Ten Questions for Future Regulation of Big Data

A Comparative and Empirical Legal Study

by **Bart van der Sloot and Sascha van Schendel***

Abstract: Much has been written about Big Data from a technical, economical, juridical and ethical perspective. Still, very little empirical and comparative data is available on how Big Data is approached and regulated in Europe and beyond. This contribution makes a first effort to fill that gap by presenting the reactions to a survey on Big Data from the Data Protection Authorities of fourteen European countries and a comparative legal research of eleven countries. This contribution presents those results, addressing 10 challenges for the regulation of Big Data.

Keywords: Big Data; Empirical; Comparative; Survey; Data Protection Authorities; Comparative Legal Research

Recommended citation: Bart van der Sloot and Sascha van Schendel, Ten Questions for Future Regulation of Big Data: A Comparative and Empirical Legal Study, 7 (2016) JIPITEC 110 para 1.

A. Introduction

1 Big Data is a buzzword used frequently in both the private and the public sector, the press, and online media. Large amounts of money are being invested to make companies Big Data-proof, and governmental institutions are eager to experiment with Big Data applications in the fields of crime prevention, intelligence, and fraud, to name but a few areas. Though the exact nature and delineation of Big Data is still unclear, it seems likely that Big Data will have an enormous impact on our daily lives. Positively, undoubtedly, but there are also inherent risks to Big Data applications, as it might result in discrimination, privacy violations, and chilling effects. The ideal situation would be to have an adequate framework in place that will ensure that the beneficial uses of Big Data are promoted and facilitated, while the negative effects are mitigated or sanctioned. This contribution provides building blocks for developing such a framework, by giving an overview of the experience in the use and regulation of Big Data in 23 countries, aiming in particular at the use of Big Data by governments.

2 The research presented in this article was conducted in two phases. The first phase involved desk research and looked at Big Data policies, legislation and regulation in a number of countries. Second, a questionnaire was sent to several European DPAs. The desk research examined eleven countries. These countries were selected on the basis of three criteria. The first was global coverage – the research sought to be as representative as possible to provide a full picture of global developments in relation to Big Data, which is by nature an international phenomenon. Therefore, at least one country from each continent (with the exception of Antarctica) was examined. The second criterion was an estimation of the potential value of the expected outcomes of the research – some countries are more innovative and ambitious

than others in terms of technological developments such as Big Data. Thirdly, the role a country plays in international politics was taken into account; on that basis, China rather than South Korea was studied, even though the latter country is often in the forefront of technological developments. Based on these three criteria Australia, Brazil, China, France, Germany, India, Israel, Japan, South Africa, the United Kingdom and the United States were selected. The desk research focused on two issues in particular. First, government policy decisions were analyzed, as were initiatives related to this topic, such as governments using Big Data themselves or stimulating the use of Big Data in the private sector, either through financial support or by engaging in partnerships. Second, research was carried out on legislation and case law revolving around Big Data in the selected countries. It should, again, be noted that this study is not exhaustive – there is, undoubtedly, a myriad of relevant laws, court cases and DPA reports that are not discussed here.

3 In studying the eleven countries, almost exclusive use was made of official sources, especially government websites. The reason for this is that it is often difficult to establish the reliability of foreign sources. This choice does, however, imply that this article mainly presents a picture of the governmental view of Big Data and of governmental regulation. Criticism of those initiatives and autonomous processes in the private sector remain largely undiscussed. This bias was accepted as a tradeoff in order to guarantee the reliability of the sources studied. When discussing Israel, however, use was made of online newspaper articles from Israeli news sources and a published online interview, because this provided vital information and because the news-source was regarded as reliable. The information from these sources was not available on government websites, but was nonetheless considered essential.

4 Publications on government websites and in press releases about new initiatives were selected by using terms related to Big Data, both in the official language of the country concerned and in English, such as 'data mining', 'data analytics', 'data projects', 'Big Data initiatives', etc. Several countries have a Ministry of Science and Technology, or a similar ministry. Those ministries were taken as the starting point of the research in those countries. General search engines were also used to scan government initiatives related to Big Data, by limiting the search to the national public domain of the country concerned. For case law and legislation, the official national search engines and general search engines were used. The search terms entered here were related to Big Data, privacy and data protection, such as 'data protection', 'privacy', 'surveillance', etc. This process yielded a list of government initiatives, legislation and relevant jurisprudence. The sources consulted and the full list of references used for this article are listed in a working paper published earlier.[1]

5 The results of the comparative desk research can be found in Appendix I and the results of the survey in Appendix II to this contribution. It has to be stressed that not all governments and governmental agencies use the term Big Data when creating, operating on, or using large scale data bases. That is why this study primarily identifies those initiatives that have been identified as Big Data by the government itself, or when it has used terms that are related to it. This means that many uses of large scale databases by governmental agencies are not included in this study. When analyzing the countries, six questions were kept in mind: 'Is a specific definition of Big Data used?', 'Is Big Data used within the government?', 'Is there a public-private partnership?', 'To what goal is Big Data used by the government?', 'Which laws are especially relevant for Big Data?' and 'Are there judicial decisions relating to Big Data?'

6 A relatively short and simple questionnaire was designed for the survey, so as to increase the potential response of the DPAs. The accompanying email, as well as the introduction to the survey, briefly explained the goal of the survey. The survey comprised six questions: 1. Are you familiar with the debate on Big Data? If so, how would you define Big Data? (max. 500 words) 2. Are there prominent examples of the use of Big Data in your country, especially in the law enforcement sector, by the police or by intelligence services? (max. 500 words) 3. Have you issued any decisions/reports/opinions on the use of Big Data? If so, could you provide us with a reference and your main argument? (max. 500 words) 4. Are there any legal cases/judgements by a court with regard to (privacy/data protection) violations following from Big Data practices in your country? If so, could you provide us with a reference and the main consideration of the court? (max. 500 words) 5. Which legal regimes are applied to Big Data/ is there a special regime for Big Data in your country? Are there any discussions/plans in parliament to introduce new legislation to regulate Big Data practices? (max. 500 words) 6. Are there any final remarks you want to make/suggestions you have for further research? (max. 500 words)

7 The reason for choosing these questions for the desk research and the survey is that the background of this study is a project by the Netherlands Scientific Council for Government Policy (WRR). The WRR

1 <http://www.wrr.nl/fileadmin/en/publicaties/PDF-Working_Papers/WP_20_International_and_Comparative_Legal_Study_on_Big_Data.pdf>. The literature studied for this article can be found here. <http://www.wrr.nl/fileadmin/nl/publicaties/PDF-Rapporten/rapport_95_Big_Data_in_een_vrije_en_veilige_samenleving.pdf>.

is an independent advisory body for the Dutch government. The task of the WRR is to advise the government on issues that are of great importance for society in the intermediate and longer term. The reports of the WRR are not tied to one policy sector but rather touch on various terrains and policy sectors; they are concerned with the direction of government policy for the longer term. The members of the WRR are established university professors who have often worked on policy related subjects and/or have made tracks in public administration themselves. The Dutch government had requested the WRR to advise on the regulation of Big Data, taking into account how privacy and security should be assessed in the deployment of big data analytics in security related policies. Questions that were suggested to be addressed include whether a distinction needs to be made between access to and use of data, how transparency and individual rights can be guaranteed in Big Data practices and what the likely impact of the emergence of quantum computing will be. In addition to the policy advice, published in the form of a report for the Dutch government,[2] a scientific book was delivered[3] and a number of working papers were written to do indicative research,[4] which were used as building blocks for the report to the government. This article is based on one of those working papers.[5]

8 The DPAs in all 28 EU Member States were emailed with a request to complete the survey. Requests were also sent to the DPAs in three non-EU countries, namely Norway, Serbia and Switzerland, because a short preliminary study had shown that they might have specific expertise in relation to Big Data. DPAs that did not respond within the period specified in the initial request were sent a reminder; those that did not respond to this mail either were sent a final reminder. In most cases, the questionnaire was sent to the general contact address as posted on DPA's website. However, since the French website lists no general email address, personal contacts were used to email two specific employees of the CNIL. For three other DPAs (Germany, the Netherlands and Norway), in addition to an email to the general email address, an email was also sent to a specific individual employee. For other DPAs, either no such personal contacts existed or they existed but it was not necessary to use them because a response had been received. Eventually, of the 31 DPAs included in the survey, 18 responded: Austria, Belgium, Croatia, Denmark, Estonia, Finland, France, Hungary, Ireland, Latvia, Lithuania, Luxembourg, the Netherlands, Norway, Slovakia, Slovenia, Sweden and the United Kingdom. Four of these (Austria, Denmark, Finland and Ireland) were negative responses, stating that the DPA in question would not participate in the study. Consequently, about half of the DPAs invited to join the survey have actually responded. The results found in this study can, therefore, not be seen as determinative but as indicative of possible trends, feelings and attitudes towards Big Data. It should be taken into account that those DPAs that have already dealt with Big Data projects would be more likely to respond to such a survey than those that haven't.

9 Rather than presenting the bare facts, listing the regulatory initiatives in the various countries studied and the answers from the DPAs, this article uses the insights gained from those results to shine light on some of the most difficult questions regulators have to answer when deciding on future regulation of Big Data. These questions are partly based on those asked in the survey and partly follow from the desk research. Additional questions have been added in order to present the most interesting findings from both the desk research and the survey in an orderly fashion. Ten issues/questions are discussed in more detail: (1) What is the definition of Big Data? (2) Is Big Data an independent phenomenon? (3) Big Data: fact or fiction? (4) What is the scope of Big Data? (5) What are the opportunities for Big Data? (6) What are the dangers of Big Data? (7) Are the current laws and regulations applicable to Big Data? (8) Is there a need for new legislation for Big Data? (9) What concept should be central to Big Data regulation? (10) How should the responsibilities be distributed? These questions will be discussed in the subsequent sections. The article will conclude with a short summary of the main findings.

B. What is the definition of Big Data?

10 The first choice when it comes to regulating Big Data is to determine a definition and delineation of Big Data. Three definitions were encountered a number of times in both the desk research and in the survey. First, the Article 29 Working Party holds that Big Data refers to the exponential growth, both in the availability and in the automated use of information. It refers to gigantic digital datasets held by corporations, governments and other large organizations, which are then extensively analyzed using computer algorithms. Big Data can, according to the Working Party, be used to identify more general trends and correlations, but it can also be

2 <http://www.wrr.nl/fileadmin/nl/publicaties/PDF-Rapporten/rapport_95_Big_Data_in_een_vrije_en_veilige_samenleving.pdf>.

3 <http://www.wrr.nl/fileadmin/en/publicaties/PDF-Verkenningen/Verkenning_32_Exploring_the_Boundaries_of_Big_Data.pdf>.

4 <http://www.wrr.nl/publicaties/working-papers/>.

5 <http://www.wrr.nl/fileadmin/en/publicaties/PDF-Working_Papers/WP_20_International_and_Comparative_Legal_Study_on_Big_Data.pdf>.

processed in order to directly affect individuals.[6] Second, the European Data Protection Supervisor (EDPS) suggests that Big Data means large amounts of different types of data produced at high speed from multiple sources, whose handling and analysis require new and more powerful processors and algorithms. Not all of these data, the EDPS points out, are personal, but many players in the digital economy increasingly rely on the large scale collection of and trade in personal information. As well as benefits, these growing markets pose specific risks to individual's rights to privacy and to data protection, the EDPS warns.[7] Third, and perhaps most well-known, the Gartner Report focusses on three matters when describing Big Data: increasing volume (amount of data), velocity (speed of data processing), and variety (range of data types and sources). This is also called the 3V model or 3V theory.[8]

11 The desk research also showed that a number of countries apply their own definition of Big Data. For example, in Germany, Big Data is defined as 'das Synonym für den intelligenten Umgang mit solchen großen oder auch heterogenen Datenmengen' (synonymous with the intelligent use of large or heterogeneous datasets).[9] The Podesta Report (United States) builds on the Gartner definition and suggests that 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. In other words, 'data is now available faster, has greater coverage and scope, and includes new types of observations and measurements that previously were not available.' More precisely, Big Datasets are 'large, diverse, complex, longitudinal, and/or distributed datasets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future."[10]

12 Finally, several DPAs also gave their own definition of Big Data when completing the survey, or referred to specific definitions used in their country. For example, the Estonian DPA describes Big Data as collected and processed open datasets, which are defined by quantity, plurality of data formats, and data origination and processing speed.[11] The French DPA refers to a definition adopted by the French General Commission on terminology and neology (Commission générale de terminologie et de néologie). The official translation of Big Data in French is 'mégadonnées', which stands for data, structured or otherwise, whose very large volume require appropriate analytical tools. The DPA of Luxembourg suggests that Big Data stems from the collection of large structured or unstructured datasets, the possible merger of such datasets, as well as the analysis of these data through computer algorithms. These datasets can usually not be stored, managed and analyzed with average technical means due to their size, it also points out. The Dutch DPA primarily points to the 'volume' aspect of Big Data and argues in particular that Big Data is all about collecting as much information as possible, storing it in ever-larger databases, combining data that is collected for different purposes and applying algorithms to find correlations and unexpected new information. The DPA from Slovenia not only refers to the use of different types of data, acquired from multiple sources in various formats, but also to predictive analytics used in Big Data. Finally, the Swedish DPA suggests the concept is particularly used for situations where large amounts of data are gathered in order to be made available for different purposes, not always precisely determined in advance.

13 It can be seen from this list of definitions that a number of components are regularly mentioned. Broadly, they relate to three states of Big Data processing, namely the collection, analysis and use of data. When it comes to collecting data, Big Data is about collecting large amounts of data (volume) from varied (variety) and often unstructured data sources. With regard to analyzing the collected data, Big Data revolves around the speed (velocity) of the analyses and the use of certain instruments such as algorithms, machine learning and statistic correlations. The results are often predictive in nature (predictive analytics) and are formulated at a general or group level. The results are usually applied by means of profiling. Many of the definitions contain some of these components; none of the definitions used mention all of these components. Consequently, none of these elements should be seen

6 <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf>.

7 <https://secure.edps.europa.eu/EDPSWEB/edps/Consultation/big_data>. See also: <https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/ <Big%20data/14-07-11_EDPS_Report_Workshop_Big_data_EN.pdf>.

8 <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.

9 References to the situation in the different countries studies might be found in Appendix I or at: <http://www.wrr.nl/fileadmin/en/publicaties/PDF-Working_Papers/WP_20_International_and_Comparative_Legal_Study_on_Big_Data.pdf>.

