[587] https://nakedsecurity.sophos.com/2011/03/10/connect-me-controversy-continues-have-your-say
[588] Camenisch and Van Herreweghen, 2002)..
[589] https://developer.mozilla.org/en-US/Persona/Protocol_Overview
[590] Hardt, M., How big data is unfair. Medium. Available at: https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de [Accessed June 24, 2015].
[591] (Sakimura et al., 2014).
[592] Hardjono, T., Deegan, P. & Clippinger, J.H., Social Use Cases for the ID3 Open Mustard Seed Platform. *Technology and Society Magazine, IEEE*, 33(3), pp.48–54.
[593] Brandão L. T. A. N. et al., (2015): Toward Mending Two Nation-Scale Brokered Identification Systems, *Proceedings on Privacy Enhancing Technologies*, (2):1–22.
[594] http://codebutler.github.io/firesheep/
[595] Brandão L. T. A. N. et al., (2015): Toward Mending Two Nation-Scale Brokered Identification Systems, *Proceedings on Privacy Enhancing Technologies*, (2):1–22.
[596] https://github.com/linkeddata/SoLiD
[597] Bhargavan et al., 2014
[598] Meyer at al., 2014
[599] http://www.academia.edu/13356778/Decentralizing_Privacy_Using_Blockchain_to_Protect_Personal_Data
[600] (Melara et al. 2014),
[601] The most secure and complex of these zero-knowledge proof techniques have been explored in the PRIMELIFE and ABC4TRUST European Projects, although more work and possibly even regulation is necessary to make them usable by client developers and standardization would be necessary to get them to work inside Web browsers. See https://abc4trust.eu/ for more information.
[602] http://srp.stanford.edu/
[603] Halpin and Cook, 2012

References

Gawer, A. (Ed.). 2011. *Platforms, markets and innovation*. Edward Elgar Publishing.

Bria, F. et al. 2013. Internet as commons or capture of collective intelligence, Internet Identity Ecosystem seminar http://www.liberosapere.org/iie-2013.pdf

Bhargavan, K., Lavaud, A. D., Fournet, C., Pironti, A., & Strub, P. Y. (2014, May). Triple handshakes and cookie cutters: Breaking and fixing authentication over TLS. In *Security and Privacy (SP), 2014 IEEE Symposium on* (pp. 98-113). IEEE.

Brandão, L. T., Christin, N., & Danezis, G. (2015). Toward Mending Two Nation-Scale Brokered Identification Systems. *Proceedings on Privacy Enhancing Technologies*, *2015*(2), 135-155.

Camenisch, J., & Van Herreweghen, E. (2002). Design and implementation of the idemix anonymous credential system. In *Proceedings of the 9th ACM conference on Computer and communications security* (pp. 21-30). ACM.

Cavoukian, A. (2009). Privacy by design: The 7 foundational principles. *Information and Privacy Commissioner of Ontario, Canada*.

Chaum, D. L. (1981). Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, *24*(2), 84-90.

de Montjoye, Y. A., Wang, S. S., Pentland, A., Anh, D. T. T., & Datta, A. (2012). On the Trusted Use of Large-Scale Personal Data. *IEEE Data Eng. Bull.*, *35*(4), 5-8.

Jones, M.,, Bradley, J., and Sakimura, N (2015). JSON Web Token (JWT). https://tools.ietf.org/html/rfc7519

Halpin, H., & Cook, B. (2014). Federated Identity as Capabilities. In *Privacy Technologies and Policy* (pp. 125-139). Springer Berlin Heidelberg.

Hardjono, T. (2014). User-Managed Access (UMA) Core Protocol, draft-hardjono-oauth-umacore-05C; 2012.

Hardt, D. (2012) OAuth 2.0 Authorization Protocol. IETF RFC. https://tools.ietf.org/html/draft-ietf-oauth-v2-31 (2012).

Kagal, L., & Pato, J. (2010). Preserving privacy based on semantic policy tools. *Security & Privacy, IEEE*, *8*(4), 25-30.

Melara, M. S., Blankstein, A., Bonneau, J., Freedman, M. J., & Felten, E. W. (2014). *CONIKS: A privacy-preserving consistent key service for secure end-to-end communication*. Cryptology ePrint Archive, Report 2014/1004.

Meyer, C., Somorovsky, J., Weiss, E., Schwenk, J., Schinzel, S., & Tews, E. (2014). Revisiting SSL/TLS implementations: New Bleichenbacher side channels and attacks. In *23rd USENIX Security Symposium (USENIX Security 14)*. USENIX Association

Murdoch, S. J., & Danezis, G. (2005). Low-cost traffic analysis of Tor. In *Proceedings-IEEE Symposium on Security and Privacy* (pp. 183-195).

Pfitzmann, A., & Köhntopp, M. (2001). Anonymity, unobservability, and pseudonymity—a proposal for terminology. In *Designing privacy enhancing technologies* (pp. 1-9). Springer Berlin Heidelberg.

Sakimura, N., Bradley, J., Jones, M., de Medeiros, B., & Mortimore, C. (2014). Openid connect core 1.0. *The OpenID Foundation*. http://openid.net/specs/openid-connect-core-1_0.html

Serjantov, A., & Danezis, G. (2003). Towards an information theoretic metric for anonymity. In *Privacy Enhancing Technologies* (pp. 41-53). Springer Berlin Heidelberg.

Story, H., Corlosquet, S., and Sambra, A. (2015). WebID+TLS; WebID Authentication over TLS. http://www.w3.org/2005/Incubator/webid/spec/tls/

W3C. (2002). Platform for privacy preferences (P3P), W3C Recommendation. http://www.w3.org/TR/P3P