Health Technol. (2017) 7:335–349
DOI 10.1007/s12553-017-0190-6

ORIGINAL PAPER

Regulation of Big Data: Perspectives on strategy, policy, law and privacy

Pompeu Casanovas[1,2] · **Louis De Koker**[1] · **Danuta Mendelson**[3] · **David Watts**[1,4,5]

Received: 25 September 2016 / Accepted: 6 April 2017 / Published online: 8 May 2017

Abstract This article encapsulates selected themes from the Australian Data to Decisions Cooperative Research Centre's Law and Policy program. It is the result of a discussion on the regulation of Big Data, especially focusing on privacy and data protection strategies. It presents four complementary perspectives stemming from governance, law, ethics, and computer science. Big, Linked, and Open Data constitute complex phenomena whose economic and political dimensions require a plurality of instruments to enhance and protect citizens' rights. Some conclusions are offered in the end to foster a more general discussion.

Keywords Big Data · Linked Data · Regulation · Law · Data protection · Privacy

This article is part of the Topical Collection on *Privacy and Security of Medical Information*

✉ Pompeu Casanovas
pompeu.casanovas@uab.cat

[1] Law and Policy Program: Data to Decisions Cooperative Research Centre and La Trobe Law School, La Trobe University, Melbourne, VIC 3086, Australia

[2] Institute of Law and Technology, Faculty of Law, Autonomous University of Barcelona, Bercelona, 08193 Cerdanyola del Vallès, Spain

[3] Law and Policy to Decisions Cooperative Research Centre, Formerly Chair in Law (Research), School of Law Deakin University, Melbourne, Australia

[4] Office of the Commissioner for Privacy and Data Protection, Level 6, 121 Exhibition Street Melbourne, VIC 3000, Australia

[5] Big Data and Open Data Lead of the UN Special Rapporteur on the Right to Privacy and Global Pulse's Data Privacy Advisory Group (United Nations), New York City, USA

This article contends that the effective regulation of Big Data requires a combination of legal tools and other instuments of a semantic and algorithmic nature. It commences with a brief discussion of the concept of Big Data and views expressed by Australian and UK participants in a study of Big Data use in a law enforcement and national security perspective. The second part of the article highlights the UN's Special Rapporteur on the Right to Privacy interest in the themes and the focus of their new program on Big Data. UK law reforms regarding authorisation of warrants for the exercise of bulk data powers is discussed in the third part. Reflecting on these developments, the paper closes with an exploration of the complex relationship between law and Big Data and the implications for regulation and governance of Big Data.

1 Perspectives on Big Data

This article focuses on the Regulation of Big Data.[1] To frame the topic we will commence with a few brief remarks about the term 'Big Data' before discussing views expressed by participants in a recently-concluded study undertaken by the Law and Policy Program of the Data to Decisions Cooperative Research Centre.

1.1 "Big Data"

'Big Data' is a concept used to label key advances, opportunities and risks in data sciences. Despite its take-up in major policy documents [1–4], the term lacks precise content. It has

[1] The article reflects papers delivered at "Regulation of Big Data", a panel discussion held at Deakin University, Australia, August 3rd 2016. While the support of the Data to Decisions Cooperative Research Centre is acknowledged, the views expressed in this article do not necessarily reflect the views of the Centre or of other members of the Law and Policy Program.

been in use since the 1990s [5], and was initially coined to refer to the rapidly increasing volume of data that presented new opportunities and also needed to be managed [6]. In 2001, Douglas Laney, currently at Gartner,[2] introduced two further elements: the velocity at which it is being developed, as well as the increasing variety of structured and unstructured data [7].

Big Data definitions abound.[3] In their systematic mapping study of Big Data definitions Ossi Ylijoki and Jari Porras[4] found that the three definitional Vs (volume, velocity and variety) were common to many [8]. Some definitions go further by including technical aspects relating to usage of the data, for example analysis or decision-making. Ylijoki and Porras argue that the inclusion of such elements often creates logical tensions, and adds to definitional vagueness.[5] Enhanced analytical capabilities are, however, such a major driver of the Big Data phenomenon that their inclusion in some definitions is understandable.

1.2 "Big Data" in the context of national security and law enforcement

A recent project of the Law and Policy Program of the Data to Decisions Cooperative Research Centre, *Big Data Technology and National Security: Comparative International Perspectives on Strategy, Policy and Law in Australia, the United Kingdom and Canada*, examined the policies, regulatory approaches, processes and strategies used by Australia, the UK and Canada, to balance the management and exploitation of Big Data for national security and law enforcement purposes. The study combined empirical research and doctrinal analysis. The remarks that follow focus on a few high level perspectives drawn from the empirical research led by Professor Janet Chan and Associate Professor Lyria Bennett Moses of UNSW Law.[6]

The design of the empirical inquiry acknowledged that technologists, stakeholders and users might see Big Data technology through different 'technological frames' [9]. These frames are assumptions, expectations and knowledge about a technology [10–12]. Such frames are informed by a range of factors, including skills, knowledge, demographics and personal experiences. Technological frames are also relevant to perceptions of legal and policy responses to Big Data.

The empirical research comprised of interviews conducted with key stakeholders, technologists, and users in each country. The interviews explored the interviewees' understanding of the capabilities and uses of Big Data and their perception of issues and challenges in relation to Big Data as well as perception of existing, proposed and recommended strategies, policies, laws and practices. In total 63 research participants took part in the research project.[7]

A comprehensive discussion of the findings lies beyond the scope of this article. To frame the topic and illustrate some of the perceptions, this discussion is limited to a selection of high level findings relating to the concept of Big Data, the barriers and challenges to using it and some of the risks relating to Big Data.

1.3 "Big Data" – The concept[8]

The definitions of "Big Data" of research participants focused largely on technical and user requirements [14]. Volume was the most frequently mentioned attribute by Australian participants (22/38). Other attributes that were mentioned included its analytical or predictive capacity (13/38) and that it comprises of aggregated or integrated data from different sources (9/38). Interestingly velocity (5/38) and variety (4/38) were mentioned but not as prominently as suggested by Laney's concept of Big Data, discussed in 1.1 [15]. Importantly a number of participants from technical organisations (5/38) viewed "Big Data" as largely a marketing term that captures the current trend of generating and using large volumes of data.

Views of "Big Data" were largely consistent among participants from the UK, Canada and Australia. However, there was much skepticism about the term among operational and policy participants in the UK. Some would prefer the more precise terminologies used in legislation, such as "bulk personal dataset" used in the UK's *Investigatory Powers Act 2016* [14, 15], as we will discuss later (section 3).

1.4 Barriers and challenges to using Big Data

The most important barriers and challenges to the use of Big Data listed by Australian participants were legal and privacy issues (13/37), public acceptance/trust in agencies wishing to use Big Data (11/37) and access to/sharing of data/data silos (9/37). A number of participants also raised the challenge of

[2] Douglas Laney is a VP and Distinguished Analyst with Gartner's Chief Data Officer Research team.

[3] [1]: "There are many definitions of "Big Data" which may differ depending on whether you are a computer scientist, a financial analyst, or an entrepreneur pitching an idea to a venture capitalist. Most definitions reflect the growing technological ability to capture, aggregate, and process an ever-greater volume, velocity, and variety of data."

[4] School of Business and Management, Lappeenranta University of Technology, Finland.

[5] See for example [2]: "Big Data refers to both large volumes of data with high level of complexity and the analytical methods applied to them which require more advanced techniques and technologies in order to derive meaningful information and insights in real time".

[6] The Canadian component was led by Dr. Alana Maurushat of UNSW Law.

[7] The division of participants among the three countries is as follows: 38 participants were from Australia (interviewed from 25 March 2015 to 13 November 2015), 14 were from the UK (interviewed from 24 February 2016 and 18 March 2016) and 11 were from Canada (interviewed from 15 October 2015 to 26 February 2016). For the methodology employed, see [13].

[8] Discussions in sub-sections 1.3, 1.4 and 1.5 are largely based on [14].

obtaining and maintaining technical, human and other resources, for example the lack of technical capacity to manage large volumes of data and competition for the small pool of good analysts [15].

Research participants in the UK reported similar types of barriers to the use of Big Data; for example, they operated under similar resource constraints. These included working with outdated systems, limited processing power and insufficient human resources. UK participants were however on average more sensitive to technical and resource concerns than Australian participants (9/14 vs 8/37). UK participants also appeared more aware than their Australian counterparts of the potential that data may be incomplete or biased.

1.5 Risks of using Big Data

Australian research participants raised privacy (12/38), misuse of data (10/38), misplaced trust in technology and assumptions behind analytics (10/38) and data security (9/38) as the most significant risks of using Big Data. Research participants from operational organisations seemed particularly sensitive to harm to their own organisations (through political and reputational risks, negative public perceptions and information overload, for example by having data that could be used to identify a criminal or prevent a terrorist attach but failing to do so). Those in operational and technical organisations appeared more conscious of misplaced trust in technology than those in policy positions. The variability of identified risks between individuals and organisations suggests that broader awareness of the diversity of risks across sectors would be beneficial.