10 <https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf>.

11 References to the answers to the survey might be found in Appendix II or at: <http://www.wrr.nl/fileadmin/en/publicaties/PDF-Working_Papers/WP_20_International_and_Comparative_Legal_Study_on_Big_Data.pdf>.

as essential – that is, if one or more of these elements do not apply, it does not follow that the phenomenon being studied is not Big Data. Rather, these elements should be seen as parameters; if none of the elements apply, the phenomenon is definitely not Big Data; if all the elements apply, the phenomenon being studied definitely is Big Data. Mostly, however, it will somewhere in between. It is impossible to say, for example, how big a dataset must be in order to qualify as Big Data; although Big Data usually works with combined datasets, it is conceivable that one enormous dataset could qualify as Big Data; although Big Data usually (partially) works with unstructured data, this is not a condition sine qua non; etc.

C. Is Big Data an independent phenomenon?

14 The overview of definitions already shows that Big Data should not be seen as an isolated phenomenon. It is a new phenomenon which by its nature is strongly connected to a number of technical, social and legal developments. This conclusion is supported by the desk research, which also found that Big Data is intertwined with several other terms. For example, lots of Big Data initiatives are linked to Open Data. As the name suggests, Open Data is the idea that (government) data should be placed in the public domain. Traditionally, it has been linked to efforts to increase transparency in the public sector and give more control over government power to media and/or citizens. The Estonian DPA is in particular very explicit about the relationship between Open Data and Big Data, as it defines Big Data as "collected and processed open datasets, which are defined by quantity, plurality of data formats and data origination and processing speed". The desk research also shows a clear link between the two concepts in countries such as Australia, France, Japan and the United Kingdom.

15 Linked to Open Data is the idea of re-use of data. Yet, there is one important difference. While Open Data has traditionally been concerned with transparency of and control over government power, the re-use of (government) data is specifically intended to promote the commercial exploitation of the data by businesses and private parties. The re-use of Public Sector Information is fostered through the PSI Directive of the European Union. More generally, re-use refers to the idea that data can be used for a purpose other than that for which they were originally collected. Obviously, the link between Big Data and re-use is often made, as appears both from the desk research and from the survey. The Norwegian DPA, for example, uses the definition of Big Data of the Working Group 29, 'but also add what in our opinion is the key aspect of Big Data, namely that it is about the compilation of data from several different sources. In other words, it is not just the volume in itself that is of interest, but the fact that secondary value is derived from the data through reuse and analysis.' The desk research also showed a link between the two concepts. In France, for example, Big Data is primarily seen as a phenomenon based on the re-use of data for new purposes and on the combination of different data and datasets.

16 The term 'Internet of Things' refers to the idea that more and more things are connected to the Internet – cars, lampposts, refrigerators, clothing, or any kind of object. This opens the way for the development of smart devices – for example, a refrigerator that records when the milk has run out and automatically reorders. By fitting all objects with a sensor, large quantities of data can be collected. As a consequence, Big Data and the Internet of Things are often mentioned in the same breath. An example is the DPA of the United Kingdom, which notes 'that Big Data may involve not only data that has been consciously provided by data subjects but also personal data that has been observed (e.g. from Internet of Things devices), derived from other data or inferred through analytics and profiling.'

17 Because of the applications of the Internet of Things and the constantly communicating devices and computers, the development of smart products and services has spiraled. Examples of such developments are smart cities, smart devices and smart robots. The desk research indicates that a number of countries – for example, the United States, China and the United Kingdom – make a link between such developments and Big Data systems. The Luxembourg DPA also emphasizes the relationship with smart systems, such as smart metering. 'At a national level, a system of smart metering for electricity and gas has been launched. The project is, however, still in a testing phase. - The CNDP has not issued any decisions, reports or opinions that are directly dealing with Big Data. The Commission has, however, issued an opinion in a related matter, namely with regard to the problematic raised by smart metering. In 2013, the CNDP issued an opinion on smart metering. The main argument of the opinion highlights the necessity to clearly define the purposes of the data processing, as well as the retention periods of the data related to smart metering.'

18 A term that is often associated with Big Data and is sometimes included as part of the definition of Big Data is 'profiling'. As increasingly large datasets are collected and analyzed, the conclusions and correlations are mostly formulated at a general or group level. This mainly involves statistical correlations, sometimes of a predictive nature. Germany is developing new laws on profiling and a number of DPAs emphasize the relationship between

Big Data and profiling; for example, the DPAs of the Netherlands, Slovenia, the UK and Belgium. The latter argues that 'we expect that de new data protection regulation will be able to provide a partial answer (profiling) to Big Data issues (legal interpretation of the EU legal framework).'

19 Similar to the term profiling, 'algorithms' is used in many definitions of Big Data. This applies to the definition by Article 29 Working Party, the EPDS and a number of DPAs responding to the survey, such as those of Luxembourg, the Netherlands and the UK. A number of countries also have a special focus on algorithms. To provide an example, in Australia, a 'Program Protocol' has been developed – a report may be issued which contains the following elements: a description of the data; a specification of each matching algorithm; the anticipated risks and how they will be addressed; the means of checking the integrity of the data; and the security measures used.

20 To provide a final example, cloud computing is also often associated with Big Data processes. In China and Israel, especially, the two terms are often connected to each other. For example, the Chinese vice-premier stressed that the government wants to make better use of technologies such as Big Data and cloud computing to support innovation; according to the Prime Minister, mobile Internet, cloud computing, Big Data and the Internet of Things are integrated with production processes, and will thus be an important engine for economic growth. In Israel, the plan is for the army to have a cloud where all data is stored in 2015 – there is even talk of a 'combat computing cloud', a data center that will make different tools available to forces on the ground. Some DPAs also suggest a relationship between cloud computing and Big Data; the Slovenian DPA, for example, states that 'new concepts and paradigms, such as cloud computing or Big Data should not lower or undermine the current levels of data protection as a fundamental human right.'

21 There are other terms that are often mentioned in connection with Big Data, such as machine learning, commodification of data, datafication, securitization and risk society. It goes beyond the scope of this article to discuss all these terms in depth. What is important to note is that Big Data should be primarily viewed in its interrelationship and in conjunction with other phenomena. Big Data is a part of and, in a certain sense, the umbrella term for many of the technological and societal developments that are already taking place. This needs to be taken into account when regulating Big Data. It seems advisable for regulators to take a holistic approach to the regulation of Big Data and related phenomena.

D. Big Data: fact or fiction?

22 There is still no clarity about the extent to which Big Data processes are already being used in practice. The reactions of a number of DPAs seem to suggest that Big Data is not yet an established practice. For example, the Austrian DPA declined to participate in the survey because it had encountered few if any Big Data processes; cautious reactions were also received from the DPAs of Latvia, Lithuania and Slovakia. The Belgian DPA suggests that there is currently a lack of clarity about Big Data and refers to Gartner's hype cycle.[12] It also adds: "Most Belgian projects seem to still be in a pilot phase and the visibility of Big Data in practice is still low." However, other DPA responses show a different picture – they confirm that Big Data is a major trend, and that Big Data is playing an increasingly significant role. Some DPAs, such as Norway, have written a special report on the regulation of Big Data practices. The United Kingdom DPA has also issued a discussion paper on this topic. Furthermore, it emerged from the desk research that projects are under way in most countries that are connected to Big Data, although it should be noted that a fairly broad approach was taken in the desk research to what qualified as 'Big Data'.

23 The picture that emerges from all of the foregoing is one in which Big Data plays a minor role in most countries at present but is set to become increasingly important. Big Data should, therefore, not be seen as either an actual practice or as a fiction, a hype that will blow over, but rather as a trend that will play a major role in five years' time and will have a significant impact on the government sector, on business, and on citizens' everyday life in the future. What is clear from the desk research is that in most countries the government feels it is missing out on this important trend. While industry is investing billions in Big Data projects, many governments are – or feel they are – lagging behind. This is why many governments are now beginning to invest heavily in Big Data projects.

24 To give a few examples, the desk research showed that in the United States, more than $200 million was reserved for a research and development initiative for Big Data, which was to be spent by six federal government departments; the army invested the most in Big Data projects, namely $250 million; $160 million was invested in a smart cities initiative, investing in 25 collaborative ventures focused on data usage. In the United Kingdom, £159 million was spent on high-quality computer and network infrastructure, there was £189 million in investments to support Big Data and to develop the UK's data infrastructure, and £10.7 million will be spent on

12 <www.gartner.com/technology/research/methodologies/hype-cycle.jsp>.

a center for Big Data and space technologies. In addition, £42 million will be spent on the Alan Turing Institute for the analysis and application of Big Data, £50 million will be set aside for the 'Digital Catapult', where researchers and industry are brought together to come up with innovative products; and lastly, in February 2014 the Minister of Universities and Science announced a new investment of £73 million in Big Data. This money will be used for bioinformatics, open data projects, research and the use of environmental data. In South Africa, the government has invested 2 billion South African Rand, approximately €126.8 million, in the Square Kilometre Array (SKA) project, which revolves around very large datasets. In France, seven research projects related to Big Data were awarded a total of €11.5 million. In Germany, the Ministry of Education and Research invested €10 million in Big Data research institutes and €20 million in Big Data research; this Ministry will also invest approximately €6.4 million in ABIDA, a four-year interdisciplinary research project focusing on the social and economic impact of large data sets.

25 These are just a few examples of what is being spent by the governmental sector. In the private sector, a multiple of these sums is being spent on Big Data projects. The expectation is that these Big Data projects will develop over the next five or ten years. Only then will many of the effects of Big Data become apparent. Consequently, when designing Big Data regulations, it seems advisable for governments to develop future-proof policies that follow and, where possible, anticipate this trend. If regulators only begin to regulate this phenomenon five or ten years from now, many of the projects will have already started. The negative impact may already have materialized, and it will be difficult to adjust and alter projects and developments that have already flourished. It should also be remembered that good, clear regulation can contribute to innovation and the use of Big Data. Since the current framework applying to new Big Data projects is not always clear, some government agencies and private companies are reluctant to use new technologies for fear of violating the law. New regulation could provide more clarity

E. What is the scope of Big Data?

26 This study, and especially the desk research, shows that Big Data projects are initiated for very different purposes. In Brazil, for example, the so called Data Viva system was initially used mainly for the formulation of economic policy. In addition, the police in Sao Paulo use a system (Detecta) that is based on Big Data technology. Detecta is an intelligent system for monitoring crime. In the United Kingdom, too, Big Data is used to fight crime. The POSTnote about Big Data and crime and safety provides an example of the use of Big Data by the police. Software has been developed as part of a pilot to predict the location of burglaries, and two British police forces use software developed for predictive policing to predict the locations of crimes. The British tax and customs authority, HMRC, also uses a Big Data system, 'Connect', in which all the data held is aggregated and analyzed. This Big Data system is used to detect tax fraud and tax evasion, and is said to have led to the recovery of £2.6 billion since April 2013. The system displays relevant information in searches that is otherwise difficult to find, allows complex analyses to be performed on the development of multiple datasets simultaneously, and enables profiles to be constructed which can help uncover patterns that may indicate particular crimes.

27 In some countries, Big Data is primarily seen as a means for the government to increase its own service to citizens; prominent examples are Australia and China. Reference can also be made in this connection to the Aadhaar project that has been developed and carried out by the 'Unique Identification Authority' of India and which involves the collection of biometric and demographic data on residents of India. One of the uses of Aadhaar is 'micropayments', a means of identification which should help improve access to financial services for people living in rural areas. The identification number makes it possible to identify people in remote regions from a long distance and also reduces costs through economies of scale, making it easier for poorer people to obtain financial services. Other sectors where Aadhaar provides solutions include demographic planning, paying security social benefits and improving the identification of beneficiaries by eliminating duplicate identities. Government administrative processes should become more efficient because the authorities now have access to all relevant information at a glance.

28 Several countries see Big Data mainly as a phenomenon that can help the private economy. Germany, for example, has launched a funding initiative to support the competitiveness of it companies, and France also feels that Big Data is set to take off, especially in the private sector, through the growth of it companies and startups which help to stimulate the economy and create jobs. There are also countries, such as Japan, Germany and the United Kingdom, where Big Data is approached primarily in relation to scientific research and innovation. Israel, finally, is unique in that it also uses new technological systems for facilitating the activities of the army. It also has to be borne in mind that many intelligence services are involved with Big Data-like projects; however, often little is

known about these projects, other than what has been leaked by whistleblowers.

29 The picture that emerges from this research is that Big Data could be used in almost every sector and for almost any task. Generally, the use of Big Data can be divided into three types. Firstly, the use of Big Data for specific government tasks – examples include the use of Big Data by intelligence services, the police, tax authorities and other public bodies, for example in the context of formulating economic policies. Second, the use of Big Data by the private or semi-public sector, helping or facilitating them in achieving their specific tasks and/or goals. Examples include the use of Big Data by companies to create risk profiles, to find statistical correlations and to personalize services and advertisements, and the use of Big Data by universities and research institutes for research-related purposes. Big Data is also widely used in the medical sector; for instance, the United Kingdom has heavily promoted the use of Big Data in the healthcare sector, and the Israeli Ministry of Health has a large dataset containing medical data on the citizens of Israel and on the healthcare system. According to the Ministry, the potential benefits lie in the facilitation of a variety of healthcare functions (including assisting in the clinical decision-making process, in monitoring diseases and in proactive healthcare). Thirdly, Big Data is used by both governments and private sector companies to improve their service to citizens or customers; this might, for example, involve increasing the transparency of their activities, strengthening the control of citizens over data processing, etc.

30 These three categories should lead to different approaches to regulation. The last category is relatively unproblematic because it serves the interests of the citizen. Here, the current legislation on aspects such as the use of personal data should suffice. The situation is different when Big Data is used by governmental agencies to support their goals. It is important to distinguish between the different fields in which Big Data is used by the government. If Big Data is used for the development of economic policies, for routinely inspecting fire installations or for epidemiological research, this should be relatively unproblematic. In these instances, general patterns and statistical correlations are used to promote the efficiency and effectiveness of public policy. However, if Big Data is used by the police, a different picture emerges – while Big Data is about processing large amounts of data and detecting general patterns, the police need to investigate and possibly arrest specific individuals on the basis of concrete facts. There is a particular danger of mismatches when general profiles are applied to specific individuals. When regulating Big Data, the potential impact on citizens must be taken into account; that impact will be greater when Big Data is used by the police, intelligence services and the army than when it is used for the development of general economic policies. It also appears from the survey that several DPAs are skeptical about the use of Big Data by the police, both because of the possible impact on the citizen and because of the potential for mismatches between general profiles and specific individuals.