1.6 Appropriate regulations and policies

A more comprehensive discussion of the findings falls outside the scope of this article. However, the views expressed by the research participants and reflected above provide some insight into the range of views held in relation to Big Data. An improved understanding of the views held by different sectors of society in relation to Big Data will improve the ability to formulate appropriate policies, especially in relation to a matter as sensitive as the use of Big Data to support national security and law enforcement.

We will address this issue in the sections below. We contend that the appropriate regulation of Big Data in the private and public spheres lies beyond the capacity of such traditional legal instruments as constitutional principles, statutes, regulations, and case law.

To be effective in the Web of Data there is an increasing need to complement them with other tools of semantic and algorithmic nature.

2 Regulating Big Data

2.1 Some doubts

In 2015, the United Nations (UN) appointed a Special Rapporteur on the Right to Privacy (SRP). The appointment was in response to the Snowden allegations, and work that was undertaken in their aftermath by the Human Rights Council.

The SRP produced his first report in March 2016 [16].[9] The report identified a number of key themes that require investigatory work under the SRP's mandate. One of these is Big Data and Open Data.

In July 2016, at the SRP's Conference on Privacy, Personality and Information Flows at the New York University Law School, one of the authors of the present article, David Watts, was appointed to lead this part of the SRP's mandate. His key task is to oversee and coordinate the production of a paper on the privacy implications of Big Data, and Open Data, for presentation to the UN General Assembly, and the UN Human Rights Council in late 2017.

As discussed in 1 above, there is no accepted single definition of Big Data: there are many descriptions and these focus mainly on the large and complex datasets that require new architectures to efficiently manage them [19]. The lack of a definition poses a number of conceptual problems for the Big Data theme — how do you go about determining risk when you don't really know what you are measuring or assessing?

To illustrate the problem, it is worthwhile taking an historical perspective. For example, the Domesday Book, compiled in 1086 as a survey of land and chattels over the whole of England, must fall within an eleventh century experience as a Big Data action of its day. So too must the inventory of English abbeys begun by Thomas Cromwell in 1536, the 1933 Prussian census that used 'Hollerith punch cards' and computing machines supplied and maintained by IBM that produced the evidence of religion, which underpinned the Holocaust.[10] The Big Data of today can easily become the little data of tomorrow.

Open Data can be seen as one of the dimensions of Big Data as an input or data source. According to the Open Knowledge International Handbook,[11] it is defined as "data

[9] See the discussion between J. Cannataci, C. Nyst, F. Patel and L. McGregor at Geneva Academy [17], and G. Greenleaf's comments on the Report [18].

[10] In 16 c. England, the visitations commenced in 1535, the inventory powers were granted by Parliament in 1536, and the process might have carried on for a few years. See for a cultural analysis of the ambivalent political roles that lists and cards can play, Werbin [65]. The author traces the history of Big Data "back to the earliest forms of punch cards, sorters and tabulators emerging in the late nineteenth century when these technologies of population control were first developed by Herman Hollerith (founder of IBM) while working at the US Census Bureau ".

[11] Open Knowledge International is a global non-profit organisation "focused on realising open data's value to society by helping civil society groups access and use data to take action on social problems". Cf. https://okfn.org/about/

that can be freely used, re-used and redistributed by anyone — subject only, at most, to the requirement to attribute and sharealike" [20].

Open Data has become a public sector article of faith over the last few years. The asserted policy basis for this is that "governments have a significant amount of data that can be published publicly. Where this data is made available in a machine-readable way, digital services can leverage it to support improved information and service delivery for users".

The SRP has expressed reservations about Open Data:

> At first sight Open Data sounds fine as a concept, a noble and altruistic approach to dealing with data as a common good, if not quite "common heritage of mankind". Who could object to data sets being used and re-used in order to benefit various parts of society and eventually hopefully all of humanity? It is what you can do with Open Data that is of concern, especially when you deploy the power of Big Data analytical methods on the data sets which may have been made publicly available thanks to Open Data policies. [21]

There are now a significant number of data sets that have been released by government in Australia under Open Data policies. One of the most recent and significant was the release of more than 1 billion lines of what is claimed to be de-identified historical health data by the Department of Health [22]. The Department stated that:

> To ensure that personal details cannot be derived from this data, a suite of confidentiality measures including encryption, perturbation and exclusion of rare events has been applied. This will safeguard personal health information and ensure that patients and providers cannot be re-identified.

There is no doubt that better research for the common good, for example population health research, carries with it community benefits. But what are open data's privacy risks and can they be mitigated appropriately? It's puzzling that the Department's announcement did not include the specific details of the nature of the de-identification process it used.

2.2 The SRP's Big Data and Open Data theme

The Big Data/Open Data theme has been divided into a number of areas of inquiry. These include:

- The benefits of Big Data and Open Data
- The associated data protection risks
- The ways that the risks can be managed/mitigated

The focus in this section is on one aspect of risk mitigation, privacy enhancing technology [PET].

Usually PET is used - to refer to that use of technology, which helps achieve compliance with data protection legislation. The rationale for using PETs does not end with privacy. PETs can protect corporate confidential information and intellectual property, as well as other categories of valuable information.

2.3 De-identification

One of the main PETs is de-identification. Privacy law only applies to personal information. If the information no longer falls within the definition of personal information, privacy law no longer applies.

De-identification is one of the most contentious international privacy issues. Its supporters acknowledge that even though no de-identification approach can be guaranteed to be successful all of the time and for all time, robust, and risk-based de-identification processes can provide sufficient protection to comply with privacy laws. They argue that there are no guarantees of anything. Ann Cavoukian[12] and Daniel Castro[13] are two of the most prominent supporters of de-identification [23]. Opponents of de-identification argue that (i) there is no evidence that de-identification works either in theory or in practice and (ii) attempts to quantify its efficacy are unscientific, and promote a false sense of security by assuming unrealistic, artificially constrained models of what an adversary might do [24].

At this stage it is difficult to know who is right and who is not, or whether a binary answer to the de-identification debate is either helpful or useful. Perhaps we need to look at the debate in a more nuanced way, accepting that in some, but not all cases, de-identification might provide acceptable answers. But even so, it is difficult to see where the boundaries lie. It is becoming easier to combine de-identified data with other data sources in ways that increase the risk of re-identification.

2.4 Distributed ledgers

High on the 'hype cycle' is distributed ledger technology of which blockchain technology is a component. A distributed ledger is a consensus of replicated, shared, and synchronized digital data geographically spread across multiple sites, countries, and/or institutions.[14] Data is stored in a continuous

[12] Former Information and Privacy Commissioner for the Canadian province of Ontario serving from 1997 to 2014. She is currently the Executive Director of the Privacy and Big Data Institute at Ryerson University.

[13] Senior Analyst, Information Technology and Innovation Foundation.

[14] See the seminal and influential white paper published in 2008 by the still unidentified author (or authors) under the pseudonym of 'Satoshi Nakamoto' [25] .

ledger but can only be added when the participants reach a validation consensus. More about the validation consensus process is discussed below.

A blockchain takes data or records, and stores them in a block. A simple analogy is the recording of a transaction on a piece of paper. Each block is then 'chained' to the subsequent block using cryptography. The chain of blocks becomes the digital version of a ledger. The ledger can be shared and examined by anyone permitted to do so. The key difference between this process, and a conventional database, is that rules can be set at the transactional level in the block chain, whereas, this does not occur with conventional databases.

In this brief discussion it is not possible to canvass all of the viewpoints about distributed ledgers and blockchains. We are in the midst of an incredible explosion of information about the uses of these technologies: experience suggests that we need to carefully examine and understand their strengths and weaknesses before reaching firm conclusions about their effectiveness.

But we need to note a few issues about privacy impacts and risks at the outset. Blockchains offer new opportunities for individuals to collaborate and to create datasets in a peer network, without a central intermediary. But what if the ledger is controlled by a single entity, or a group of affiliated interests, who control the validation and permissions process? What happens if they exercise their majority powers? Further, what if the data in each block contains personal information – such as your health records, or your recent bankruptcy, or your change of gender? In an open blockchain, chances are this information is available to anyone, and forever. In a closed blockchain, the information will be available to those with the relevant permissions. Although encryption can be used to protect personal information embodied in each block, at some point the permissions and validation processes require that the information be decrypted.

The distributed ledger technology model operates in a way that challenges one of the main information privacy assumptions – that an organisation, whether public or private, collects, uses and discloses personal information, and thus is both accountable and responsible for those activities. Implicit in this is the assumption that a hierarchy exists: that stewardship of personal information can be traced to a source. In a distributed system that relies on a community of membership permissions and validation processes, such an assumption breaks down.

The outlined technologies might have a potential to deliver privacy benefits; however, they need to be better understood and tested before being declared the answer to our privacy dreams.

2.5 Consent technologies

Finally, another body of work is exploring ways in which technology can be used to underpin the core privacy concept of consent, and the collection, use and disclosure of personal information for the purpose for which it was collected. This body of work relies on applying forms of digital rights management – a technology that fell into disrepute after the way it was used by the entertainment industry to "prop-up" its decaying business model – to personal information. It also can, within some implementations, use semantic web principles, to the same end. Essentially, these approaches attach permissions to personal information, and enable automated negotiations between information subjects and information recipients about the collection and subsequent use and disclosure of the subjects' personal information. This type of approach has been advocated by Professor Alex Pentland of MIT, who has supported placing "the individual much more in charge of data that's about them. This is a major step in making Big Data safer and more transparent, as well as more liquid and available, because people can now choose to share data [26]."

Pentland also believes that personal information can be owned by data subjects, that there should be a proprietary right in personal information. This is a view that is expressed fairly often in US privacy discourse, but not elsewhere. While an interesting ideal, there is perhaps less chance of such a proposition becoming a reality. That said, the idea of giving individuals technological tools to negotiate personal information transactions needs to be explored and considered carefully.

The challenge for readers is to understand and scrutinize the technologies and their implications realistically, and from multiple viewpoints. The conduct of the recent Australian census serves as a useful case study.

The Australian Bureau of Statistics (ABS) is responsible for undertaking a periodic census in Australia. In the past, the ABS has not retained, for any significant period of time, names and addresses. For the 2016 census, it decided that it would do so. There are various accounts of the duration of the retention period. They appear to have shifted as public concern grew.

Another decision was that the census should be undertaken primarily online. The ABS gave assurances that the security measures would be the best in the world.

On the day that the census began, the online platform was hacked, and was taken down. Various, and sometimes conflicting explanations were provided. Assurances were offered that no personal information was compromised. But the damage had been done. The trust in, and the reputation of, a hitherto highly regarded Australian institution had been shattered, perhaps irreparably. A variety of inquiries are currently being undertaken to understand "what" happened, and "why" it happened.

The Australian census debacle has shown that Australians care very much about the privacy of their personal information and are unlikely to trust solutions that do not strike the right balance between functionality and protecting their rights.

They are also sceptical of government. Despite the Commonwealth Minister responsible for the census, Michael McCormack, commenting that the census was just like Facebook, and dismissing concerns about the census enabling government to track the population as being "much ado about nothing," [27] the intensity of the public debate indicates that this viewpoint is a contentious one.

The same argument was used in the recent Australian controversy about the retention of bulk telecommunications metadata for law enforcement and national security purposes, when the (then) Head of the Australian Security Intelligence Organisation said:

> Are you arguing that it is OK for Microsoft or Google to profile you in order to sell you a new BMW, or some beauty product, that is alright for them, but it's not alright for the government on a very selective basis to access telecommunications metadata in order to save lives? That to me is a very distorted and worrying argument. [28]

One of the key issues in the data retention debate was whether individuals' web browsing history would be retained for two years. Public trust and confidence was undermined when Australia's first law officer, the Attorney General, George Brandis, was unable to answer this question clearly:

> Brandis: "The web address, um, is part of the metadata."
> Journalist: "The website?"
> Brandis: "The well, the web address, the electronic address of the website. What the security agencies want to know, to be retained is the, is the electronic address of the website that the web user is ..."
> Journalist: "So it does tell you the website?"
> Brandis: "Well, it, it tells you the address of the website." [29]

There is a body of opinion in Australian government – to the effect that 'if you let the private sector do it, then government should be able to do it as well.' Frequently, this is the opinion expressed by those responsible for 'innovation' or 'disruption' agendas, and who see no boundaries to government information sharing.

Our scepticism grows in proportion to the claims about personal information that are made by government that prove to be inaccurate, such as the 'opt-in' promise for the Personally Controlled Electronic Health Records system [30] or claims that the Census had the best security features.

2.6 The path ahead

David Watts was tasked by the UN Special Rapporteur, to be critical, however, also open-minded: the focus will be on evidence-based critical analysis. The approach is to be inclusive and collaborative.

False trade-offs pervade privacy discourse, such as the supposed trade-off between privacy and security. It is important, when considering privacy issues as they pertain to Big Data and Open Data, to avoid simplistic analysis and to take account of all of the risks and benefits of each in formulating a response that respects privacy rights. Let's examine now with more detail application of legal instruments, and how they may change in the future. The two sections that follow are devoted to this objective. We will introduce them through the examination of the *Investigatory Powers Act 2016* (UK).

3 An improved legal framework

Technical developments, including Big Data analytics have profoundly changed the way and the extent to which law enforcement and security agencies access and examine information. However, legislation governing their functions and practices is still grounded in, and focused upon the twentieth century's responses - to the 'then existing' technologies of access to, automated aggregation, processing and distribution of data, including personal data. This is why, governments and legislatures are reconceptualising what role the law can play in controlling and regulating the use and misuse of data. The task is not easy.

What follows describes an improved legal framework for Law Enforcement and Security Agencies in the UK: the *Investigatory Powers Act 2016* (UK) and proportionate controls incorporated into the system of warrants.

3.1 The Investigatory Powers Act 2016 (UK)

In the United Kingdom, the much debated *Investigatory Powers Act 2016* (UK) ch 25 consolidates and streamlines statutory controls and safeguards that were previously contained in statutes governing nine law enforcement and national security and intelligence agencies into a single statute.[15] It systemically embeds a system of warrants based on the tests of proportionality and necessity into provisions about the interception, interference with, acquisition, examination, and management of communications and other digitized data. It also establishes and oversight scheme involving the Investigatory Powers Commissioner and Judicial Commissioners.

The Act, as compared with, for example, Australian and Canadian legislation[16] in this field, provides the most advanced and comprehensive legislative framework of controls

[15] Including *Regulation of Investigatory Powers Act 2000* (UK), *Police Act 1997* (UK), *Justice and Security Act 2013* (UK), *Counter-Terrorism and Security Act 2015* (UK), and *Data Retention and Investigatory Powers Act 2014* (UK).

[16] The United Kingdom, Australia, and Canada belong to the common law family of legal systems; are constitutional monarchies; they are bound by the multilateral United Kingdom – United States of America Agreement for co-operation in signals intelligence, known as Five Eyes.

on access/interception, examination/analysis of the material, and distribution of information from Big Data. The *Investigatory Powers Act 2016* (UK) attempts to deal with the unprecedented volume of accessible data, referred to as "bulk".[17] This "bulk" data, as available to national security and law enforcement agencies, has been conceptualised in terms of a hierarchical order of categories that can be regulated by way of warrants, authorisations and notices. In turn, these legal instruments enabling access, collection and management instruments include criteria for certain general privacy protections based on consideration of:

data access/collection minimisation "(a) whether what is sought to be achieved by the warrant, authorisation or notice could reasonably be achieved by other less intrusive means";

differentiation of protective privacy levels "(b) whether the level of [privacy] protection to be applied in relation to any obtaining of information by virtue of the warrant, authorisation or notice is higher because of the particular sensitivity of that information";

and the tension between two public interests:

"(c) the public interest in the integrity and security of telecommunication systems and postal services, and (d) any other aspects of the public interest in the protection of privacy."[18]

Statutory duties to consider limits on access, collection and distribution of bulk data are context-driven, and not exhaustive. In particular, the statute (s 2(4)) identifies the following additional factors that, depending on context, must be considered:

> "(a) the interests of national security or of the economic well-being of the United Kingdom,
> (b) the public interest in preventing or detecting serious crime,
> (c) other considerations which are relevant to:
> (i) *whether the conduct authorised or required* by the warrant, authorisation or notice *is proportionate*, or
> (ii) *whether it* [conduct] *is necessary to act for a purpose* provided for by this Act,
> (d) the requirements of the *Human Rights Act 1998* (UK), and
> (e) other requirements of public law."

3.2 The "double-lock" mechanism

It is in this specific statutory context that the tests of proportionality and necessity serve as controls for all operational provisions of the Act. They are also central to the special "double-lock" mechanism for the issuance and approval of warrants relating to "bulk" data. The rule is that before the issued warrant becomes operative, it must be approved.

The double-lock mechanism applies to approval of all warrants involving targeted interception, targeted equipment interference, and all "bulk" surveillance measures. They are set out below and apply to:

1. Equipment interference warrants: Part 5 and Part 6, Chapter 3,[19] (their issuance must be *necessary in the interests of national security*)[20];
2. Communication interception warrants: Part 2 and Part 6, Chapter 1;
3. Obtaining communications warrants: Part 3 and Part 6, Chapter 2;
4. Bulk Personal Dataset Warrants (large datasets containing personal information about a wide range of people): Part 7.

Decisions to issue warrants for bulk interception,[21] bulk acquisition,[22] bulk equipment interference,[23] and bulk personal datasets[24] are to be taken personally by the Secretary of State.

The double-lock mechanism works as follows:

1. In order to obtain a relevant warrant, the investigative agency has to provide the *Secretary of State* with materials, (including evidence of *necessity* and *proportionality*) in support of the proposed warrant [31][25];
2. The *Secretary of State* then must apply specific statutory tests of "*necessity*" and "*proportionality*" before deciding whether to issue the warrant.[26]

[17] Provisions of the *Investigatory Powers Act* include not only statutory controls on the issuance and approval of warrants, but also framework for oversight of the access to and gathering of communications and bulk sets of data, their use and management (distribution, retention and destruction).