31 Finally, the use of Big Data in the private sector can also be problematic. It emerged from this study that two things in particular need to be taken into account. First, use can be made of data or profiles that are based on sensitive information, such a data about race, medical conditions or religious beliefs; use can also be made of categories that appear neutral but are, in fact, based on these types of information – a practice known as redlining. Second, the consequences of the use of Big Data in the private sector may also be substantial, irrespective of whether or not sensitive information is used. Where advertisements are personalized through the use of Big Data-like applications, the impact will, of course, be relatively small; however, when Big Data is used to develop risk profiles on the basis of which banks decide who may be eligible for a loan and on what terms, or by health insurers to decide who they are prepared to insure and on what terms, the consequences can be significant. Factors that could be taken into account when regulating Big Data are the impact of its use on the individual, the types of data and data analysis that are used and the potential danger of a mismatch between general profiles and specific individuals. A distinction could also be made between the type of organization that uses Big Data and the specific purpose for which it is used. The general interest that is served by the use of Big Data naturally also has an impact on what should be considered legally admissible.

F. What are the opportunities for Big Data?

32 From both the desk research and the results from the survey it appears that Big Data represents both significant opportunities and significant risks. For example, in 2013, 'France Stratégie', an advisory body to the French Prime Minister, performed an analysis of the advantages and disadvantages of Big Data. It emphasized that, on the one hand, Big Data provides for more knowledge and opportunities, but that, on the other, it may cause problems in relation to the protection of privacy and confidentiality. John Podesta also stressed this duality. He published a blog on 1 May, 2014, which discussed the results of the Working Group Review. In his blog, Podesta describes Big Data as a vital technology. He refers to the devastation and suffering caused by tornadoes

and, more implicitly, to the predictive powers of Big Data in preventing these adverse events. Big Data could provide opportunities for virtually every sector of the economy, Podesta suggests, and could make the government more efficient. The report of the US Working Group recognized in addition that Big Data carries risks, noting the fact that 'how we protect our privacy and other values in a world where data collection is increasingly ubiquitous and where analysis is conducted at speeds approaching real time.'

33 The opportunities for Big Data can be discussed relatively briefly; they follow from the field of application as discussed earlier. The first opportunity that Big Data offers lies in improving the service to the citizen or customer, improving transparency in the public or private sector and giving more control to individuals. Second, particularly in the private sector, it is expected that Big Data will lead to substantial growth in the number of companies, especially start-ups, the number of jobs and the profits generated by those companies. For example, according to the roadmap developed by the Comité de Pilotage de la Nouvelle France Industrielle (Steering Committee of the New Industrial France) headed by the French Minister for Industry, Big Data activities in France represented €1.5 billion in 2014 and would reach approximately €9 billion in 2020, with Big Data activities also generating an additional 137.000 jobs. The EDPS report on Big Data also stresses the economic potential of Big Data. 'According to the OECD, 'Big Data related' mergers and acquisitions rose from 55 in 2008 to 134 in 2012. The internet sector is hugely successful with revenue per employee in 2011, among the top 250 companies, of over $900 – over twice as high as for the ICT industry overall (OECD). Internet companies could enjoy 'economies of scope', network effects of more data attracting more users attracting more data, culminating in winner-takes-all markets and near monopolies which enjoy increasing returns of scale due to the absolute 'permanence' of their digital assets.'[13]

34 Finally, Big Data can also be used for achieving the specific objectives of organizations, institutions and government departments. Yet, the question is to what extent Big Data is actually used within the public sector. The underlying research for this article seems to indicate that most countries and DPAs mainly recognize the opportunities for Big Data in the private sector, in relation to economic growth, stimulating businesses and increasing the number of jobs. The use of Big Data by the government, and especially by governmental institutions involved with maintaining public order or protecting national security, is viewed with skepticism. The Hungarian DPA, for example, emphasizes that Big Data is primarily used in the business sphere, such by as banks, supermarkets, media and telecommunication companies. In similar fashion, the Luxembourg DPA states explicitly that it has no knowledge of prominent examples of the use of Big Data in the law enforcement sector or by police or intelligence services in Luxembourg, but points out that other actors do engage with Big Data. The Norwegian DPA argues along the same line: 'There is, as far as we know, no usage of Big Data within the law enforcement sector in Norway. In 2014, the intelligence service addressed in a public speech the need to use Big Data techniques in order to combat terrorism more efficiently. However, politicians across all parties reacted very negatively to this request and no formal request to use such techniques has since been launched by the intelligence service. The companies that are most advanced when it comes to using Big Data may be found within the telecom (e.g. Telenor) and media (e.g. Schibsted and Cxence) sectors. The tax and customs authorities have also initiated projects in which they look at how Big Data can be used to enhance the efficiency of their work.'

35 In similar fashion, the Slovenian DPA stresses that it has not seen prominent examples of the use of Big Data in Slovenia; it suggests that Big Data applications are mainly of interest in insurance, banking and electronic communications sectors, mostly to combat fraud and other illegal practices. Another important field is scientific and statistical research. 'Law enforcement use is to our knowledge currently at development stages (e.g. in the case of processing Passenger Name Records), whereas information about the use of Big Data at intelligence services is either not available or confidential in nature.' The Swedish DPA states that it has not carried out any specific supervision related to the concept of Big Data and does not have any statistics or specific information on how this is used. 'In our opinion, the law enforcement sector does not use Big Data. Their personal data processing is strictly regulated in terms of collection of data, limited purposes, etc.' Finally, the British DPA indicates that it knows 'that companies are actively investigating the potential of Big Data, and there are some examples of Big Data in practice, such as the use of telematics in motor insurance, the use of mobile phone location data for market research, and the availability of data from the Twitter 'firehose' for analytics. We do not have any specific information on the use of Big Data in law enforcement or security.'

36 Noteworthy is that many DPAs suggest that Big Data is used particularly in the private sector and less so in the public sector – in particular, the use

13 <https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Big%20data/14-07-11_EDPS_Report_Workshop_Big_data_EN.pdf>.

of Big Data for security-related activities by the government is rejected. Only a few DPAs, such as the Dutch DPA, refer to the use of Big Data by the government for security purposes. The desk research, however, reveals a different picture, showing that governments do, indeed, use Big Data technologies, including for security purposes. Australia is an example of a country that is already quite well-advanced in using and applying Big Data processes. Among other things, it operates a prototype of the 'Border Risk Identification System' (BRIS). This system can be used at international airports to better estimate which travelers might cause problems. Reference can also be made to the 'Developmental Pathways Project', in which data on children from a variety of sources are linked. Among other things, an assessment will be made of the influence of factors relating to family and the environment on the health of children, the risk of juvenile delinquency, and education. Finally, there is a data tool, Vizie, which has been designed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO), an Australian government corporate entity. This tool follows activity on social media and analyses social media behaviour. A number of government agencies and public sector actors would also like to use this tool, at least according to CSIRO. Some examples can also be found of trials with Big Data in the area of security in the United States. For example, police forces used Big Data analytics to predict the odds that an individual will become involved in criminal activity. An example is Philadelphia, where the police used a tool to predict the chance of repeated offences. In addition, as indicated in the previous paragraph, countries such as Brazil, Israel and the United Kingdom promote the use of Big Data by the police, the intelligence and security services, and the military.

37 All in all, no clear picture has yet emerged as to where the opportunities for the use of Big Data lie. It seems clear that both the public and private sectors agree that Big Data will be used in the private sector and will lead to economic and jobs growth. There is less certainty about both the desirability and effectiveness of the use of Big Data by the government, particularly for security-related purposes. This also relates to the questions that have already been raised regarding the effectiveness of Big Data-type data collections by intelligence services such as the NSA in the United States in the fight against terrorism. Yet, a number of countries have actually implemented such projects involving the intelligence services, the armed forces and the police; for example, in connection with predictive policing. In conclusion, it seems advisable that regulators make an explicit assessment of the desirability and effectiveness of the use of Big Data in the public sector, especially when used for the promotion of national security or public order.

G. What are the dangers of Big Data?

38 This study shows that the dangers of Big Data are mainly assessed along two lines: first, a possible violation of the right to privacy and/or the right to data protection, and second, the danger of discrimination and stigmatization. With regards to the first point, most countries appear to be well aware of the risks that Big Data might pose for the privacy of citizens. For example, the current legal framework is based on the principles of purpose and purpose limitation. Article 7 of the EU Data Protection Directive contains an exhaustive list of the legitimate grounds for processing ordinary personal data; Article 8 does the same with regard to the processing of sensitive personal data (e.g. about race, religion, sexual orientation, etc.). Article 6 states that personal data must be processed fairly and lawfully, and must be collected for specified, explicit and legitimate purposes, and not further processed in a way that is incompatible with those purposes. The prohibition on further processing for different purposes is also known as the 'purpose limitation principle', from which it follows that 'secondary use' is in principle not permitted. The results of both the desk research and the survey show that it is this principle (along with the data minimization principle) that is cited the most when it comes to the tension between Big Data and data protection. Big Data processes often have no fixed purpose – large amounts of data are simply collected and it may only become clear what the value or potential use of that data is after it has been collected. Moreover, in Big Data analysis, different kinds of databases with different types of data are often linked or merged. The original purpose for which the data was collected is then lost. For example, the Swedish DPA argues that the concept of Big Data 'is used for situations where large amounts of data are gathered in order to be made available for different purposes, not always precisely determined in advance.'

39 The second principle that is often mentioned is the principle of data minimization. This principle requires that as little data as possible should be collected, and that the amount of data should, in any event, not be excessive in relation to the purposes for which it is collected. Additionally, personal data must be removed once the goal for which they were gathered has been achieved, and data should be rendered anonymous when possible. This principle, which mainly follows from Article 6 of the Data Protection Directive, obviously clashes with Big Data. The core idea behind Big Data is that as much data as possible is collected and that new purposes can always be found for data already gathered. Data can always be given a second life. This also challenges the requirement that data should be deleted or anonymized when it is no longer needed for achieving the purpose for which it was collected.

Almost all DPAs mention this principle when it comes to the dangers of Big Data. The Luxembourg DPA, among others, refers to a decision in which it stressed the importance of a retention period for data storage. The Dutch DPA summarizes the tension between Big Data and data minimization in very clear terms: 'Big Data is all about collecting as much information as possible.'

40 Articles 16 and 17 of the Data Protection Directive espouse the principle that data should be treated confidentially and should be stored in a secure manner. Many DPAs also mention this principle when discussing the dangers of Big Data; this holds especially for countries and DPAs that establish a link between Big Data and Open Data. The Slovenian DPA, for example, argues that the 'principles of personal data accuracy and personal data being kept up to date may also be under pressure in Big Data processing. Data may be processed by several entities and merged from different sources without proper transparency and legal ground. Processing vast quantities of personal data also brings along higher data security concerns and calls for strict and effective technical and organizational data security measures.'

41 The current framework also requires that the data that is accurate and kept up-to-date. This ensures that profiles created of or applied to an individual person, and any decisions taken on the basis of them, are appropriate and accurate. This study shows that many countries are aware of this tension and that DPAs are concerned about how this principle can be maintained in Big Data processes. Often, Big Data applications do not revolve around individual profiles, but around group profiles; not around retrospective analyses, but around probability and predictive applications with a certain margin of error. Moreover, it is supposedly becoming less and less important for data processors to work with correct and accurate data about specific individuals, as long as a high percentage of the data on which the analysis is based provides a generally correct picture. 'Quantity over quality of data', so the saying goes, as more and more organizations become accustomed to working with 'dirty data'. In the public sector, too, it seems that working with contaminated data or unreliable sources is becoming more common. Examples include the use by government agencies of open sources on the Internet, such as Facebook, websites and discussion forums. The Dutch DPA, for example, refers to the fact that in Holland, there 'has been a lot of media attention for Big Data use by the Tax administration scraping websites such as Marktplaats [an eBay-like website] to detect sales, mass collection of data about parking and driving in leased cars, including use of ANPR data, and profiling people to detect potentially fraudulent tax filings.'

42 An important principle of the Data Protection Directive and the upcoming General Data Protection Regulation is transparency. It includes a right of the data subject to request information about whether data relating to him/her are processed, how and by whom; the controller has a duty to provide the data subject with this information on its own initiative. This principle is also at odds with the rise of Big Data, partly because data subjects often simply do not know that their data is being collected and are therefore not likely to invoke their right to information. This applies equally to the flipside of the coin: the transparency obligation for data controllers. For them, it is often unclear to whom the information relates, where the information came from and how they could contact the data subjects, especially when the processes entail the linking of different databases and the re-use of information. As the Slovenian DPA puts it: 'Big Data has important information privacy implications. Information on personal data processing may not be known to the individual or poorly described for the individual, personal data may be used for purposes previously unknown to the individual. The individual may be profiled and decisions may be adopted in automated and non-transparent fashion having more or less severe consequences for the individual.'

43 The current legal system also puts much emphasis on subjective individual rights and does so to an increasing degree. For example, the forthcoming Regulation gives data subjects additional individual rights, such as the right to be forgotten and the right to data portability. In their response to the survey, DPAs also frequently referred to the principle of informed consent. Individual rights traditionally also come with individual responsibility, namely to protect individual rights and to invoke them if they are undermined. The question is whether this focus can be maintained in the age of Big Data. It is often difficult for individuals to demonstrate personal injury or an individual interest in a case; individuals are often unaware that their rights are being violated, even if they do know that their data has been gathered. In the Big Data era, data collection will presumably be so widespread that it is impossible for individuals to assess each data process to determine whether it includes their personal data; if so, to determine whether or not the processing is lawful; and, if that is not the case, to go to court or file a complaint. This tension appears both from the desk research and from the output of the survey. The British DPA holds, for example, that it 'may be difficult to provide meaningful privacy information to data subjects, because of the complexity of the analytics and people's reluctance to read terms and conditions, and because it may not be possible to identify at the outset all the purposes for which the data will be used. It may be difficult to obtain valid consent, particularly in circumstances where

data is being collected through being observed or gathered from connected devices, rather than being consciously provided by data subjects.'