[18] *Investigatory Powers Act 2016* (UK) s 2(1).

[19] To access data from computers, smartphones etc. by the security and intelligence agencies, law enforcement and the armed forces.

[20] The new *Draft Code of Practice on Equipment Interference* for the security and intelligence agencies identifies the following objectives:

"a) obtain information from the equipment in pursuit of intelligence requirements;

b) obtain information concerning the ownership, nature and use of the equipment with a view to meeting intelligence requirements;

c) locate and examine, remove, modify or substitute equipment hardware or software which is capable of yielding information of the type described in (a) and (b);

d) enable and facilitate surveillance activity by means of the equipment;

"Information" may include communications content, and communications data."

[21] *Investigatory Powers Act 2016* (UK) s 141.

[22] *Investigatory Powers Act 2016* (UK) s160.

[23] *Investigatory Powers Act 2016* (UK) s182.

[24] *Investigatory Powers Act 2016* (UK) s 211.

[25] *Investigatory Powers Act 2016* (UK), s 18 and s 20.

[26] *Investigatory Powers Act 2016* (UK), s 19.

3. The *Secretary of State* (or in the case of a warrant to be issued by the Scottish Ministers, a member of the Scottish Government), is to *personally* make decisions to issue[27] three specific kinds of warrants on behalf of an "intercepting authority",[28] namely *targeted interception warrant, targeted examination warrant*, and *mutual assistance warrant.*

The statutory test of necessity lists grounds that are applicable to each specific category of warrant. For example: to fulfil the *necessity* precondition, a *targeted interception warrant* or *targeted examination warrant*, must be "necessary —

(a) in the interests of national security,
(b) for the purpose of preventing or detecting serious crime, or
(c) in the interests of the economic well-being of the United Kingdom so far as those interests are also relevant to the interests of national security."

A mutual assistance warrant[29] under sections 15(4) and 15(5) will be deemed necessary if —

(a) it is necessary for the purpose of giving effect - to the provisions of an EU mutual assistance instrument, or an international mutual assistance agreement, and
(b) the circumstances appear to the Secretary of State to be equivalent to those in which the Secretary of State would issue a warrant for the purpose of preventing or detecting serious crime.

However, subsection 20(4) specifies that "information which it is considered necessary to obtain is information relating to the acts or intentions of persons outside the British Islands" [i.e., it cannot pertain to people in Britain].[30]

The second part of the "double-lock system" brings in the judiciary. *Judicial Commissioners* must be judges who hold, or have held, high judicial office.[31] They have statutory power *to approve* warrants issued by the Secretary of State.[32] While Judicial Commissioners must apply the (s 19) statutory necessity test already applied by the Secretary of State, their determination regarding the proportionality criterion is different. Section 23(2)(a) of the Act mandates that in considering whether the conduct to be authorised by the warrant is proportionate: "the Judicial Commissioner must apply the same principles as would be applied by a court on an application for judicial review".[33] In other words, with the exception of instances implicating EU treaties, Judicial Commissioners must apply the common law test of proportionality.[34]

3.3 Discussion: Controls and protections

The double-lock mechanism is based on a system of checks and balances involving *executive decision-making* directed to

[27] *Investigatory Powers Act* s 30 Renewal (s 33(9)(b)), notification and major modifications (s 37(3) and s 38) must be personally approved by the Secretary of State, or in the case of a warrant to be issued by the Scottish Ministers, a member of the Scottish Government. Decision to issue warrants to intelligence services are to be taken personally by the Secretary of State or, where relevant, by a member of the Scottish Government Ministers (105); as are decisions involving renewals of warrants (ss 117), major modifications (s 120, s 122).

[28] *Investigatory Powers Act 2016*, s 18: "(1) Each of the following is an "intercepting authority" for the purposes of this Part—
(a) a person who is the head of an intelligence service; (b) the Director General of the National Crime Agency; (c) the Commissioner of Police of the Metropolis; (d) the Chief Constable of the Police Service of Northern Ireland; (e) the chief constable of the Police Service of Scotland [separate warrantry regime]; (f) the Commissioners for Her Majesty's Revenue and Customs; (g) the Chief of Defence Intelligence;
(h) a person who is the competent authority of a country or territory outside the United Kingdom for the purposes of an EU mutual assistance instrument or an international mutual assistance agreement."

[29] Warrants made under the relevant mutual legal assistance treaty to which the United Kingdom is party for the purpose of gathering and exchanging information/data.

[30] *Investigatory Powers Act 2016* (UK) s 20(2)(c) provides that a "targeted interception warrant or targeted examination warrant is necessary "in the interests of the economic well-being of the United Kingdom so far as those interests are also relevant to the interests of national security", but only "if the information which it is considered necessary to obtain is information relating to the acts or intentions of persons outside the British Islands" (s 20 (4).

[31] Appointments of Judicial Commissioners will be made by the Prime Minister after consultation with the Lord Chief Justice of England and Wales, the Lord President of Scotland, the Lord Chief Justice of Northern Ireland, the Scottish Ministers, and the First Minister and deputy First Minister in Northern Ireland.

[32] Judicial Commissioners have the power to approve both, warrants issued by the Secretary of State and those issued by Scottish Ministers under s 23 of the Investigatory Powers Act 2016. (1) In deciding whether to approve a person's decision to issue a warrant under this Chapter, a Judicial Commissioner must review the person's conclusions as to the following matters— (a) whether the warrant is necessary on relevant grounds (see subsection (3)), and (b) whether the conduct that would be authorised by the warrant is proportionate to what is sought to be achieved by that conduct. (2) In doing so, the Judicial Commissioner must— (a) apply the same principles as would be applied by a court on an application for judicial review, and (b) consider the matters referred to in subsection (1) with a sufficient degree of care as to ensure that the Judicial Commissioner complies with the duties imposed by section 2 (general duties in relation to privacy). (3) In subsection (1)(a) "relevant grounds" means— (a) in the case of a decision of the Secretary of State to issue a warrant, grounds falling within section 20; (b) in the case of a decision of the Scottish Ministers to issue a warrant, grounds falling within section 21(4)." The Advocate-General for Scotland (Lord Keen of Elie) [31].

[33] There are two procedural control mechanisms: Section 23(4) of the *Investigatory Powers Act 2016* requires the Judicial Commissioner who refuses to approve a person's decision to issue a warrant to provide written reasons for the refusal; and s 23(5) provides that where "a Judicial Commissioner, other than the Investigatory Powers Commissioner, refuses to approve a person's decision to issue a warrant …, the person may ask the Investigatory Powers Commissioner to decide whether to approve the decision to issue the warrant".

[34] The common law test of proportionality differs from the EU formulation of proportionality. *Lumsdon & Ors, R v Legal Services Board* [2015] UKSC 41; [2015] 3 WLR 121; *Bank Mellat v HM Treasury* (No 2) [2013] UKSC 39, [2013] 3 WLR 179.

the issuance of the warrant, and *judicial oversight* directed to approval of the warrant.

In Chapter 1 of Part 8, the legislation creates the Office of the Investigatory Powers Commissioner[35] whose duties include review (by way of audit, inspection and investigation) the agencies' functions relating to the interception of communications; the acquisition or retention of communications data; the acquisition of secondary data [i.e. metadata] or related systems data; equipment interference (whether under warrants and authorisations or otherwise). The effectiveness of the Investigatory Powers Commissioner's oversight will depend on allocation of resources by successive governments to this Office.

Finally, ensuring the security of the accessed and managed data-sets is placed under *personal responsibility* of the designated Cabinet Minister. Does the statutory notion of Ministerial personal responsibility have practical significance? The answer to this question will depend on judicial construction of the relevant provisions. Section 1(5) of the Act refers to further statutory protections for privacy,[36] including "the common law offence of misconduct in public office", and "elsewhere in the law". It is unclear whether the imposition of personal responsibility on Cabinet Ministers means that they have no immunity from charge of misconduct in public office or a suit for damages at common law by individuals whose reputation or other legal interests have been harmed by a privacy breach.

The control scheme of the *Investigatory Powers Act* increases legal controls over national security and law enforcement access and usage of bulk data. It is doubtful, however, whether they will prove sufficient to adequately control and monitor data processing flows. The legal mechanisms embedded in the Act are based on public law principles, e.g. necessity and proportionality. The "double-lock" mechanism relies on a system of checks and balances between executive decisions and judicial control, reinforced by the "personal responsibility" of the Cabinet Minister.

However, the Act does not directly regulate the nature and use of automated algorithms that aggregate, link, examine, and process (filter, classify, draw inferences from) the "bulk" data.

As detailed as the Act may be, this means that there is a fundamental dimension of the Web of Data that has not been directly addressed. As argued in the next section, the nature of knowledge and information processes does not fit well into the mould of existing legal solutions. The latter are built on the basis of the rule of law and are nationally, culturally, and linguistically bounded. The former, on the contrary, are transnational in nature, and are usually represented within formal languages and methods.