44 Finally, the current system is primarily based on the legal regulation of rights and obligations. Big Data challenges this basis in several ways. Data processing is becoming increasingly transnational. This implies that more and more agreements must be made between jurisdictions and states. Making this legally binding is often difficult due to the different traditions and legal systems. Rapidly changing technology means that specific legal provisions can easily be circumvented and that unforeseen problems and challenges arise. The legal reality is often overtaken by events and technical developments. The fact that many of the problems resulting from Big Data processes, as also highlighted by a number of DPAs, predominantly revolve about more general social and societal issues makes it difficult to address all the Big Data issues within specific legal doctrines, which are often aimed at protecting the interests of individuals, of legal subjects. That is why more and more national governments are looking for alternatives or additions to traditional black letter law when regulating Big Data – for example, self-regulation, codes of conduct and ethical guidelines. The DPA of the United Kingdom states, for example, that it is notable 'that there is some evidence of a move towards self-regulation, in the sense that some companies are developing what can be described as an 'ethical' approach to Big Data, based on understanding the customer's perspective, being transparent about the processing and building trust.'

45 Besides privacy and data protection principles, DPAs also place a good deal of emphasis on profiling and the risk of discrimination, stigmatization and inequality of power resulting from Big Data. The desk research shows that a number of countries specifically acknowledge this danger. The best overview of these types of dangers is provided in the Working Paper 'Big Data and Privacy: Privacy principles under pressure in the age of Big Data analytics' by the International Working Group on Data Protection in Telecommunications. Four points are made in the working paper in this respect. First, there is a risk of power imbalance between those that gather the data (multinationals and states) and citizens. Second, there is a risk of determinism and discrimination, because algorithms are not neutral, but reflect choices, among others, about data, connections, inferences, interpretations, and thresholds for inclusion that advances a specific purpose. Big Data may, the Working Group makes clear, consolidate existing prejudices and stereotyping, as well as reinforce social exclusion and stratification. Third, there is the risk of chilling effects, which is the effect that people will restrict and limit their behavior if they know or think that they might be surveilled. Fourth and finally, the Working groups signal the chance of echo chambers, which may result from personalized advertising, search results and news items. 'The danger associated with so-called 'echo chambers' or 'filter bubbles' is that the population will only be exposed to content which confirms their own attitudes and values. The exchange of ideas and viewpoints may be curbed when individuals are more rarely exposed to viewpoints different from their own.'[14]

46 It, therefore, appears that in addition to opportunities, there are significant risks associated with Big Data processes. It should be emphasized that these threats again vary with respect to their impact on citizens according to their application. Instances of discrimination are always problematic, but if the police discriminates, this may obviously be more serious than in the case of personalized advertisements. Consequently, when regulating Big Data, account should be taken of the likelihood and the magnitude of potential problems relating to privacy and/or discrimination, and this must be weighed against the potential benefits.

H. Are the current laws and regulations applicable to Big Data?

47 Both the desk research and the results of the survey show that in most countries, the current rules in the area of privacy and data protection, as developed in their respective jurisdictions, are applied to Big Data processes. There is Germany with its distinctive personality right, the United States without an umbrella law for the regulation of privacy, but with sectoral legislation, and most other countries with relatively similar rules concerning privacy and data protection. In addition, a number of countries have specific laws on telecommunications and special rules for organizations such as the intelligence services and archives. In Australia, for example, there is specific regulation covering data matching in terms of tax records by governmental agencies, in which protocols are established for linking this data. Government departments working with files from the tax department must fulfill the requirements of the 'Data-matching Program (Assistance and Tax) Act 1990'. There are also mandatory guidelines for the implementation of the data-matching program.

48 It appears that current legislation is generally applied to Big Data projects, including in several court cases. In July 2015, for example, the French Constitutional Court, the Conseil Constitutionnel, gave its opinion on the French law governing the

14 <www.datenschutz-berlin.de/attachments/1052/WP_Big_Data_final_clean_675.48.12.pdf>.

intelligence and security services. In this ruling, the court specifically stated which provisions of this law are in line with the French Constitution and which parts or provisions of the law are not. Some provisions were declared unconstitutional, including a provision regarding the permission given by the Minister to monitor communications sent from abroad or received from abroad. In the United States, the case of the United States v Jones from 2011 may be of importance because this lawsuit had a limiting effect on the large-scale data gathering of location data by the police. In ACLU v Clapper, the Second Circuit Court of Appeals ruled that the mass collection of metadata about phone records by the NSA is illegal – this activity is not covered by section 215 of the Patriot Act. Meanwhile, however, the Foreign Intelligence Surveillance Court has ruled that the collection of metadata may continue. In the United Kingdom, in the case of Google Inc. v Vidal-Hall & Others, the Court of Appeal was asked to rule on the interpretation of the Data Protection Act 1998. The case revolved around the complaint by users of Apple's Safari browser, who believed that Google was gathering data through that browser in violation of the Data Protection Act 1998. The Court ruled that browsing information may be personal information and abuse of personal information should be considered as a tort.

49 From the survey among the DPAs, it also appears that current legislation is considered to be generally applicable to Big Data. They mostly refer to the national implementation of the Data Protection Directive. Yet, there are a number of countries with specific laws. Because the Estonian DPA sees Big Data as part of the Open Data movement, it refers to the Open Data legislation, namely the Public Information Act, which is currently pending in Parliament. In Hungary, the Information Self-Determination and Freedom of Information ('Privacy Act') applies. The Swedish DPA refers to special legislation for public services, such as the tax authorities, and to telecommunications law which partially constitutes an implementation of the European e-Privacy Directive. The survey also shows that the current legislation is applied in legal cases by national courts and in the opinions of the DPAs. The Belgian DPA refers to its advice on profiling, the DPA of Luxembourg to a report on smart metering and the Dutch DPA to lawsuits regarding the Tax Authorities and the use of data collected by the police through traffic cameras operated by the Tax Authorities.

50 In conclusion, it seems that the current legislation is generally declared to be applicable to Big Data; both courts and DPAs have successfully applied current principles when assessing Big Data-related projects. This should be taken into account when regulating Big Data. Replacing the current regulation with new 'Big Data' regulation would be to throw the baby out with the bathwater. If additional regulation is required, it seems more logical to develop new rules that could be applied in addition to the current regulatory framework. Whether, and to what extent, there is a need for such additional legislation will be discussed next.

I. Is there a need for new legislation for Big Data?

51 It is evident from the foregoing sections that in most countries, Big Data initiatives are treated under existing legislation with regard to issues such as privacy and data protection. Furthermore, the DPAs are agreed that the current data protection principles must be maintained. The Slovenian DPA, for example, explicitly points out that Big Data brings substantial challenges 'for personal data protection and these challenges must firstly be well understood and adequately addressed. In our view, new concepts and paradigms, such as cloud computing or Big Data should not lower or undermine the current levels of data protection as a fundamental human right. Existing central data protection principles, such lawfulness, fairness, proportionality, rights of the data subjects and finality should not be undermined with the advent of Big Data. The rights of the individuals to informational self-determination should be cornerstone in modern information society, protected by modern data protection framework delivering efficient data protection for the individual, while allowing lawful and legitimate interests, often also in the interest of the individual, to be attained.' Yet, most DPAs are also aware of the fundamental clash between Big Data and data protection principles, as discussed previously.

52 It is remarkable from the survey it appears that despite this fact, as of yet, little new legislation seems to be being developed that specifically addresses the new dangers posed by Big Data. Some DPAs refer to the forthcoming General Data Protection Regulation and indicate that they hope that those rules will help them to adequately curb the dangers of Big Data. For example, the British DPA suggests 'that the proposals for the new EU General Data Protection regulation incorporate some of the measures we have identified as being important in ensuring compliance in Big Data e.g. clearer privacy notices, privacy impact assessments and privacy by design. We welcome the fact that these measures are being foregrounded, although we are concerned that that they should not be seen as simply a bureaucratic exercise.' Moreover, the Estonian parliament is discussing new legislation on Open Data (including Big Data). Also, a number of DPAs refer to co-regulation and self-regulation as a possible solution.

53 Yet, the desk research supports the idea that governments are, in fact, actively thinking about new legislation, partly because current laws are seen as hindering technological innovation. Japan may be a case in point here. In 2013, the Strategic Headquarters for IT produced an amendment to various statutory provisions on privacy and data protection: 'Directions on Institutional Revision for Protection and Utilization of Personal Data'. A summary containing the main points of its policy, issued in 2014, discusses technological developments, including Big Data, that have occurred since the introduction of the Data Protection Act of 2003. According to the Strategic Headquarters for IT, there are now several barriers to the use of personal data. Furthermore, even organizations that respect the law and do not infringe rights are worried about criticism over potential privacy violations and the use of personal data; as a consequence, data are not used optimally. The growth envisaged by the Japanese government can only be achieved if personal data is used optimally and if Big Data flourishes. That is why the government wants to remove these barriers. An environment must be created in which violations of rights are prevented and in which personal information and privacy are protected, but in which, at the same time, personal information can be used for innovation. Furthermore, the UK Parliament has commissioned a study on the legislative framework for sharing data between public authorities. In July 2014, a commission published a report with three recommendations, suggesting among other things that the legal reform should go beyond simply stipulating rules for the sharing of data between public authorities; it should also regard the sharing of information between government agencies and organizations with public tasks. Finally, reference can be made to Germany. The Minister of the Interior has proposed a new principle for forthcoming legislation: the minimization of risk. He has also announced that Germany will propose the inclusion of provisions about pseudonymisation and profiling.

54 Consequently, when answering the question of whether it is desirable to formulate new rules for Big Data processes, three specific issues seem important. First, almost all countries and DPAs acknowledge that Big Data poses new and fairly fundamental risks to the current regulatory framework, and in particular the underlying principles. Second, the current regulatory framework is perceived as being (too) restrictive in relation to the deployment of new technologies and technological innovation, particularly in the private sector. Thirdly, many stakeholders are unsure how the current regulatory framework should actually be applied and interpreted in relation to Big Data. Two dangers might follow from this: on the one hand stakeholders, for fear of breaking the law, might forgo many technological innovations and data uses that would in fact be legitimate. On the other hand, parties might use – or rather, abuse – the existing grey area to deploy certain technologies that would not be in accordance with the current regulatory framework. Whether and how a new regulatory framework might provide a solution for these challenges needs to be assessed carefully by regulators.

J. What concept should be central to Big Data regulation?

55 In short, a diffuse picture emerges, with respect to the extent to which developing a special regulatory Big Data regime is necessary or even desirable. What is evident is that regulating Big Data will be especially difficult for two reasons. First, it is difficult to choose a good starting point for the regulation of Big Data; this will be discussed in this section. Second, it will be difficult to pinpoint a specific person or institution to serve as data controller or, more generally, a natural or legal person that is responsible for compliance with the regulatory principles in Big Data processes. This will be discussed in the next section. Regarding the starting point, it should be noted that the current regulation is primarily based on the individual and their interests – this holds for human rights such as privacy and for data protection, which is based on the concept of 'personal data', i.e. data that enables someone to identify or individualize a natural person. However, Big Data processes do not so much revolve around the storage and processing of data at an individual level – rather, the trend is to work increasingly with aggregated data, general patterns and group profiles. Consequently, it is questionable whether the focus on the individual, on personal data, can still be maintained in the Big Data era. The statistical correlations and group profiles do not qualify personal data, but can be used inter alia to alter, shape or influence the living environment of people to a great extent. Furthermore, the trend towards the use of metadata also ties into this problem, because it is unclear to what extent metadata will always qualify as personal data.

56 In addition, many DPAs point out that in Big Data processes, personal data or profiles may be created through the use, combination or analysis of data that do not qualify as personal data. The EPDS states explicitly that a lot of data is gathered in Big Data processes, but also suggests: 'Not all of these data are personal, but many players in the digital economy increasingly rely on the large scale collection of and trade in personal information.' The Working Party 29 states that: 'In addition, Big Data processing operations do not always involve personal data. Nevertheless, the retention and analysis of huge amounts of personal data in Big Data environments require particular attention and care. Patterns

relating to specific individuals may be identified, also by means of the increased availability of computer processing power and data mining capabilities.' The DPA from Luxembourg suggests that Big Data 'allows for the correlation of information which previously could not be linked. From a data protection point of view it can raise many concerns, when it contains personal data, such as the respect of data subjects' rights – for example in the context of data mining – and their ability to exercise control over the personal data or the respect fundamental principles of data protection such as that of data minimization or purpose limitation. Moreover practices such as linking separate databases or computer analytics can turn anonymous data or any kind of non-identifiable information into personal data which would need to be protected under data protection law.' As a final example, reference can be made to the DPA from Slovakia, which argues: 'As a research topic, we would like to suggest examining boundaries between personal and non-personal information. In the Big Data environment, you are able to connect non-personal information and, based on this information, identify the data subject which represents potential risk to rights of the data subjects.'

57 Consequently, it is questionable whether the individual, individual interests and concepts such as personal data, which are explicitly linked to individual natural persons, still serve as a good starting point for building a regulatory framework in the Big Data era. Irrespective of whether the regulator chooses to leave the current legislation largely intact, whether it opts to amend current legislation or chooses to develop a new Big Data framework, it seems that at a certain point in time it will be necessary to address the fact that it is increasingly difficult to take 'personal data', or a related concept, as the basis for rules and obligations. It should finally be noted that the nature of the data is also becoming less and less static; rather, data increasingly goes through a lifecycle in which its nature might change constantly. While the current legal system is focused on relatively static stages of data, and linked to them specific forms of protection (e.g. for personal data, sensitive data, private data, statistical data, anonymous data, non-identifying information, metadata, etc.), in reality, data go through a circular process: data is linked, aggregated and anonymized and then again de-anonymized, enriched with other data and profiles, so that it becomes personally identifying information again, and potentially even sensitive data, and is then once again pseudonymised, used for statistical analysis and group profiles, etc.

K. How should the responsibilities be distributed?

58 A final question that needs to be answered when regulating Big Data is who should bear responsibility for enforcing the rights and obligations; or, in data protection terms, who should be the data controller. This issue exists irrespective of whether the regulator chooses to leave the existing legislation untouched, seeks to amend current legislation or opts to develop new Big Data legislation. The problem of allocating responsibility was prominent both in the desk research and the survey and, in general, manifests itself on three different levels. Firstly, there was already a fair degree of awareness of the increasingly transnational nature of data processing activities. The problem is that different countries have different levels of data protection. The danger is that private parties will settle in those countries where the regulatory pressure is low. But public sector organisations might act in similar ways as well. For example, in the Netherlands, there is a court case pending on the cooperation between the Dutch intelligence services and their counterparts abroad. Although the Netherlands limits the capacities of its intelligence services to collecting information about Dutch citizens, the US intelligence services, which are less constrained regarding the collection of data on Dutch nationals, might collect such data and then pass it on to the Dutch intelligence services. This might work the other way around, too. Consequently, intelligence services might effectively circumvent the rules that apply to them, by cooperating with other international actors that are not bound by those rules.