4 Defining digital regulations

4.1 Privacy and Big Data[37]

Quite recently in contemporary societies, Big Data encountered the Semantic Web, and the Internet of Things. In 2015, eight Zettabytes (Zetta = 10^{21}) were generated, which consisted mostly of unstructured (e-mails, blogs, Twitter, Facebook posts, images, and videos) data [34]. Twitter users generate more than half a billion of daily tweets. e-Bay Online Dispute Resolution system alone solves about 80 million disputes annually. Collecting data and producing metadata — i.e., machine understandable information — constitutes the new stage. But metadata can be collected and structured as well. Big Data entails a belief that combining data from multiple sources may lead to better decisions (or not!). The quality of Big Data-made decisions depends on the structure of context, the organisation, principles and criteria applied.

Furthermore, these decisions are the product of their context: Big Data actually builds a new individual and collective *identity*, and a new kind of *hybrid political culture*. With the many sensors within the Internet of Things, and the organization of Smart Cities, our world may be managed and possibly ruled almost automatically - in the short run. Information processing interacts with programs that not only simulate human intelligence, but virtually act like humans.

The term "hype cycles" refers to "graphic representation of the maturity and adoption of technologies and applications" ("industry noise"), indicating where things are moving to.[38]

[35] Under s 227 of the *Investigatory Powers Act 2016*, the complex process of appointment is as follows: "Investigatory Powers Commissioner and other Judicial Commissioners (1) The Prime Minister must appoint— (a) the Investigatory Powers Commissioner, and (b) such number of other Judicial Commissioners as the Prime Minister considers necessary for the carrying out of the functions of the Judicial Commissioners. (2) A person is not to be appointed as the Investigatory Powers Commissioner or another Judicial Commissioner unless the person holds or has held a high judicial office (within the meaning of Part 3 of the *Constitutional Reform Act 2005*). (3) A person is not to be appointed as the Investigatory Powers Commissioner unless recommended jointly by— (a) the Lord Chancellor, (b) the Lord Chief Justice of England and Wales, (c) the Lord President of the Court of Session, and (d) the Lord Chief Justice of Northern Ireland. (4) A person is not to be appointed as a Judicial Commissioner under subsection (1)(b) unless recommended jointly by— (a) the Lord Chancellor, (b) the Lord Chief Justice of England and Wales, (c) the Lord President of the Court of Session, (d) the Lord Chief Justice of Northern Ireland, and (e) the Investigatory Powers Commissioner. (5) Before appointing any person under subsection (1), the Prime Minister must consult the Scottish Ministers. (6) The Prime Minister must have regard to a memorandum of understanding agreed between the Prime Minister and the Scottish Ministers when exercising functions under subsection (1) or (5). (7) The Investigatory Powers Commissioner is a Judicial Commissioner and the Investigatory Powers Commissioner and the other Judicial Commissioners are to be known, collectively, as the Judicial Commissioners".

[36] In the *Investigatory Powers Act 2016* (UK), s 1, Parts 2 to 7 and Part 8; as well as by virtue of the *Human Rights Act 1998* (UK); *Data Protection Act 1998* (UK), s 55 (unlawful obtaining etc. of personal data); *Wireless Telegraphy Act 2006* (UK), s 48 (offence of interception or disclosure of messages); *Computer Misuse Act 1990* (UK), s 1 to 3A (computer misuse offences).

[37] This discussion is based on Casanovas [32] and Casanovas et al. [33].

[38] http://www.gartner.com/technology/research/hype-cycles/

The Gartner Hype Cycle for Emerging Technologies (GHENT) of August 2014 located Big Data on the edge of already known, but, yet non-mature technologies. The Internet of Things occupied the place of Big Data, at the peak of emerging expectations curve. Huge amounts of data are produced daily through smartphones' sensors, automatically sending information - regardless of the will of their owners. Mobile technology outnumbered the use of personal computers in 2008 [35].

All of this has fostered new applied research. The GHENT for 2015 dropped Big Data from its peak curve, and signalled that autonomous vehicles and the Internet of Things are at the peak of inflated expectations. New related technologies with social and economic applications are emerging: among them, *digital dexterity* (employee cognitive ability and social practice seeking digital business success) and *citizen data science.*[39] These are social areas. Society and security are the highlighted emergent fields. The recently delivered GHENT for 2016 confirms these trends, stressing their cognitive side (e.g. perceptual smart machines).[40] Cognitive business [36] and management strategic studies [37] give support to the same idea, as industry aims at reducing latency and data transfer costs.

Mimicking the Hyper Gardner Cycle, Daniel Castro[41] and Alan McQuinn,[42] from the influential *US Information Technology and Innovation Foundation*, wrote an essay published in September 2015 with the title, "*The Privacy Panic Cycle: A Guide to Public Fears about New Technologies"* [38]. It shows the growing tensions between Human Rights lawyers (and the so-called whistle-blowers) on the one side, and government and administrators on the other. While we are not in complete agreement with the aforementioned authors, we agree that they plot the battlefield; at least from the US point of view. Still, conversations are better served, which focus not so much on the tension between *privacy* vs. *innovation,* rather on the necessary balance among personal needs, decisions, fears and wisdoms (liberty), and collective and public risks (security).

Citizens should be informed and, more than that, act upon incorporating self-regulation into practice for collective purposes, because rules and regulations will become increasingly embedded into programs and devices through interactive workflows between humans and machines [39].

[39] See GHENT 2015, ibid. "A data scientist can be defined as a person who creates or generates models that leverage predictive or prescriptive analytics but whose primary job function is outside of the field of statistics and analytics."

[40] http://www.gartner.com/technology/research/hype-cycles/

[41] Vice president at the Information Technology and Innovation Foundation (ITIF) and director of ITIF's Center for Data Innovation.

[42] Research analyst with The Information Technology and Innovation Foundation (ITIF).

4.2 Citizens' experiences

How do we acknowledge and express that these issues matter? How do we, the citizens, experience privacy and security in our everyday lives? Currently, it is by *surprise, by confusion, and by exhaustion*.

Let's consider the following situation. Imagine that you discover that your family tree (which by extension includes your children and yourself) is available on an online family tree service. It is provided by a company that created an online platform, where users can build family trees based on data collected from their relatives. The platform has aggregated data from 1.5 billion family tree profiles. What can you do?

This is one side of the so-called Open Source Intelligence (OSINT) [40]. Collected data could be kept private by design, and by default. But they are not: data of deceased people are located in the public domain. Otherwise the information could not be legally structured, properly linked, and sold. Only when you click on the names of living people, registration is required. This presents a kind of paradox: the public domain - at the service of market' agents. Apparently, they offer a solution, if you are unhappy about 'your inclusion' in their database. However, reversing this situation is time-consuming, painstaking, and costly, and few people choose going along with the consumer opt-out option.

With the presence of Big Data in everyday life, other examples come to mind, the three Vs (volume, velocity, and variety) foster health applications that may collect a continuous flow of physical personal information, opening new possibilities. For example, Barrett et al. predict:

> "In addition to simple monitoring, a more sophisticated programs would include algorithms that provide personalized feedback to assist with behaviour modification at key moments of decision making (e.g., suggesting healthy recipes while the patient is shopping; encouraging exercise at the end of the workday, or giving a personalized warning about location based environmental triggers for asthma). The real-time velocity sets this application of big data apart from traditional public health uses of behavioural or health data." [41].

Who will control the access and reuse of this personal data flow?

4.3 Linked Data[43]

From a semantic point of view, Big Data can in fact be Linked Data, and Open Linked Data, when they are made accessible. Let's go to the "global giant graph" envisaged by Tim Berners-Lee. Since 2007, there has been a DBpedia project linking

[43] This discussion is based on Casanovas et al. [33].

databases according to the best practices and guidelines of the World Wide Web Consortium [W3C], and building a large-scale, multilingual knowledge base by extracting structured data from Wikipedia editions, now existing in 125 languages.[44] As stated by the organisation: "The English version of the DBpedia knowledge base describes 4.58 million things, out of which 4.22 million are classified in a consistent ontology, including 1,445,000 persons, 735,000 places (including 478,000 populated places), 411,000 creative works (including 123,000 music albums, 87,000 films and 19,000 video games), 241,000 organizations (including 58,000 companies and 49,000 educational institutions), 251,000 species and 6,000 diseases.".[45]

References are tied using Semantic Web languages, especially Resource Descriptive Framework [RDF]. The search language is the Protocol and RDF Query Language [SPARQL], currently being drawn 3000 million triples — subject / object / relation in all natural languages — describing some four and a half million objects [42].

In 2011, a sister project Wikidata was launched to be "a free linked database that can be read and edited by both humans and machines". It contains more than 24,263,995 data items that anyone can edit (October 2016) in all Wikimedia languages.[46] Wikidata aims to provide statements given in a particular context. But few databases contain explicit licenses and rule-based provisions to allow data and metadata be regulated through a rights-driven workflow. Nearly 40% do not yet have licenses [43].

How could personal information workflows be controlled? How can data and metadata be *personalised* in a safe manner? Such possibility can exist only by balancing liberty and security, and by embedding regulations in semi-automated systems. Some authors envisage a "ubiquitous pragmatic web" (encompassing humans and multi-agent systems, i.e. emotions, and processing languages) [33, 44].