59 Secondly, it is also apparent from the desk research that there is increasing cooperation between the public and the private sectors, voluntary or otherwise. For example, in Australia, there is collaboration between industry and academia; the Brazilian police use a system that was originally developed by Microsoft and the New York police; China stresses the need for cooperation between the public and the private sector; and the Estonian DPA refers to the cooperation between public and private parties with respect to the development of regional policies. Again, the question is which responsibilities should be borne by which party. Often, it is not clear at first sight what role an organization has played in the value chain of the data processing activity. Also, very different regulatory frameworks often apply to public sector and private sector institutions, as also noted by a number of DPAs in their response to the survey.

60 Thirdly and finally, there is also a trend towards sharing data and linking databases between governmental organisations. This implies that

governmental agencies that have a limited legal capacity to gather and store data may still obtain a wealth of information from other governmental organisations that have a greater legal capacity to gather and store such data. For example, the Dutch DPA refers to a lawsuit that revolves around the use by the Tax Authorities of information gathered by the police. Again, the question is which party should bear responsibility for enforcing the legal regime and the restrictions it imposes. More generally, it should be noted that data flows are becoming more fluid and elusive, meaning that more and more organizations are involved and more and more parties share partial responsibility. This complicates the attribution of responsibilities.

61 Just as the lifecycle of data is becoming increasingly circular, so the division of responsibilities is a clearly shifting from a rather static reality, in which one party collects and processes data, is the main controller of the data and should therefore enforce the different rules and obligations encapsulated in the legislative framework, to a world in which different parties collect, share and link data; in which parties from the private and the public sectors cooperate; in which different governmental institutions share data and databases; and in which international data flows are becoming increasingly common. Consequently, when regulating Big Data, it seems logical to make a choice regarding the distribution and attribution of responsibility. The regulator may, despite these developments, opt for a relatively static model in which one party is the main controller and is responsible for enforcing the legal obligations; or it could opt for a more dynamic model, in which the distribution and attribution of responsibilities is shared and might change as the nature of the data processing activities change. The Data Protection Directive could provide a basis for the latter option, as it defines the controller as 'the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data.'

L. Summary of main findings

1. *What is the definition of Big Data?* It is impossible to give an exact definition of Big Data. From the research conducted for this report, it follows that a number of different phases must be taken into account when defining Big Data, namely the collection, analysis and use of data. Big Data revolves around collecting large amounts of data (volume), from varied (variety) and often unstructured data sources. Big Data refers to the speed (velocity) of the analyses, often with the use algorithms, machine learning and statistical correlations. The results are often predictive in nature (predictive analytics) and are formulated on a general or group level. The use of the results is usually carried out through profiling. Many of the definitions used in the field contain some of these concepts; none of them mentions all of them. It therefore seems premature to give an exact and precise definition. Two things must be taken into account when regulating Big Data. First, the fact that Big Data cannot be easily defined; this will complicate the making of specific Big Data regulations or laws. Second, the fact that the Big Data process occurs at three levels: collection, analysis and use. These are communicating vessels and must be treated and possibly regulated in connection to each other.

2. *Is Big Data an independent phenomenon?* Big Data should be viewed in its interrelationship and in conjunction with other phenomena. Big Data is part of and in some sense the umbrella term for many of the technological developments that are taking place right now. Terms that are often mentioned as part of the definition of Big Data or as related to Big Data are: Open Data, Re-Use, Internet of Things, smart applications, Profiling, Algorithms and Cloud Computing. Also, machine learning, commodification, datafication, securitization and risk society are sometimes brought up. If the government chooses to regulate Big Data, it should take into account that Big Data is not an isolated phenomenon, but is a development which by its nature very strongly correlates with a number of technical, social and legal developments that are already taking place. The government will have to take a holistic approach when regulating Big Data and related phenomena.

3. *Big Data: fact or fiction?* Right now, Big Data plays a small role, but it will, nevertheless, become increasingly important as time progresses. Consequently, Big Data should not be seen as either an actual practice or fiction, a hype that will blow over, but mainly as a trend that will play a major role of significance in 5 or 10 years from now and will have a significant impact on the operations of governments and businesses and will significantly affect the everyday life of citizens. Only then will many of the effects of Big Data become clear. The government should develop future-oriented policies that follow and preferably anticipate this trend. If it starts to regulate Big Data only in about 5 or 10 years, many of the projects will already have started. The potential negative consequences will have materialized, and it will be difficult to adjust or cancel the projects that have already started. It should also be remembered that good and clear regulation can contribute to innovation

and the use of Big Data. Because the frameworks for Big Data projects are not always clear at the moment, some government agencies and companies are reluctant to use new technologies for fear of breaking the law. New regulation may give more clarity on this point.

4. *What is the scope of Big Data?* Generally speaking, the use of Big Data can be divided into three types. First, the use of Big Data for specific government tasks - examples include the use of Big Data by intelligence services, the police, tax authorities and other public bodies; for example, in the context of formulating economic policies. Second, the use of Big Data by the private or semi-public sector for achieving their tasks and/or goals. Examples include the use of Big Data by companies to create risk profiles, the use of Big Data in the healthcare sector and the use of Big Data in scientific projects. Thirdly, Big Data is used by both governments and companies to improve their service to citizens and customers - for example, this could involve increasing the transparency of activities, strengthening the control citizens have on data processing, etc. The regulation of Big Data will have to take into account the impact the use of Big Data has on the individual, the type of data and data analysis that is used, and the possible danger of a mismatch between a general profile and a specific individual. A distinction must be made between the type of body that executes Big Data projects and the specific purpose for which it is used - the general interest served by the use of Big Data should also have an impact on what is legally permissible.

5. *What are the opportunities for Big Data?* The first opportunity that Big Data offers is to improve the service to the citizen or customer, to improve transparency in the public or private sector, and to give more control to individuals. This practice is generally unproblematic as it serves the interests of the citizen. The second possibility is the use of Big Data in the private and semi-public sector. Big Data is expected to provide a substantial growth in the number of companies, especially start-ups, the number of jobs and the profits generated by these companies. Both the public and the private sector see the biggest opportunities for Big Data in this field of application. However, the use of Big Data in the private sector is not unproblematic. When advertisements or services are personalized through the use of Big Data, the impact on the individual will be relatively small, but this may be different when risk profiles are created by banks or health insurers when deciding who may get a loan or insurance, and on what condition. There exists controversy about the question whether governments should make use of Big Data, especially with respect to security-related purposes. On the one hand, some countries already use Big Data, also for security-related purposes. On the other hand, there are considerable doubts about both the efficacy and the desirability of these projects. The regulator should particularly assess the efficacy and the desirability of the use of Big Data by the public sector institutions when used for security-related purposes. With regard to the use of Big Data by the private sector, a distinction should be made between the type of application.

6. *What are the dangers of Big Data?* This study shows that the dangers of Big Data are assessed mainly along two lines. First, a possible violation of the right to privacy or the right to data protection. Second, the danger of discrimination and stigmatization. Regarding the first point, it appears from underlying research that most countries are well aware of the risks to the privacy of citizens. With regard to the risk of discrimination and stigmatization, this appears to be true to a lesser extent. Consequently, the government will have to weigh the dangers of a breach of privacy and of discrimination against the potential benefits. It should be stressed that both the right to privacy, the right to data protection and the right to freedom from discrimination are fundamental human rights that may be limited only in exceptional circumstances, if necessary in a democratic society.

7. *Are the current laws and regulations applicable to Big Data?* From both the desk research and the results of the survey, it appears that, in most countries, the current regulations in the area of privacy and data protection are applied to Big Data processes. Germany with the distinctive personality right, the United States without an umbrella law for the regulation of privacy, but with sectoral legislation, and most other countries with relatively similar rules concerning privacy and data protection. In addition, a number of countries has specific legislation in the field of telecommunications; also, there are often special rules for organizations such as the intelligence services and archives. Current legislation is generally applicable to Big Data; both courts of law and DPAs are not empty-handed when confronted with Big Data-like processes. This should be taken into account by the government when regulating Big Data. Replacing the current regulation by new 'Big Data' regulation would be to throw the baby out with the bathwater. Rather, it should consider formulating new rules in addition to the current regulatory framework.

8. *Is there a need for new legislation for Big Data?* In most countries, the existing laws are applied to Big Data initiatives. Also, the DPAs are in agreement that the current privacy and data protection principles must be safeguarded. Yet, most DPAs are also aware of the fundamental tension between Big Data and data protection principles. It is remarkable that despite this fact, little new legislation seems to be developed that specifically addresses the new dangers posed by Big Data. Some DPAs refer to the upcoming General Data Protection Regulation and hope it will contain new rules that could help to tackle the dangers posed by Big Data. A number of DPAs refer to co- and self-regulation as a possible solution. Still, some countries seem to be thinking about new regulations for data processing techniques, such as Estonia, France, Japan and Great-Britain. This is partly motivated by concerns over the protection of privacy, but also by the thought that the current laws hinder technological innovation. When answering the question whether it is desirable to formulate new rules for Big Data processes, the government will need to take into account three issues. First, almost all countries and DPAs see new and fundamental risks for the current regulatory framework and, in particular, its underlying principles in the Big Data era. Second, it appears that the current regulatory framework is regarded by some to be too restrictive, muffling the use of new technologies and technological innovation, particularly in the private sector. Third, many parties are unsure how the current rules and laws should be applied to and interpreted in the light of Big Data processes. There are roughly two dangers: on the one hand, for fear of breaking the law, parties may forgo many technological innovations that would be legitimate to use; on the other hand, parties may abuse the existing gray area and take steps that circumvent basic constitutional principles. Whether and how a new regulatory framework can solve these problems needs to be considered by the government.

9. *What concept should be central to Big Data regulation?* Current regulations are often based on the individual and his interests - this applies to individual human rights and to data protection, which regulates the processing of personal data, that is, data that can identify or individualize a natural person. Since increasingly, data are not collected and processed at an individual level, and rather, use is made of aggregated data, which lead to general patterns or group profiles, the question is whether the focus on the individual can still be maintained. This ties up to the use of metadata – it is often unclear to what extent metadata can qualify as personal data. Finally, it should be noted that the nature of the data is less and less static and that data increasingly go through a circular life. While the current legal system is focused on relatively static stages of data and attaches to these stages a specific protection regime (such as for personal data, sensitive data, statistical data, private data, anonymous data, metadata, etc.), in practice, data go through a circular process: data are linked, aggregated and anonymized and then again de-anonymized, enriched with other data for the making of personal or even sensitive profiles, and then again pseudonymised, used for statistical analysis and group profiles, etc. It seems to go too far to simply regulate 'data', but the direct connection to a specific individual, such as is the case with 'personal data', also seems difficult to sustain in the Big Data era. The government will have to determine whether 'personal data' as a concept is still adequate to serve as a basis for data regulation in the Big Data era.

10. *How should the responsibilities be distributed?* Like the life cycle of data that is increasingly circular, with regard to the attribution of responsibilities, a clear shift may be seen from a world in which one controller collects, processes and uses the data and is, therefore, the party solely or primarily responsible for respecting the legal principles, to a world in which data are increasingly shared between governmental organizations, between the private and the public sector and between international public and private sector parties. With regard to the attribution and distribution of responsibilities in the Big Data era, the government has to make a principled choice. Will it, despite the observed trend, maintain the model in which one party has the sole or primary responsibility, and if so, who will bear the burden, or will it choose for a more dynamic model, and if so, how will the responsibility of the parties be divided and established?

Appendix I

	Is a specific definition of Big Data used?	Is Big Data used within the government?	Is there a public-private partnership?
Australia	For the purpose of the Big Data Strategy, the following definition is used: *"1. The data analysis being undertaken uses a high volume of data from a variety of sources including structured, semi-structured, unstructured or even incomplete data; and* *2. The size (volume) of the data sets within the data analysis and velocity with which they need to be analysed has outpaced the current abilities of standard business intelligence tools and methods of analysis".*	The Australian Public Service Big Data Strategy is one of the most prominent examples. This strategy, and accompanying documents, were drafted by the Australian Department of Finance. Parallel to this, a center for the entire the government was set up, headed by the Department of Finance, for improving the data analytics capacity of the government. In the Strategy, several current Big Data projects or pilots of Big Data projects are listed, such as: Border Risk Identification System (BRIS) and the Development Pathways Project.	There is a law that facilitates the use of data from the private sector for the tax authorities, called the Data-matching Program. This law can facilitate a public-private partnership.
Brazil	-	One of the most prominent examples from Brazil is the Big Data tool, 'DataViva', used by the government of the province of Minas Gerais. DataViva combines data from databases belonging to three Ministries and an U.N. database on trade, concerning exports and imports, labour and education, from all over the country. Another prominent example is the system that is used by the Sao Paulo police, 'Detecta'. Detecta is an intelligent system for monitoring crime. Large datasets held by the Sao Paulo police are combined in this tool and subsequently, Detecta makes connections between the data. The system gives of warning signals to relevant authorities and reveals patterns in the crimes committed in the region.	-
China	-	According to the State Council, Big Data is used to make the government more efficient. This entails more personalized service delivery by the government, greater efficiency in the administrative approvals process, with preference being given to companies with a good credit score and those with a poor credit rating being restricted. The premier of the State Council also announced that the government is working on Big Data. An example can be found in the new credit system that will be introduced in China. Another example is the judicial Big Data center, linking all China's judicial bodies.	-

France	-	This is unclear. The French government considered the future challenges for 2025 for the national mail delivery system. The research suggests that the government has five options for resolving these problems. One possible strategy is to focus the service delivery more on e-commerce and to use Big Data analytics to improve the chain of production. It is not yet clear which of these five directions is preferred	-
Germany	The Federal Ministry for Education and Research is the Ministry that is most concerned with Big Data in Germany. According to this Ministry, Big Data is synonymous with: "*den intelligenten Umgang mit solchen großen oder auch heterogenen Datenmengen*" (intelligent use of large or heterogeneous datasets).	This is unclear. There are investments and research projects concerning Big Data. In 2014, the Ministry announced that it would be providing financial support for the construction of two Big Data centers: the Berlin Big Data Centre and the Competence Center for Scalable Data Services in Dresden. In addition to building the two centers, the Ministry will promote further research in support of Big Data, as illustrated by the funding initiative launched in 2013. Specifically, the Ministry will focus attention on 'Industry 4.0' projects and on the bio- and geosciences. A research project focusing on Big Data is ABIDA ('Interdisziplinäre Analyse der gesamtgesellschaftlichen und wirtschaftlichen Folgen beim Umgang mit großen Datenmengen'), funded by the Ministry of Education and Research.	-
India	-	The Indian Ministry of Science and Technology has started a Big Data initiative. The Ministry lists four focus areas for the development of a sustainable data analysis system. Aadhaar is a government-wide project being implemented by the Unique Identification Authority of India. It involves the collection of biometric and demographic data of the Indian population. The Indian Government has not specifically labelled this as a Big Data project.	Not in the sense of a partnership, but the Indian government does make datasets publicly available online to make large amounts of non-sensitive data available to society.
Israel	-	C4i is the department of the IDF that is specifically engaged in information and computer technology. An interview with the commander of this unit makes clear that it is no longer just about passing on information to divisions of the armed forces. Rather, C4i should be seen as a tool which can be deployed in the area of Big Data analytics. The IDF makes use of several Big Data systems such a 'Crystal	The Israeli Ministry of Health sent out a tender in August 2015 for a partner in Big Data analytics. The Ministry has an enormous dataset containing all the medical data on the Israeli population as well as data on the health care system. The Ministry wants to put this dataset to good use and to be able to translate it into specific recommendations.

		Ball' and a GPS system to direct the troops.	
Japan	-	The Japan Science and Technology Agency (JST) is the body responsible for implementing the technology policy of the Japanese government. One of JST's research programmes, 'CREST', involves team-based research to achieve the strategic goals of the government. The programme involves research on Big Data, under the auspices of two main projects: 'Advanced Application Technologies to Boost Big Data Utilization for Multiple-Field Scientific Discovery and Social Problem Solving' and 'Advanced Core Technologies for Big Data Integration'.	There is not a specific partnership, but the sharing of data between the two sectors is encouraged by the government, especially data relating to earthquakes.
South-Africa	-	With the Square Kilometre Array (SKA), a large multi-radio telescope project, South Africa is seeking to put itself on the map as a Big Data hub. The data science capacity that comes with the SKA project must be provided by a network of universities, grouped together in the 'Inter-University Institute for Data-Intensive Astronomy (IDIA)'.	-
United Kingdom	-	The British government published a strategy for Big Data: 'Seizing the data opportunity. A strategy for uk data capability', and made several large investments in Big Data Research Councils. One of the projects funded by a Council is the 'Big Data for Law' initiative, allowing Big Data research on legislation. There are several Big Data projects scattered over various sectors, these projects are described in 'POSTnotes' by the Parliamentary Office of Science and Technology.	The government has founded several Big Data centers which are used by the private sector, in which data from the government sector and private sector is used, or in which researchers and the business sector work together. The British government also makes use of said data.
Unites States	The Podesta report refers to the definition given by Gartner and adds that: "More precisely, Big Datasets are 'large, diverse, complex, longitudinal, and/or distributed datasets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future".	In March 2012, the Obama Administration launched the 'Big Data Research and Development Initiative'. Under this initiative, six federal government departments and agencies announced the investment of 200 million dollars in additional improvements to the processing of enormous volumes of data. In the fact sheet dated 29 March 2012, 'Big Data Across the Federal Government', dozens of ongoing government projects and partnerships related to Big Data are mapped, in all sectors. Some examples can also be found of trials with Big Data in the area of security in the United States.	The US government appeals to the private sector to "join with the Administration to make the most of the opportunities created by Big Data. Clearly, the government can't do this on its own". Whether this should take the form of a partnership between both sectors remains unclear.

	To what goal is Big Data used by the government?	Which laws are especially relevant for Big Data?	Are there judicial decisions relating to Big Data?
Australia	According to the Australian Public Service Big Data Strategy, the strategy is intended to advance the possibilities of Big Data while safeguarding the privacy of the individual. Improving the possibilities for Big Data analytics for the government should lead to improved services and better policy advice. In this Strategy, the mission of the Australian government in relation to Big Data is described as: "The Australian Government will be a world leader in the use of Big Data analytics to drive efficiency, collaboration and innovation in the public sector".	The Freedom of Information Act 1982, the Archives Act 1983, the Telecommunications Act 1997, the Electronic Transactions Act 1999, the Data-matching Program (Assistance and Tax) Act 1990, the Privacy Act 1988, the Privacy Amendment (Enhancing Privacy Protection) Act 2012, the Privacy Regulation 2013.	-
Brazil	At first, the aim of the DataViva tool was to help in drafting economic policy, but it became clear that it offered opportunities as a Big Data tool as such; the relationships and dynamics that the tool exposes provide an insight into the economy for public and private actors and support them in their decision-making. The Detecta system is used to combat and prevent crime.	An amendment to the legislation on data protection is currently being developed. The government has released a draft bill for this law, entitled: "On the processing of personal data to protect the personality and dignity of natural persons".	-
China	There is an emphasis on the use of Big Data to make government services more efficient and to stimulate economic growth.	China does not have overarching privacy legislation such as is present in many European countries. At the end of 2012, the Chinese parliament drafted a resolution consisting of 12 articles and regulating privacy and data protection: the 'Decision of the Standing Committee of the National People's Congress to Strengthen the Protection of Internet Data'.	-
France	Big Data is highlighted by the French government as one of the key developments for modern reforms in French industry.	The 'Loi Informatique et libertés'1978, which has been amended several times since its introduction.	The highest French constitutional court, the Conseil Constitutionnel, issued a ruling in July 2015 regarding the French law governing the intelligence and security agencies. In this ruling, the court declared specifically which provisions of this law are in accordance with the French Constitution and which provisions or parts of provisions are not. What is for example permitted, subject to certain conditions, is the collection of data in real time in order to prevent terrorism, and obliging service providers to identify connections (the parameters of which are set out in the order) which suggest a terrorist threat.

Germany	Several research initiatives for Big Data are aimed at researching how Big Data can be used sensibly and how to handle Big Data. Big Data is seen as a great opportunity for the ICT sector and can improve the competitive position of the German business and science sector, but is also seen as "one of the major challenges" of our time.	The central data protection legislation in Germany is the Bundesdatenschutzgesetz, originally dating from 1990.	-
India	The Indian government uses its Big Data strategy to focus on a sustainable system of data analysis.	The "Information Technology (Reasonable security practices and procedures and sensitive personal data or information) Rules, 2011. A new bill is in the making, the 'Privacy Bill 2014'. In 2012 the Ministry of Science and Technology developed a national policy for data sharing and accessibility.	-
Israel	The focus of Big Data initiatives in Israel lies on making the best use possible of the government's data and using Big Data to protect the country and make the military system more efficient.	The right to privacy is enshrined in Section 7 of the Basic Law on Human Dignity and Liberty. In 1981 a law was also introduced which is tailored specifically to this right, the Protection of Privacy Law 5741 – 1981. To implement this this law, special legislation was drafted governing data flows from Israel to other countries. In 2010 an amendment to the privacy legislation was introduced, adding provisions relating to the security of databases.	-
Japan	The Japanese Prime Minister stated that in order to achieve its economic goals the Japanese government was among other things making changes to optimize the it sector. The law on the protection of personal data would be changed to make it easier to use personal information as part of Big Data. The 'it Strategic Headquarters', established within the Japanese Cabinet, published an open data strategy for the government, in which it argued that government data is a public asset and that the sharing and use of that asset should be encouraged.	The Act on the Protection of Personal Information from 2003, in 2013 amendments were made to this law, inter alia because of Big Data.	-
South-Africa	South Africa is seeking to put itself on the map as a Big Data hub, further goals of the Big Data project are to reduce poverty and improve the country's economic competitiveness.	The right to privacy is explicitly enshrined in Article 14 of the South African Constitution. The Protection of Personal Information Act 2013 is relevant.	-
United Kingdom	Big Data is used for various purposes, such as: creating efficient motorways and traffic flows, predicting crime, researching diseases and facilitating Big Data research on legislation. There is no	The Data Protection Act 1998, the Human Rights Act 1998 (section 8), the 2000 Regulation of Investigatory Powers Act and the Intelligence Services Act 1994.	In 2015, the case of Google Inc v Vidal-Hall & Others was heard by the Court of Appeal. The case related to data protection and the Data Protection Act 1998. The Court ruled that browser

	focus on one specific goal. The Minister for Universities and Science and the Minister for Skills and Enterprise state the following about data: Governments around the world must change the way they engage with citizens, the way they develop policy and deliver services, and the way they are held to account (...) The UK government is determined to position the UK to make the most of the data revolution."		information can be regarded personal data and that abuse of personal data should be regarded as a tort. With regard to data protection, the High Court pronounced a verdict in July 2015 in the case of Davis & Others v SSHD in relation to the Data Retention and Investigatory Powers Act 2014. In this case the Court declared this law partially invalid due to conflicts with European law, and specifically the section in which the competence is established to request telecommunications service providers to retain communications data.
Unites States	The Big Data review produced five overarching conclusions which can be seen as goals the government can aim for in following the report: First, more research must be carried out on the protection of privacy, and action should be taken in the area of legislation on the protection of privacy. Second, there should be more attention for the responsible handling of data collected in the context of education, especially data regarding children. Third, the federal government is advised to be on its guard for discrimination of citizens, which can be caused by Big Data analytics. Fourth, the authorities responsible for enforcement and safety are advised to make maximum use of the legal possibilities for Big Data analytics. The Big Data initiatives that are already in place focus on several goals, varying with the sector of the government that they are used within.	The United States does not have an overarching law for the regulation of privacy, and certainly not for the specific regulation of Big Data. Besides the constitutional protection, the United States has a system of sector-specific regulation of privacy risks. The Consumer Bill of Privacy Rights was introduced in 2012. This is not legislation in the sense of being enforceable, but more of a guideline for the business sector.	A court case on limiting the effects on large-scale location data collection by the police was The United States v. Jones from 2011. Another interesting case is Sorrell v. IMS Health Inc., which was also heard by the Supreme Court in 2011. In this case, involving the commercial use of medical data, the Court ruled that there is a limited scope for datamining when in breach of the freedom of expression. On 7 May 2015, the Second Circuit Court of Appeals ruled in ACLU v. Clapper that the large-scale collection of metadata concerning telephone records by the NSA is unlawful. However, the Foreign Intelligence Surveillance Court ruled that the collection of metadata could continue.

Appendix II Responses from the DPAs to the survey

	1. Are you familiar with the debate on Big Data? If so, how would you define Big Data?	2. Are there prominent examples of the use of Big Data in your country, especially in the law enforcement sector, by the police or by intelligence services?	3. Have you issued any decisions/ reports/opinions on the use of Big Data? If so, could you provide us with a reference and your main argument?
Belgium	We have no official national definition. However we follow closely the definitions; The EDPS states on its website "Big Data means *large amounts of different types of data produced at high speed from multiple sources, whose handling and analysis require new and more powerful processors and algorithms. Not all of these data are personal, but many players in the digital economy increasingly rely on the large scale collection of and trade in personal information. As well as benefits, these growing markets pose specific risks to individual's rights to privacy and to data protection"* (<https://secure.edps.europa.eu/EDPSWEB/edps/Consultation/big_data>) Also, the Working Party 29 has issued a general statement on Big Data. (<http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp221_en.pdf>) The Consultative Committee of the Convention 108 has appointed an expert that has to write a report on Big Data, expected to become public in 2016 (<www.coe.int/t/dghl/standardsetting/dataprotection/TPD_documents/OJ_TPD32(2015)_11%2006%2015_Fr.asp>)	Not to our knowledge for the indicated sectors in the strict meaning (there is no obligation to notify our DPA of such projects in these sectors). However, in the approach of the fiscal and social fraud, the projects and discussion on the use of Big Data or the steps in this process (profiling, data mining,...) exist since 2012. We have addressed several opinions since 2012 that address a part of the Big Data issue (mainly data mining and profiling)	On profiling by facebook : Aanbeveling 04/2015 van 13 mei 2015 uit eigen beweging met betrekking tot 1) Facebook, 2) de gebruikers van internet en/of Facebook alsook 3) de gebruikers en aanbieders van Facebook diensten, inzonderheid social plug-ins, gepubliceerd op <www.privacycommission.be/sites/privacycommission/files/documents/aanbeveling_04_2015.pdf> At the request of our Commission the inter-university research center EMSOC/SPION (see <www.law.kuleuven.be/icri/en/news/item/icri-cir-advises-belgianprivacy-commission-in-facebook-investigation>) conducted a detailed study into the way in which Facebook deals with its members' personal data. And that of citizens who do not use Facebook or who explicitly opted out of its service. On profiling of energy and water clients: Advies nr. /2015 van 17 juni 2015 betreffende Hoofdstuk II van het Ontwerp van wet houdende diverse bepalingen, betreffende de verbruiksgegevens van nutsbedrijven en distributiebeheerders
Croatia	The Republic of Croatia is familiar with the concept of Big Data, and a definition /explanation with which we most agree is from the text "What is really Big Data and where is it used?" By Luka Stepinac from 12. May 2014. published at the www.ictbusiness.info in which stands „Definition that we can find the most often refers to "3V": Volume - a large amount of data collected, processed and made available for analysis; Velocity - continuous collection of large amounts of data in real time; Variety - the data are available in various forms and sources, and in fact are usually unstructured, or, in one sentence, Big Data is a technology that enables the collection and processing of large amounts of structured and unstructured data in real time."It is	At this moment we do not have an appropriate/adequate information.	No.