For this purpose, we should distinguish *semantic metadata* (human or automated annotations added to the content) from *structural metadata*. The latter adds information — about creation, purpose, origin, time, author, location, network, and language and data standards. Metadata is data that refers and describes data. As it is defined by the W3C, it has the feature of being automatable: for the Web, metadata is machine-understandable information, expressible into a programming language.

Besides, we should also distinguish at least three types of languages expressing knowledge: *(i) natural language, (ii) technical (expert) language, (iii) formal language*. Expert language is essential, as rules and norms are usually formulated in natural languages (English, Spanish, French, etc.). Formal language is the only one that machines can understand. All three kinds of languages can be integrated to create a controlled regulatory ecosystem. For example, Creative Commons licenses incorporate a "three layer design" to make them more comprehensible and to facilitate their greater usage —legal code, human readable, machine-readable.[47]

Contextually, it is also important to distinguish *law, the rule of law*, and *the meta-rule of law,* as analytical dimensions that are always pivotal to human-machine-human or machine-human-machine communication. Law refers to rules laid down by official bodies with regulatory and binding powers (such as parliaments and courts). The rule of law protects citizens' rights, restricting the arbitrary exercise of power. When such protections are embedded into formal systems, and represented through the languages of the Web of Data we face the layer of the meta-rule of law. That is, with the development of the web, the rule of law needs to evolve to a meta-rule of law, based on legal knowledge, and incorporating tools to regulate and monitor the semantic and algorithmic layer of the web [32].

Semantic patterns (and ontology design patterns)[48] can be used and reused to create new types of (interactive, hybrid) regulations by design.[49] W3C Open Digital Rights Language (ODRL) supports open publishing, distribution, and consumption of content, applications and services on the Web. This model respects and puts into the users' hands the management of their rights: it requires their explicit permission. Otherwise, moves are forbidden by default.[50]

4.4 Regulatory algorithms

Focusing on categories and content is not the only strategy. Cryptography and Privacy-enhancing Technologies (PETs) have been developed nearly in parallel of Semantic Web approaches. Statistical disclosure control, inference control, privacy-preserving data mining, private data analysis, differential data privacy (privacy-preserving data analysis)[51] are algorithmic techniques applied to large databases using statistical methods [47].[52] However, they are usually designed to

[44] http://wiki.dbpedia.org/Datasets, http://wiki.dbpedia.org/
[45] http://wiki.dbpedia.org/about
[46] https://www.wikidata.org/wiki/Wikidata:Main_Page

[47] https://learn.canvas.net/courses/4/pages/creative-commons-licenses
[48] http://ontologydesignpatterns.org/wiki/Main_Page
[49] These semantic tools are constructed within a cooperative and collective work of knowledge engineering (with end-users' cooperation). Semantics, constructing ontologies and ODP, means eliciting and sharing knowledge, making it explicit. See some examples [45] [46].
[50] See the work by Renato Iannella et al. at https://www.w3.org/community/odrl/
[51] Differential privacy aims to provide means to maximize the accuracy of queries from statistical databases while minimizing the chances of identifying its records.
Cfr. https://en.wikipedia.org/wiki/Differential_privacy
[52] See Dwork [47]: "Differential privacy is a strong privacy guarantee for an individual's input to a (randomized) function or sequence of functions, which we call a privacy mechanism. Informally, the guarantee says that the behaviour of the mechanism is essentially unchanged independent of whether any individual opts into or opts out of the data set. Designed for statistical analysis, for example, of health or census data, the definition protects the privacy of individuals, and small groups of individuals, while permitting very different outcomes in the case of very different data sets".

protect privacy in a slightly different situation than those described above: "it addresses the paradox of *learning nothing about an individual while learning useful information about a population*" [48, 49]. Thus, they try to neutralise risks of Linked Data, protecting private information against linkage attacks, because aggregate statistical information about the data may reveal some information about the individuals. But even if accurate, they are not infallible.[53]

Industry has been developing some algorithmic solutions based on machine learning and Artificial Intelligence as well. The general idea is that it would be better that data never reach corporate servers in the first place. Privacy-sensitive intelligence is being introduced in mobile operating systems for iPhones and iPads [50, 51]. For instance, Apple is implementing "local" differential privacy. Thus, the company never gets hold of the raw data.

4.5 Legal approaches

As described above (section 3), there are some legal steps that regulate possible privacy risks in UK. Yet, it is worth noticing that the USA and Europe are handling Linked and Big Data in a quite different way. It has been suggested that a perceived loss of control of societal developments by governments, inhibit the effective protection of essential values in democratic societies [52].

Europe embraced the General Data Protection Reform (GDPR) five years ago.[54] Legal scholars pointed out the Copernican legal turn of GDPR compared to the previous situation [53]. This means that privacy and data protection is being aligned with the construction of the so-called European digital single market.[55] The set of principles contained in the recent EU Regulation 2016/679 and in the EU Directive 2016/680 extend citizens' protections - covered by the EU Charter of Fundamental Rights (2000). Personal data protection has emerged in Europe as a specific *fundamental right* [54]. Transparency, data minimisation, proportionality, purpose limitation, consent, accountability, data security, rights of access, rights of correction, third country transfers, rights of erasure, can be now *enforced* through economic sanctions and instruments of monitoring and control by EU agencies [55]. This holds for citizens, but it is clear that the new Regulation is mainly aimed at companies handling large amounts of data. Breaches of privacy and data protection provisions will be severely fined.

This is not the case in the USA, where privacy is under the general protection of Courts and three Constitutional Amendments (the First, Fourteenth and especially the Fourth Amendment). Constitutional law is a way of enhancing citizens' rights, in which, technology and privacy at the County, National, and Federal levels, deal with Court rulings, and the fulfilment of Federal conditions (such as the accomplishment of *Fair Information Practices*). Personal data is aligned with the "right to be left alone" in the tradition of Justice Louis Brandeis, and Thomas Cooley's *A Treatise on Law of Torts* (1888),[56] and its dimension of data protection based on Alan Westin's work on databases [57, 58].

Actually, it has been noticed that the reaction of Warren and Brandeis[59] was a kind of legal transplant, an early attempt to incorporate the European tradition into American culture [60].[57] Thus,

"Why do these sensibilities differ? Why is it that French people won't talk about their salaries, but will take off their bikini tops? Why is it that Americans comply with court discovery orders that open essentially all of their documents for inspection, but refuse to carry identity cards? Why is it that Europeans tolerate state meddling in their choice of baby names? Why is it that Americans submit to extensive credit reporting without rebelling?"

In the twenty-first century Big Data, the starting point is the value of datatasets and how much added value metadata brings to data, using data analytics. From this point of view, there is more room for mass surveillance and for data commodification, because markets are allowed to set the rules of the game. Online Services use clickstream data to optimise operations, and the state is collecting and using citizens' data under security laws and Open Source Intelligence (OSINT). These are two sides of the same coin.

Languages and semantic tools are the subject of technical protocols and standards (e.g. W3C Recommendations, Oasis standards, best practices). They are not mandatory. Legal scholars, computer and social scientists have identified the main elements of this new regulatory framework — including

[53] As stated by Dwork and Roth [49]: "*Differential privacy* describes a promise, made by a data holder, or curator, to a data subject: 'You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, data sets, or information sources, are available.' At their best, differentially private database mechanisms can make confidential data widely available for accurate data analysis, without resorting to data clean rooms, data usage agreements, data protection plans, or restricted views. Nonetheless, data utility will eventually be consumed: the Fundamental Law of Information Recovery states that overly accurate answers to too many questions will destroy privacy in a spectacular way. The goal of algorithmic research on differential privacy is to postpone this inevitability as long as possible."

[54] On 8 April 2016 the Council adopted the Regulation and the Directive. On 14 April so did the European Parliament. On 4 May, the official texts were published in the EU Official Journal (in all the official languages). The Regulation entered into force on 24 May, and it shall apply from 25 May 2018. The Directive, on 5 May 2016, and the EU states should transpose it into their national laws before 6 May 2018.

[55] https://ec.europa.eu/priorities/digital-single-market_en

[56] This American judicial tradition of property under some constraints to protect individual rights has been interpreted by Morton Horwitz [56] as a benefit for the economic development of entrepeneurs and companies, creating the legal conditions for 20 c. liberal capitalism.

[57] Two different Western cultures: "On the one hand, a European interest in personal dignity, threatened primarily by the mass media; on the other hand, an American interest in liberty, threatened primarily by the government" [60].

the results of fifteen years of research on legal XML, Legal RuleML, and legal ontologies; linked data publication and consumption, copyright related terms, licensing, patents, privacy risks, and the emergence of data and metadata markets [61, 62]. But their legal function, and how they will contribute to create sustainable social ecosystems are still uncertain at a general level.

The relationship between the rule of law and its semantic and algorithmic counter-part, the meta-rule of law, should be explored further. The intermediate level between macro- and micro-regulations opens a space for intermediate anchoring institutions, compliance by design and regulatory modelling, to minimise risks and to implement individual and collective rights. Mobile health applications, websites (such *PatientsLikeMe* and the *Health Tracking Network*), crowdsourcing platforms, programs (such *Flu Near You*), *Electronic Health Records* [EHR] databases… could benefit from this kind of intermediate regulatory institutions.