	necessary to point out that the Republic of Croatia regularly monitors technological innovations which in most cases allows the use of information from the field of Big Data, and most often in commercial purposes.		
Estonia	Estonian Data Protection Inspectorate is familiar with the debate on Big Data. In our opinion Big Data could be defined as collected and processed open datasets, which are defined by quantity, plurality of data formats and data origination and processing speed.	Yes, some public sector authorities in cooperation with the private sector (e.g. mobile operators) and universities have applied Big Data to their analysis. For example, *Bank of Estonia* (Eesti Pank) and *Statistics Estonia* on tourism statistics, *Ministry of the Interior* with municipalities have used Big Data in the development of regional policy. Based on open datasets, private company *Big Data Scoring* provides background information to loan companies.	No.
France	The CNIL is familiar with the debate on Big Data and is actively working on the subject. In August 2014, a definition of the term 'Big Data' was adopted by the French General Commission on terminology and neology (*Commission générale de terminologie et de néologie*). The official translation of this term in French is '*mégadonnées*' and the definition is 'structured data or not whose very large volume require appropriate analytical tools'. The Gartner definition is also a reference: 'Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making and process automation'. With reference to this definition, three 'Vs' are generally associated with Big Data: volume, variety and velocity. Our Data protection authority (DPA), as other actors, considers that other 'Vs' are also relevant, in particular value and veracity. Many examples of Big Data operations involve processing of personal data, in various business sectors. The projects have different goals and use different categories of data. But, beyond this diversity of projects and objectives, the notion of 'Big Data' reveals a new approach of the data, appeared with the development of new storage and analytical capacities. And privacy challenges are associated to Big Data because, thanks to sophisticated algorithms, Big Data can ultimately be used to identify profiles, predict the behavior of individuals or groups of individuals, and take decision affecting them.	There are various examples of the use of Big Data in France, for instance in the fields of marketing, insurance, credit scoring, anti-fraud mechanisms, tourism or research. Data controllers can use specific compliance tools *i.e.* simplified standards or single authorizations that allow interconnecting databases (See AU39 fraud detection in insurance sector for a recent example <www.cnil.fr/documentation/deliberations/deliberation/delib/318/>). Regarding the law enforcement sector, different data processing operations can be considered as Big Data analysis. For example, opinions of the CNIL on such processing operations are available on our website (<www.cnil.fr/nc/linstitution/actualite/article/article/publication-de-lavis-sur-le-projet-de-loi-relatif-aurenseignement/;www.cnil.fr/documentation/deliberations/deliberation/delib/302/>).	At this stage, there is no report on the use of Big Data drafted by our DPA. However, different presentations were made during conferences on this topic as well as analytical articles (see, for example, the article '*Big Data et protection des données personnelles : quels enjeux ?*', Sophie Vulliet-Tavernier, Revue Statistique et société <www.statistique-et-societe.fr>). The CNIL also participated in the elaboration of International opinions (Statement of the WP29 on the impact of the development of Big Data on the protection of individuals with regard to the processing of the personal data in the EU; Working paper on Big Data and Privacy of the International Working Group on Data Protection in Telecommunications, Berlin Group). Besides, in 2011, the CNIL issued a warning against the company *Pages Jaunes* (*deliberation n° 2011-203, September 21, 2011*), for having obtained personal data contained in profiles available on different social media websites, without data subjects' knowing. This online directory proposed a 'webcrawl' function on its website enabling to add information from the accounts of web users to the search results provided by the directory. About 25 million people were concerned and the captured data included the names and first names, pseudonyms, photographs, the names of their school, the names of their employer, their geographical location... In particular, the CNIL considered that the fact that the data were public on the internet did not authorize a third party to massively, repetitively and indiscriminately collect

			such data without informing the data subjects before posting these information on its website. Consequently, the collection of the personal data was unfair. Moreover, it was difficult for the data subjects to exercise their rights. *Pages Jaunes* (Solocal Group) introduced an appeal before the *Conseil d' État* against the warning of the CNIL but the Supreme Court for administrative justice confirmed the analysis of the CNIL (*Conseil d'État, 10ème et 9ème sous-sections réunies, 12/03/2014, 353193*).
Hungary	The Hungarian National Authority for Data Protection and Freedom of Information accepts the Big Data definition of the International Working Group on Data Protection and Telecommunications. According to the Working Group's Working Paper on Big Data and Privacy: "Big Data is a term which refers to the enormous increase in access to and automated use of information. It refers to the gigantic amounts of digital data controlled by companies, authorities and other large organizations which are subjected to extensive analysis based on the use of algorithms." Big Data is, to a certain extent, used to analyze data in order to identify and predict trends and correlations.	As far as we know, there are no prominent examples in Hungary for the use of Big Data in law enforcement sector, by the police or intelligence services.	The Hungarian National Authority for Data Protection and Freedom of Information has not issued any decision, report or opinion on the use of Big Data so far. Besides that our Authority participated in the drafting of the working paper on Big Data by the International Working Group on Data Protection and Telecommunications. It is available online on the following address: <https://datenschutz-berlin.de/content/europa-international/international-working-group-on-data-protection-in-telecommunications-iwgdpt/working-papers-and-common-positions-adopted-by-the-working-group/>
Latvia	We do not have a specifically determined definition for Big Data, even though we are familiar with the debate on it.	No, there aren't.	No, we have not.
Lithuania	The State Data Protection Inspectorate is involved in discussions on Big Data, insofar as regards the performance of supervisory functions.	In Lithuania there is a Home Affairs Information System, which is a system performing data processing in which on the basis of the joint infrastructure of information technology and telecommunications operates the state and institutional registers and information systems (Criminal Offences register, Police information systems and etc.) managed by the MI and institutions under the MI.	Not yet.
Luxembourg	Big Data stems from the collection of large structured or unstructured datasets, the possible merger of such datasets as well as the analysis of these data through computer algorithms. It usually refers to datasets which cannot be stored, managed and analysed with average technical means, due to their size. Personal data can also be a part of Big Data but Big Data usually extends beyond that, containing aggregated and anonymous data. It allows	To our knowledge, there are no prominent examples of the use of Big Data in the law enforcement sector or by police or intelligence services in Luxembourg. There are however other actors which deal with Big Data. At a national level, a system of smart metering for electricity and gas has been launched. The project is, however, still in a testing phase. At the level of the University of Luxembourg, the Luxembourg Centre for Systems	The CNPD has not issued any decisions, reports or opinions that are directly dealing with Big Data. The Commission has however issued an opinion in a related matter, namely with regard to the problematic raised by smart metering. In 2013, the CNPD issued an opinion on smart metering (Avis de la Commission nationale pour la protection des données relatif au projet de règlement grand-ducal relatif aux modalités du comptage de

	for the correlation of information which previously could not be linked. From a data protection point of view it can raise many concerns, when it contains personal data, such as the respect of data subjects' rights – for example, in the context of data mining – and their ability to exercise control over the personal data or the respect fundamental principles of data protection such as that of data minimization or purpose limitation. Moreover, practices such as linking separate databases or computer analytics can turn anonymous data or any kind of non-identifiable information into personal data which would need to be protected under data protection law.	Biomedicine uses Big Data in the health sector. The Interdisciplinary Center for Security, Reliability and Trust (SnT) is also involved in Big Data projects. A partnership with Choice Technologies allows the SnT to conduct research into the new analytical methods in the domain of "Big Data". Moreover there are private companies that use Big Data. NeXus for example is, a company "which surfs the wave of Big Data and security by developing services that fall in the pure concept of "Industry 4.0". "With objects, people and data in constant move, nexus creates a dynamic identity for each end point and keeps track, connects and provides security to the information shared."200	l'énergie électrique et du gaz naturel, Délibération n° 566/2013 du 13 décembre 2013 (<www.cnpd.public.lu/fr/decisions-avis/2013/12/comptage-energie-gaz/566_2013_Deliberation_MinistereEconomie_avis-prj-rgd-comptage-energie-electrique-et-gaz-naturel.pdf>). The main argument of the opinion highlights the necessity to clearly define the purposes of the data processing as well as the retention periods of the data related to smart metering.
Netherlands	Yes, we are familiar with the broad concept of Big Data. Big Data is all about collecting as much information as possible; storing it in ever larger databases; combining data that is collected for different purposes; and applying algorithms to find correlations and unexpected new information. We refer to the speech of our chairman on Big Data, at URL: <https://cbpweb.nl/sites/default/files/atoms/files/2._sp ech_jko_panel_ii_privacy_with_no_territorial_bounds.pdf>	Yes, there are examples of the use of Big Data in the Netherlands. There has been a lot of media attention for Big Data use by the Tax administration (scraping websites such as Marktplaats to detect sales, mass collection of data about parking and driving in leased cars, including use of ANPR-data, and profiling people to detect potentially fraudulent tax filings, see for example the interview with the general manager of the IRS, at <https://decorrespondent.nl/2720/Baas-Belastingdienstover-Big-Data-Mijn-missie-is-gedragsverandering/83656320f6e78aaf>). Next to that, there are many pilots currently being conducted by different municipalities to combine different statistical, social care and medical care data, related to a shift in financial responsibility for social care duties. Recently, an interview was given by high ranking police officers describing the introduction of datamining tools for preventive policing. See URL: <www.politieacademie.nl/kennisenonderzoek/kennis/mediatheek/pdf/89539.pdf>	Next to the speech of our chairman, we refer to international opinions and resolutions from The International Working Group on Data Protection and Telecommunications (<www.datenschutz-berlin.de/attachments/1052/WP_Big_Data_final_clean_675.48.12.pdf?1407931243 The Article 29 Working Party (http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp221_en.pdf>) and The resolution from the International Commissioners conference (<https://cbpweb.nl/sites/default/files/atoms/files/resolution_big_data.pdf>). Our key concern is that data protection should be about surprise minimisation, while Big Data entails the risk of surprise maximization. There is a real risk that those who are involved in the development and use of Big Data are ignoring the basic principles of purpose limitation, data minimisation and transparency. And an additional frightening fact is that the statistical information, even if the data used is properly anonymised, can lead to such precise results that it essentially constitutes re-identification. When Big Data are used to profile people, it has the potential of leading us on to a - predetermined and maybe sometimes dangerous - path. A path that may in the end undermine the values that underpin our democratic societies, by depriving people of their free choice, of their right to personal development and equal treatment.

Norway	The Norwegian DPA issued a report on Big Data in 2013. The report was very well received and we have been giving talks on this topics for representatives from all sectors, covering finance, health, law enforcement, marketing, telecom etc. In the report we use the definition of Big Data as it was phrased by the the Article 29 Group:201 *Big Data is a term that refers to the enormous increase in access to and automated use of information: It refers to the gigantic amounts of digital data controlled by companies, authorities and other large organisations which are subjected to extensive analysis based on the use of algorithms. Big Data may be used to identify general trends and correlations, but it can also be used such that it affects individuals directly.* We use this definition as a basis, but also add what in our opinion is the key aspect of Big Data, namely that it is about the compilation of data from several different sources. In other words, it is not just the volume in itself that is of interest, but the fact that secondary value is derived from the data through reuse and analysis. This aspect of Big Data, and the consequences it has, is in our opinion the most challenging aspect from a privacy perspective.	There are, as far as we know, no usage of Big Data within the law enforcement sector in Norway. In 2014, the intelligence service addressed in a public speech the need to use Big Data techniques in order to combat terrorism more efficiently. However, politicians across all parties reacted very negatively to this request and no formal request to use such techniques has since been launched by the intelligence service. The companies that are most advanced when it comes to using Big Data may be found within the telecom (eg. Telenor) and media (eg. Schibsted and Cxence) sector. The tax and customs authorities have also initiated projects in which they look at how Big Data can be used to enhance the efficiency of their work.	The Norwegian DPA published a report on Big Data in 2013. In 2014 we drafted a working paper on Big Data for the International Working Group on Data Protection in Telecommunications (aka the Berlin Group). Following on from this work we were later responsible for drafting a Resolution on Big Data for the 36th International Conference of Data Protection Authorities and Privacy Commissioners. Report on Big Data: <www.datatilsynet.no/Global/04_planer_rapporter/big-dataengelsk-web.pdf> Working Paper on Big Data and Privacy: <www.datenschutz-berlin.de/content/europa-international/international working-group-on-data-protection-in-telecommunications-iwgdpt/working-papers-and-common-positions-adopted-bythe-working-group Resolution> on Big Data: <http://privacyconference2014.org/media/16602/Resolution-Big-Data.pdf> Our main argument in the report can be summarized as follows: "Big Data is challenging key privacy principles, in particular the principles of purpose limitation and data minimisation. The protection provided by these privacy principles is more important than ever at a time when an increasing amount of information is collected about us. The principles provide the foundation for safeguards against extensive profiling in an ever increasing array of new contexts. A watering down of key privacy principles, in combination with more extensive use of Big Data, is likely to have adverse consequences for the protection of privacy and other fundamental rights."
Slovakia	We are following the debate, but we have not adopted any definition yet.	We are not aware of special example of the use of Big Data in Slovakia.	No, we have not issued any documents about the use of Big Data yet.
Slovenia	The Information Commissioner is closely following the debate on Big Data. In terms of definitions of Big Data, we believe that established definitions and descriptions (e.g. Wikipedia) adequately describe the issue. Big Data is a broad term for processing of large amounts of different types of data, including personal data, acquired from multiple sources in various formats. Big Data revolves around predictive analytics – acquiring new knowledge from large data sets which requires new and more powerful processing applications.\ Big Data has important information privacy implications. Information on personal data processing may not be known to the	We have thus far not seen prominent examples of the use of Big Data in our country. To our knowledge, Big Data applications are particularly of interest in insurance, banking and electronic communications sector, mostly to battle fraud and other illegal practices. Another important field is scientific and statistical research. Law enforcement use is to our knowledge currently at development stages (e.g. in the case of processing Passenger Name Records), whereas information about the use of Big Data at intelligence services is either not available or of confidential nature.	So far, given that the use of Big Data in our country has not attained greater acceptance, we have not issued particular papers on Big Data at national level. On the other hand, we co-operate in international fora of privacy advocates and supervisory authorities, such as Article 29 Working Party202, International Working Group on Data Protection in Telecommunications203, European and International Privacy Commissioners conference204, which have already provided their views on the issues surrounding Big Data in resolutions, working papers and opinions.

	individual or poorly described for the individual, personal data may be used for purposes previously unknown to the individual. The individual may be profiled and decisions may be adopted in automated and non-transparent fashion having more or less severe consequences for the individual. Decisions about the individual may be biased, discriminatory and even adopted on grounds of statistics, averages and predictions that could have little or even nothing to do with individual's actual data. Such uses could have severe consequences for the individual particular when used by law enforcement, but also in other sensitive fields, such as health services and health insurance, social transfers, employment and in particularly situations where processing of sensitive personal data may be involved. The principles of personal data accuracy and personal data being kept up-to-date may also be under pressure in Big Data processing. Data may be processed by several entities and merged from different sources without proper transparency and legal ground. Processing vast quantities of personal data also brings along higher data security concerns and calls for strict and effective technical and organisational data security measures.		
Sweden	We are familiar with the debate on Big Data, but we have not produced any definition of this concept ourselves. As we see it, the concept is used for situations where large amounts of data are gathered in order to be made available for different purposes, not always precisely determined in advance.	We have not carried out any specific supervision related to the concept Big Data and do not have any statistics or specific information on how this is used. In our opinion, the law enforcement sector does not use Big Data. Their personal data processing is strictly regulated in terms of collection of data, limited purposes etc.	No
United Kingdom	We are familiar with current debates on Big Data and have contributed to them. We consider that the accepted Gartner definition based on the "three V's" (volume, variety and velocity) provides a useful starting point for defining Big Data. We also consider that other key characteristics of Big Data analytics include: repurposing data; using algorithms to find correlations in datasets rather than constructing traditional queries; and bringing together data from a variety of sources, including structured and unstructured data. Furthermore, we note that Big Data may involve not only data that has been consciously provided by data subjects, but also personal data that has been observed (eg from Internet	We have not carried out a comprehensive market assessment of Big Data but, from our contacts with business and our desk research, our impression is that the take up of Big Data is still at a relatively early stage in the UK. Nevertheless, we know that companies are actively investigating the potential of Big Data, and there are some examples of Big Data in practice, such as the use of telematics in motor insurance, the use of mobile phone location data for market research, and the availability of data from the Twitter 'firehose' for analytics. We do not have any specific information on the use of Big Data in law enforcement or security. The UK Data Protection Act includes a wide-ranging exemption from the data	In July 2014, we published a discussion paper on Big Data and data protection. We invited feedback on this and in April 2015, we published a summary of feedback, together with our response. In our work we have noted that Big Data poses a number of challenges to data protection, in particular: It may be difficult to provide meaningful privacy information to data subjects, because of the complexity of the analytics and people's reluctance to read terms and conditions, and because it may not be possible to identify at the outset all the purposes for which the data will be used. It may be difficult to obtain valid consent, particularly in circumstances where data is being collected through being observed or gathered from

	of Things devices), derived from other data or inferred through analytics and profiling. Given the range of features listed here, we think that it is difficult to produce a comprehensive definition of Big Data which fits all use cases. It is better to see Big Data as a phenomenon, rather than a specific technology. In our discussions with companies about Big Data, they have tended to see the defining characteristics of Big Data as the use of new data sources (eg social media data) and the use of existing data for new purposes, rather than simply the volume of data.	protection principles where it is required for safeguarding national security.	connected devices, rather than being consciously provided by data subjects. Big Data tends to use data for new and unexpected purposes, which may conflict with the purpose limitation principle. Big Data tends to use "all the data", which may conflict with the data minimization principle. Nevertheless, we have stressed that the data protection principles still apply in the world of Big Data; it is not a game that is played by different rules. We have said that organisations need to carry out a realistic assessment of what they are trying to achieve, and balance the benefits of the analytics to the organisation, to the individual and to society against the impact on data privacy. They also need to be innovative in seeking new ways to provide privacy notices. We think that privacy impact assessments (PIAs) have an important role to play in helping to ensure that Big Data analytics meets data protection requirements. We are currently doing further work with organisations to explore how PIAs can be used in the context of Big Data as part of privacy by design approach. We also advocate that, wherever possible and appropriate, the data used for the analytics should be anonymised, so that it can no longer be considered to be personal data. We are planning to publish a new version of our Big Data paper later this year.

	4. Are there any legal cases/judgements by a court with regard to (privacy/data protection) violations following from Big Data practices in your country? If so, could you provide us with a reference and the main consideration of the court?	**5. Which legal regimes are applied to Big Data/ is there a special regime for Big Data in your country? Are there any discussions/plans in parliament to introduce new legislation to regulate Big Data practices?**	**6. Are there any final remarks you want to make/suggestions you have for further research?**
Belgium	We have no judgment, yet, in the Facebook case. We expect that the main discussion will be on the competence of our DPA. See the media of 15 june 2015 (<www.theguardian.com/technology/2015/jun/15/belgium-facebook-court-privacy-breaches-ads>).	No. The general data protection law applies, and we expect that the new data protection regulation will be able to provide a partial answer (profiling) to Big Data issues (legal interpretation of the EU legal framework)	Most Belgian projects seem to be still in a pilot phase and the visibility of Big Data in practice is still low (competition issue). Often, the practice is still labeled differently (data mining, profiling,...) Conclusions seem to be premature at this stage until more experience has been obtained on the practical uses of this new practice. (Gartner's 2013 Hype Cycle for Emerging Technologies, <www.gartner.com/newsroom/id/281991>). Follow-up research seems necessary.

Croatia	At this moment, we do not have an appropriate/adequate information.	At the moment, in Republic of Croatia, there is no separate regulations governing the area of the Big Data, but certainly the part referring to the personal data of natural persons applies the Law on Protection of Personal Data.	No.
Estonia	Inspectorate is not aware of legal cases/ judgements by a court, related to Big Data practices in Estonia.	Estonian Data Protection Inspectorate consider Open Data as a part of Big Data. General requirements of Open Data processing are described in the Public Information Act, which new draft bill is in the parliament.	No additional comments.
France	Please refer to the aforementioned case.	Like the WP29, the CNIL considers that the EU and national legal framework for data protection is applicable to the processing of personal data in Big Data operations, even if the challenges of Big Data might require, in some cases, innovative thinking on how some of the key data protection principles are applied in practice. Regarding the discussions at the national level to introduce new legislation to regulate Big Data operations, we can mention the works relating to a new law for a 'Digital Republic' and a report published by the French Digital Council. At present, the French government is preparing a new law for a 'Digital Republic'. An online consultation was launched on the draft bill on September 2015, and the public was invited to suggest amendments to 30 proposed measures, ranging from net neutrality to open data (until 17 October 2015, < http://www.economie.gouv.fr/projet-loi-numerique>). The draft bill proposes, in particular, an open-data policy for the French state that would make official documents and public-sector research accessible to all online. The bill should be submitted to the parliament at the beginning of 2016. The French Digital Council (*Conseil national du numérique, CNNum*) is an independent advisory commission. The Council issues independent opinions and recommendations on any question relating to the impact of digital technologies on economy and society. The government can consult the Council on new legislation or draft regulations. The Council's thirty members come from across the digital spectrum, and include researchers and activists. In its report handed over on 13 June 2014 to Arnaud MONTEBOURG (Minister of Economy, of Productive Recovery and of the Digital) and to Axelle LEMAIRE, (Secretary	-

		of State charged of the Digital), the French Digital Council held an expanded approach to the neutrality principle: consecrate Internet neutrality and take into account the digital platforms that became the new entrance doors of the digital society. The report recommends to establish guidelines on transparency in the way services operate, particularly algorithms. The relevance criteria and governing principles of algorithms should be explained to users as part of a digital literacy effort. The report is available in English on the website of the French Digital Council (<www.cnnumerique.fr/wp-content/uploads/2014/06/PlatformNeutrality_VA.pdf>).	
Hungary	As far as we know, there hasn't been any legal cases or judgments by Hungarian court with regard to violation following from Big Data practices so far.	In Hungary Act CXII of 2011 on Information Self-Determination and Freedom of Information ("Privacy Act") should be applied to any data protection issues including data protection problems concerning Big Data. Neither the aforementioned act nor other laws includes special regulation on Big Data, so the general legal regulation on data protection and privacy should be applied. There aren't any plans or discussions now in the parliament to introduce special legislation for Big Data practices.	We would like to raise to attention that according to the working paper on Big Data by the International Working Group on Data Protection and Telecommunications the application of Privacy-by-Design principles are crucial for legitimate Big Data practices in most cases. Furthermore, a Privacy Impact Assessment could be also recommended and effective before the installation and use of Big Data services in order to avoid future privacy incidents. Furthermore, we would like to point out that in Hungarian business sphere more and more enterprises, such as banks, supermarkets, media and telecommunication companies use and take advantage of the possibilities in Big Data. Moreover, several international conferences are being organized in Budapest in the topic.
Latvia	We do not have such information.	We do not have information on this issue at this point.	No. But we would like to be informed on the outcome of this survey.
Lithuania	Not yet.	Not yet.	-
Luxembourg	No	There is no legislation directly addressing Big Data. The general data protection legislation applies (Amended Act of 2 August 2002 concerning the protection of individuals with regard to the processing of personal data). To our knowledge, there are no plans in Parliament to introduce new legislation to regulate Big Data practices.	-
Netherlands	Yes, there has been a court procedure in two instances about access to parking data for the IRS (case number HD 200.139.173/01, URL: <http://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:GHSHE:2014:2803>). Furthermore,	The current data protection regime also applies to the use of Big Data, but enforcement of the key values cannot be solely made dependent of the supervisory authority. Our chairman has called for a fierce social dialogue, to make people	-

	complaints about the use of police data from traffic cameras for the investigation of road vehicle usage in compliance with tax law have led to complaints and court cases. In March 2015, the Court of Appeal in Den Bosch ruled that the data that is collected with road surveillance camera's of the police that are installed for safety purposes, may be used by the tax authorities to monitor compliance with the law on road vehicle tax. (The ANPR data case, See: <http://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:GHSHE:2015:1087>)	aware of the risks to our intrinsic values that is posed by Big Data and to think together about how we can effectively address these risks and unwanted consequences. With regard to the security and intelligence services, a Bill has been consulted publicly and will be introduced to parliament soon to extend powers to allow for mass interception of communications data. With regard to scientific and academic research, sector- specific rules apply. For example, the law on higher education and scientific research.	
Norway	There are no legal cases	There are no special regimes for Big Data in Norway or plans to introduce new legislation. We rely on the national "Personal Data Act" which builds on the European Data Protection Directive.	Knowledge and awareness of the privacy challenges associated with Big Data are important among the enterprises that implement the technology. We urge the trade organisations to place these challenges on their agendas, and provide training in how they can be handled, for example through the use of privacy by design. Knowledge of data protection and the privacy challenges associated with the use of Big Data should be part of the curriculum for universities and colleges where data analysis or data science are taught. It is also crucial that supervisory authorities possess the necessary knowledge and awareness of the potential that lies in Big Data. This is important so that they can function as efficient and effective enforcers of the regulations that have been established to protect key societal assets. Research on the social and privacy consequences of Big Data is also of great importance. Big Data is still a relatively new phenomenon. It will be important to research how access to ever-increasing volumes and additional types of data will affect how we make decisions and organise our society in the future. At the Norwegian DPA we are currently looking into how it affects our privacy when personal data is more and more turning into a valuable commodity in all sectors of the economy. We are writing a report on how Big Data is used within the advertising industry, and how the use of automated, personalised marketing triggers an enormous appetite for and exchange of personal data.
Slovakia	We have no knowledge about the case or judgements about the Big Data in our country to this date.	We have no special regime for Big Data so far. General data protection law will apply when the personal data will be processed within the Big Data. We are not planning to issue a new legislation connected with	We think that the issue of Big Data is a very challenging topic. Finding the right balance between protection of personal data and the business models based on Big Data will need to be examined and

		Big Data practices yet.	legislated. As a research topic we would like to suggest examining boundaries between personal and non-personal information. In the Big Data environment, you are able to connect non-personal information and based on this information identify the data subject which represents potential risk to rights of the data subjects.
Slovenia	Not to our knowledge.	There is no special regime for Big Data. If processing of personal data is involved, then Personal Data Protection Act applies with its existing provisions. To our knowledge, there are no plans to introduce new legislation to regulate Big Data practices. The Information Commissioner has the competence to issue non-binding decisions regarding proposals for new legislation and will and would be able to comment on such proposals.	Big Data brings substantial challenges for personal data protection and these challenges must, firstly, be well understood and adequately addressed. In our view, new concepts and paradigms, such as cloud computing or Big Data should not lower or undermine the current levels of data protection as a fundamental human right. Existing central data protection principles, such lawfulness, fairness, proportionality, rights of the data subjects and finality should not be undermined with the advent of Big Data. The rights of the individuals to informational self – determination should be cornerstone in modern information society, protected by modern data protection framework delivering efficient data protection for the individual while allowing lawful and legitimate interests, often also in the interest of the individual, to be attained. Further research issues could cover the following topics: Understanding and managing privacy risks arising from the concept of Big Data. Adequacy and effectiveness of the notion of consent in the age of Big Data. Benefits and pitfalls of the notion of "legitimate interests" as legal ground for processing personal data in Big Data environments. The principle of finality vis a vis exploiting the benefits offered by Big Data. Privacy by design and privacy enhancing technologies in connection with Big Data. Accountability and other notions of demonstrative and effective data protection vis a vis Big Data. Automated decision making and profiling – which privacy safeguards are needed?
Sweden	No	Personal data processing in general is regulated in the Personal Data Act, which in principle applies to all sectors of society. However, many public agencies have their own personal data legislation which is specifically adapted to each agency's particular activity and needs. To the extent that public agencies collect large amounts of data, this is therefore usually specifically regulated (e.g. the Tax	-

		authority which processes data for taxation purposes but also for population register purposes). Telecom and Internet service providers' collection of data may involve collection of large amounts of data and this is specifically regulated in an act that implements the e-Privacy directive. This personal data processing does not fall under our supervision but, instead, under supervision of the National Post and Telecom Agency. It might also be worth noting that further to the aim to strengthen the right to privacy, the Swedish Constitution was amended in 2010 and now explicitly mentions the right to protection against privacy infringements by surveillance or mapping of the individual's personal circumstances without his/her consent. This means that the creation of large databases which contain information that provides a comprehensive image of an individual person, must be specifically permitted in an Act by the Parliament. We are not aware of any specific plans for Big Data regulation.	
United Kingdom	We are not aware of any cases specifically to do with Big Data. This may be due to the fact that Big Data analytics can be opaque to the data subject, and so people do not necessarily realise how their data is being used.	There is no specific legal regime for Big Data, other than the Data Protection Act. It is notable, however, that there is some evidence of a move towards self-regulation, in the sense that some companies are developing what can be described as an 'ethical' approach to Big Data, based on understanding the customer's perspective, being transparent about the processing and building trust.	We note that the proposals for the new EU General Data Protection regulation incorporate some of the measures we have identified as being important in ensuring compliance in Big Data, eg. clearer privacy notices, privacy impact assessments and privacy by design. We welcome the fact that these measures are being foregrounded, although we are concerned that that they should not be seen as simply a bureaucratic exercise.

* *LL.M. MPhil Bart van der Sloot* is a senior researcher at the Tilburg Institute for Law, Technology, and Society (TILT), Tilburg University, the Netherlands. During 2014 and 2015, he participated in the 'Big Data, Privacy and Security' project of the Netherlands Scientific Council for Government Policy.

Sascha van Schendel (LL.M.) is a Phd student at the Tilburg Institute for Law, Technology, and Society (TILT), Tilburg University, the Netherlands, she worked as a trainee in this project. The basis for this article was published as a Working Paper on the internet: <http://www.wrr.nl/fileadmin/en/publicaties/PDF-Working_Papers/WP_20_International_and_Comparative_Legal_Study_on_Big_Data.pdf>.