It is worth noticing that underlying ethics have a new positive regulatory value in this field, both for Artificial Intelligence broad applications [63] and in the digital space for biomedical Big Data [64]. Traditional legal tools are not enough, and there is an urgent need for us to understand it.

5 Conclusions

Law is facing significant new challenges that should be discussed. *Personalisation* (the use of services, information, and knowledge) in the Web of Data, create new unregulated contexts and scenarios. Some boundaries arise within emerging data markets. Others unfold under non-harmonised jurisdictions and rules, while other boundaries relate to safety and collective security. There are at least these pending topics on the list:

1. *Digital legal pluralism* (how to deal with different legal cultures, jurisdictions, and the binding power of national states).
2. *Legal forum-shopping* (how to deal with markets, companies and corporate power).
3. Balancing *citizens' rights* (privacy) and *security* (how to deal with individual rights and collective needs).
4. *Creating a global, digital, legal culture* (how to deal with the general human-machine framework in which agency, risk scenarios and social ecosystems emerge).
5. *Improving National and International legal frameworks* to include new digital terminologies and concepts, harmonising them and allowing legal interoperability across jurisdictions and legal cultures.
6. *Consciousness*. We should acknowledge the change, and accept that privacy is a *public and collective* issue. Building privacy means accepting that there is no absolute privacy. It deals rather with building communities and generating trust in citizens and consumers —within national and transnational markets.
7. *Aligning civil, legal and technological knowledge*. Identity means building, controlling and monitoring *knowledge* about safety, security, and personage (data and metadata).
8. Solving the *algorithmic-semantic puzzle*. From the technical point of view, differential privacy, and (semantic) privacy and data protection by design should be developed alike. Encryption, de-identification, and self-enforcing protocols should be encompassed with ethical and legal protections.
9. *Ethics matter*. Principles and values should be fleshed out, and dynamically *anchored* into normative and institutional systems (not only considering them as preliminary requirements for enginery-building).
10. We should acknowledge that there is a political dimension too. Whistle-blowers, activists (including ethical hackers), perform significant socio-political functions. Having an open mind and accepting innovation means adding *linked democracy* as a new dimension of liberal, deliberative, and epistemic democracy.

Acknowledgements This article is an outcome of Melbourne-based researchers of the Law and Policy Program of the Australian government-funded Data to Decisions Cooperative Research Centre (http://www.d2dcrc.com.au/), with the cooperation of the UAB Institute of Law and Technology (DER2012-39492-C02-01, and DER2016-78108-P).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Funding Data to Decisions Cooperative Research Centre (D2D CRC Ltd., ABN 45168769677; Project DC160051 - Practical perspectives on a balanced, enabling regulatory framework for data-based decision-support technologies used by law enforcement and national security in Australia.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study (not applicable).

References

1. Executive Office of the US President. Big Data: Seizing Opportunities, Preserving Values (White House). Published May 1. Council on Foreign Relations (mar 2017). http://wwwcfrorg/technology-and-science/white-house-big-data—seizing-opportunities-preserving-values/p32916 Cited. October 5 2016.

2. UK Government. Emerging Technologies Big Data Community of Interest, 'Emerging Technologies: Big Data: A Horizon Scanning Research Paper' (HM Government Horizon Scanning Programme. Cabinet Office. First published: 18 November 2014. Available at: https://www.gov.uk/government/publications/emerging-technologies-big-data. Cited October 5 2016.
3. Australian Government Information Management Office. 'Australian Public Service Big Data Strategy - Improved Understanding Through Enhanced Data-Analytics Capability'. 2013. Department of finance and deregulation. ISBN: 978–1–922096-27-2 Available at: http://wwwfinancegovau/sites/default/files/the-australian-public-service-big-data-strategy-archivedpdf Cited 5 October 2016.
4. Australian Government. 'Australian Public Service Better Practice Guide for Big Data'. 2015. Version 2.0. January 2015. Joint work of the Data Analytics Centre of Excellence (chaired by the Australian Taxation Office) and the Big Data Working Group (chaired by the Department of Finance). ISBN: 978–1–922096-31-9 Available at: http://www.finance.gov.au/sites/default/files/APS-Better-Practice-Guide-for-Big-Data.pdf. Cited 5 October 2016.
5. Diebold FA. Personal Perspective on the Origin(s) and Development of 'Big Data': The Phenomenon, the Term, and the Discipline, PIER Working Paper No 13–003, 2nd Version. 2012. Available at: http://ssrn.com/abstract=2202843. Cited 5 October 2015.
6. Cox M, Ellsworth D. Application-Controlled Demand Paging for Out-of-Core Visualization. Proceedings of the 8th Conference on Visualization, IEEE Computer Society Press; 1997.
7. Laney D. 3D data management: controlling data volume, velocity and variety. META Group Research Note No. 2001;6:2001.
8. Ylijoki O, Porras J. Perspectives to definition of Big Data: A mapping study and discussion. Journal of Innovation Management. 2016;4(1):69–91.
9. Bennett Moses L. Bridging distances in approach: Sharing ideas about technology regulation. In R. Leenes, E. Kosta (editors) Bridging distances in technology and regulation, Oisterwijk: Wolf Legal Pub.; 2013. 37–51.
10. Bijker WE. Of Bicycles, Bakelites, and Bulbs: Towards a Theory of Sociotechnical Change, MIT Press; 1997.
11. Orlikowski WJ, Gash DC. Technological frames: Making sense of information technology in organisations. ACM Trans Inf Syst. 1994;12(2):174–207.
12. Chan J, et al. The technological game: How information technology is transforming police practice. Criminal Justice. 2001;1(2):139–59.
13. Law and Policy Program, Data to Decisions CRC. Big Data Technology and National Security - Comparative International Perspectives on Strategy, Policy and Law: Methodology (Data to Decisions CRC). Adelaide. 2016.
14. Law and Policy Program, Data to Decisions CRC. Big Data Technology and National Security- Comparative International Perspectives on Strategy, Policy and Law: Comparative Study (Data to Decisions CRC, 2016). Adelaide 2016.
15. Law and Policy Program, Data to Decisions CRC. Big Data Technology and National Security - Comparative International Perspectives on Strategy, Policy and Law: Australia (Data to Decisions CRC). Adelaide. 2016.
16. Cannataci JA. Report of the Special Rapporteur on the Right to Privacy. 2016. Available at: http://www.ohchr.org/EN/Issues/Privacy/SR/Pages/SRPrivacyIndex.asp. Cited 5 October 2016.
17. Geneva Academy of International Humanitarian Law and Human Rights, Report of the Side event: The Right to Privacy in the Digital Age: Challenges and Ways Forward. 8 March 2016. Available at: http://www.geneva-academy.ch/docs/events/2016/GenevaAcademy%20FINAL%20REPORTsideevent%20privacy_8March2016-1.pdf. Cited 5 October 2016.
18. Greenleaf G. UN Privacy Rapporteur Sets High Goals, 140 *Privacy Laws & Business International, Report* 10–12, 30 April 2016. UNSW *Law* Research Paper No. *2016*–5. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2801981. Cited 5 October 2016.
19. NIST Big Data Public Working Group Definitions and Taxonomies Subgroup. NIST Special Publication 1500–1, NIST Big Data Interoperability Framework: Volume 1, Definitions. Final Version 1. 2015. Available at: doi:10.6028/NIST.SP.1500-1. Cited 5 October 2016.
20. OpenKnowledge International. Open Data Handbook. What is Open Data?. 2016. Available at: http://opendatahandbook.org/guide/en/what-is-open-data/. Cited 5 October 2016.
21. Australian Digital Transformation Office. Open data Improving services through accessible machine-readable data. Updated 21 July 2015. Available at: https://www.dto.gov.au/standard/design-guides/open-data/. Cited 5 October 2016.
22. Australian Department of Health. Linkable de-identified 10% sample of Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Schedule (PBS). 2016. Available at: https://data.gov.au/dataset/mbs-sample-10pct-1984-gz Data temporally unavailable, 5 October 2016.
23. Cavoukian A, Castro D. Big Data and Innovation, Setting the Record Straight: De-identification Does Work. 2014. Available at: http://www2.itif.org/2014-big-data-deidentification.pdf. Cited 5 October 2016.
24. Narayanan A, Felten EW. (Princeton). No silver bullet: De-identification still doesn't work. 2014. Available at: http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf. Cited 5 October 2016.
25. Nakamoto S. [pseudonym]. Bitcoin: A peer-to-peer electronic cash system, White paper. Published for the first time in October 31st. 2008. Available at: http://satoshi.nakamotoinstitute.org/quotes/bitcoin-design/ Also available at: https://bitcoin.org/bitcoin.pdf. Cited 10 March 2017.
26. Pentland Alex (MIT). Big Data's Biggest Obstacles. 2012. Available at: https://hbr.org/2012/10/big-datas-biggest-obstacles. Cited 5 October 2016.
27. Martin P, Hunter F. Minister says Census 'no worse than Facebook' as Nick Xenophon risks jail. The Sydney Morning Herald. 2016. Available at: http://www.smh.com.au/federal-politics/political-news/minister-says-census-no-worse-than-facebook-as-nick-xenophon-risks-jail-20160808-gqnobg.html. Cited 5 October 2016.
28. Dent J. Ex-head of ASIO, David Irvine, on Data Retention Laws. 2015. The ethics Centre. Available at: http://wwwethicsorgau/on-ethics/blog/march/exclusive-ex-head-of-asio,-david-irvine,-on-data-r Cited 5 October 2016.
29. Dingle S. Attorney-General George Brandis struggles to explain Government's metadata proposal. ABC News. Available at: http://www.abc.net.au/news/2014-08-07/brandis-explanation-adds-confusion-to-metadata-proposal/5654186. Cited 5 October 2016.
30. Australian Government. Review Panel. Review of The Personally Controlled Electronic Health Record. Available at: https://health.gov.au/internet/main/publishing.nsf/Content/17BF043A41D470A9CA257E13000C9322/$File/FINAL-Review-of-PCEHR-December-2013.pdf. Cited 5 October 2016.
31. The Advocate-General for Scotland (Lord Keen of Elie) (Con), "Investigatory Powers Bill" House of Lords Hansard, 13 July 2016, Volume 774 Column 228. Availllable at: https://hansardparliamentuk/lords/2016-07-13/debates/16071337000437/InvestigatoryPowersBill Cited 5 October 2016.
32. Casanovas P. Conceptualisation of Rights and Meta-rule of Law for the Web of Data, Democracia Digital e Governo Eletrônico (Santa Caterina, Brazil) vol.12. 2015: 18-41; repr. Journal of Governance and Regulation / Volume 4, Issue 4, 2015; p. 118–129. Available at:

http://buscalegis.ufsc.br/revistas/index.php/observatoriodoegov/article/viewFile/34399/33229. Cited 5 October 2016.
33. Casanovas P, Rodríguez-Doncel V, González-Conejero J. The Role of Pragmatics in the Web of Data; in F. Poggi, A. Capone (Eds.) Pragmatics and Law. Practical and Theoretical Approaches, Berlin: Springer Verlag. 2016. pp. 293–330. DOI: 10.1007/978-3-319-44601-1_12. Available at SSRN: http://ssrn.com/abstract=2697832. Cited 5 October 2016.
34. Rajaraman V. Big Data Analytics. Resonance. 2016;21:695. doi:10.1007/s12045-016-0376-7.
35. Poblet M. (ed.). Mobile Technologies for Conflict Management. Online dispute resolution, governance, participation. LGTS n. 2, 2011, Dordrecht: Springer Verlag
36. Corea, F. Big Data Analytics: A Management Perspective, Studies in Big Data 21, 2016, Switzerland, Springer Nature
37. Williams S. Business intelligence strategy and Big Data analytics: a general management perspective. Amsterdam: Morgan Kaufmann, Elsevier; 2016.
38. Castro, D., McQuinn, The Privacy Panic Cycle: A Guide to Public Fears about New Technologies, US Information Technology and Innovation Foundation. 2015. Available at: http://www2.itif.org/2015-privacy-panic.pdf. Cited 5 October 2016.
39. Casanovas P. The Future of Law: Relational Law and Next Generation of Web Services, in M. Fernández-Barrera, P. de Filippi et al. (Eds.) The Future of Law and Technology: Looking into the Future. Selected Essays. European Press Academic Publishing, Legal Information and Communication Technologies Series, vol. 7, Florence. 2010. pp. 137–156. Available at: http://www.ejls.eu/6/205UK.htm. Cited 5 October 2016.
40. Casanovas P. Open Source Intelligence, Open Social Intelligence and Privacy by Design. In ECSI-2014, CEUR 183, pp. 174–185, Available at: http://ceur-ws.org/Vol-1283/. Cited 5 October 2016.
41. Barrett MA, Humblet O, Hiatt RA, Adler NE. Big Data and disease prevention: From qualified self to quantified communities. Big Data. 2013;1(3):168–75. doi:10.1089/big.2013.0027. http://online.liebertpub.com/doi/abs/10.1089/big.2013.0027. Cited 5 October 2016
42. Schmachtenberg M, Bizer C, Jentzsch A, Cyganiak R. The Linking Open Data Cloud Diagram. 2014. Available at: http://Lod-Cloud.Net. Cited: 5 October 2016. See as a published paper: Schmachtenberg, M. et al. Adoption of the Linked Data Best practices in different topical domains, 13th International Semantic Web Conference, LCNS 8796, pp. 245-260. Springer International Publishing.
43. Rodriguez-Doncel V, Gómez-Pérez A, Mihindukulasooriya N, Rights declaration *in* Linked Data. In: O. Hartig et al. (ed.) Proceedings of the Fourth International Workshop on Consuming Linked Data (COLD2013), CEUR 1034. 2014. Available at: http://ceur-ws.org/Vol-1034/RodriguezDoncelEtAl_COLD2013.pdf. Cited 5 October 2016.
44. Paschke, A. Pragmatic Web 4.0. Towards an Active and Interactive Semantic Media Web, W3C Aspect of Semantic Technologies. 2013. Available at: http://www.slideshare.net/swadpasc/pragmatic-web-paschke. Cited 5 October 2015.
45. Casellas N, Blázquez M, Kiryakov A, Casanovas P, Poblet M, Benjamins VR. OPJK into PROTON: Legal domain ontology integration into an upper-level ontology. In OTM Confederated International Conferences, On the Move to Meaningful Internet Systems, Springer Berlin, Heidelberg. 2005, pp. 846–855.
46. Casanovas P, Casellas N, Tempich C., Vrandecic D., Benjamins VR. OPJK and DILIGENT: ontology modeling in a distributed environment. Artificial Intelligence and Law, 2007; 15 (2): 171–186.
47. Dwork C. The differential privacy frontier. In: Theory of Cryptography Conference, Springer, LNCS 5444. Berlin: Springer; 2009. pp. 496–502.
48. Dwork C. Differential privacy: a survey of results. In: In international Conference on theory and applications of models of computation, LNCS, Springer, Berlin, 4978; 2008.,pp. 1–19.
49. Dwork C, Roth A. The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science. 2014;9(3–4):211–407.
50. Simonite T. Apple Rolls Out Privacy-Sensitive Artificial Intelligence, MIT Technology Review. 2016. Available at: https://www.technologyreview.com/s/601688/apple-rolls-out-privacy-sensitive-artificial-intelligence/. Cited 5 October 2016.
51. Simonite T. Apple's New Privacy Technology May Pressure Competitors to Better Protect Our Data, MIT Technology Review. 2016. Available at: https://www.technologyreview.com/s/602046/apples-new-privacy-technology-may-pressure-competitors-to-better-protect-our-data/?utm_campaign=content-distribution&utm_source=dlvr.it&utm_medium=twitter. Cited 5 October 2016.
52. Hijmans, H. 2016. The European Union as Guardian of internet privacy: the story of art 16 TFEU, LGTS vol. 31, Dordrecht: Springer.
53. Gutwirth S, Leenes R, De Hert P (Eds.). Data Protection on the Move. Current Developments in ICT and Privacy/Data Protection, Springer Verlag, Dordrecht, 2016.
54. González-Fuster G. The emergence of personal data protection as a fundamental right of the EU. Springer, Dordrecht, LGT. 16: 2014
55. De Hert P, Papakonstantinou V. The new General Data Protection Regulation: still a sound system for the protection of individuals? Computer Law & Security Review. 2016;32:179–94.
56. Horwitz MJ. The transformation of American law, 1870–1960: The crisis of legal orthodoxy. Oxford: Oxford University Press; 1992.
57. Westin AF. Privacy and freedom. New York: Atheneum Books; 1963.
58. Westin AF. (dir.). Computers, personnel administration, and citizen rights. 1979. 500 (50), US Department of Commerce, National Bureau of Standards.
59. Warren SD, Brandeis LD. The right to privacy. Harvard Law Review. 1890;4(5):193–220.
60. Whitman JQ. The two Western cultures of privacy: Dignity versus liberty. Yale Law Journal. 2004;113:1151–221.
61. Casanovas P, Palmirani M, Peroni S, van Engers T, Vitali, F. Special Issue on the Semantic Web for the Legal Domain, Guest Editors Editorial: The Next Step. Semantic Web Journal. 2016, 7 (2): 1–13. Available at: available at SSRN: http://ssrn.Com/abstract=2765912. Cited 5 October 2016.
62. Rodríguez-Doncel V, Santos C, Casanovas P, Gómez-Pérez A. Legal aspects of linked data – The European framework, Computer Law & Security Review: The International Journal of Technology Law and Practice, 2016;32(6): 799–813, doi: 10.1016/j.clsr.2016.07.005.
63. Pagallo U. The Law of Robots. Crime, contracts, and torts. Dordrecht: Springer; 2013.
64. Mittelstadt BD, Floridi L, (ed.). The Ethics of Biomedical Big Data. Dordrecht: Springer; 2016.
65. Werbin KC. The List Serves: Population Control and Power. Amsterdam: Institute of Network Cultures; 2008.