International Working Group
on Data Protection
in Telecommunications

675.48.12

Working Paper on Big Data and Privacy

Privacy principles under pressure in the age of Big Data analytics

55th Meeting, 5 – 6 May 2014, Skopje

Introduction[1]

1. Big Data is a term which refers to the enormous increase in access to and automated use of information.[2] 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[3].

2. Big Data entails a challenge to key privacy principles. Some claim that it will be impossible to enforce these principles in an age characterised by Big Data.[4] According to this view, the protection of privacy must primarily be safeguarded through enterprises providing clear and comprehensive information on how personal data is handled. The Working Group is of the opinion, however, that the protection of privacy is more important than ever at a time when increasing amounts of information are collected about individuals.[5] The privacy principles constitute our guarantee that we will not be

[1] This paper contains references to legal requirements that may not be relevant in all jurisdictions represented in the Working Group.

[2] Cf. White (2012). Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand databases management tools or traditional data processing applications.

[3] Article 29 Working Party, Opinion 03/2013 on purpose limitation,p.35.

[4] For example in Tene, Omer and Jules Polonetsky (2012), "Big Data for All: Privacy and User Control in the Age of Analytics", Northwestern Journal of Technology and Intellectual Property, Forthcoming, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2149364 , in World Economic Forum (2013), "Unlocking the Value of Personal Data: From Collection to Usage", http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf, and in Cate, Fred H. and Viktor Mayer-Schönberger (2013), "Tomorrow's privacy. Notice and consent in a world of Big Data", International Data Privacy Law, 2013, Vol. 3, No. 2

[5] Similar views are expressed by, among others, FTC Commissioner Julie Brill (2013): "We can unlock the potential of big data and enjoy its benefits. But we can do so and still obey privacy principles

Secretariat
Berliner Beauftragter für
Datenschutz und Informationsfreiheit
An der Urania 4- 10
D-10787 Berlin
Phone +49 / 30 / 13889 0
Fax: +49 / 30 / 215 5050

E-Mail:
IWGDPT@datenschutz-berlin.de

Internet:
http://www.berlin-privacy-group.org

subjected to 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, may have adverse consequences for the protection of privacy and other important values in society such as freedom of expression and the conditions for exchange of ideas.

3. Some core principles have been stipulated by the OECD and in the European Data Protection Directive for how personal data may be processed in a reasonable, correct and legitimate manner.[6] The following principles in particular are of relevance for Big Data: purpose limitation, relevance and data minimisation, completeness and quality, transparency and right of access to information.[7]

Scope

4. The purpose of this working paper is to highlight the privacy challenges associated with Big Data, primarily within the realm of the telecommunication industry, to help ensure these are placed on the agenda by data protection authorities and other stakeholders. The paper is intended for decision-makers, public authorities, industry and civil society.

5. Big Data covers a wide range of challenges. Many of the challenges, such as the risk of re-identification, could easily have been the subject of extensive reports in themselves. The purpose of this working paper, however, is to point out the key privacy challenges and not to address individual issues of a technical nature in detail.

Background

6. Data is everywhere. The amount of data on the global level is growing by 50 per cent annually. 90% of the world's data has been generated within the past two years alone.[8] Most of this data is generated by consumers through interaction with Internet-based services. With the emergence of the Internet of Things[9] , new data streams will be

---

that protect consumers", Reclaim Your Name: Privacy in the Age of Big Data, Sloan Cyber Security Lecture, Polytechnic Institute of NYU, October 23, 2013, and by Ann Cavoukian, Alexander Dix and Khaled El Emam (2014), "The Unintended Consequences of Privacy Paternalism", March 5, 2014, http://www.privacybydesign.ca/content/uploads/2014/03/pbd-privacy_paternalism.pdf.

[6] Cf. OECD Guidelines governing the Protection of Privacy and Transborder Flows of Personal Data (2013) and Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Corresponding principles are also laid down in the Recommendation CM/Rec(2010)13 of the Council of Europe on the protection of individuals with regard to automatic processing of personal data in the context of profiling.

[7] The corresponding principles in the OECD Guidelines are: the Collection Limitation Principle, Purpose Specification Principle, Data Quality Principle, Use Limitation Principle and the Individual Participation Principle.

[8] http://www-01.ibm.com/software/data/bigdata/

[9] The Internet of Things is the trend whereby an increasing number of objects and people are equipped with sensors in wireless communication with each other in networks

added. It is estimated that there will be more than 50 billion sensors by 2015.[10] These sensors will upload information to cloud computing services on how humans interact with the things surrounding us. This may bring about a change in markets and business models.

7. There is little doubt that the ability to store and analyse vast quantities of data will prove beneficial to society in many different ways.[11] Big Data is, to a certain extent, already used to analyse data in order to identify and predict trends and correlations. Big Data may be used, for example, to predict the spread of epidemics, uncover serious side effects of medicines and combat pollution in large cities. In principle, such analyses do not entail a challenge in terms of privacy, provided that the data have been properly anonymised (the concept of anonymization is discussed in more detail later in the paper). There are also Big Data analyses which do not involve use of personal data at all, such as analysis of weather data or sensor data from equipment on oil platforms.

8. Big Data may also be utilised in ways that may directly affect individuals. Techniques exist that can be used to prepare profiles and predict the behaviour of individuals and groups of individuals by compiling and analysing personal data collected from many different sources. Although the information may be aggregated and de-identified, the result of the analysis may still be of consequence for individuals.

9. "Personal data" means any information relating to an identified or identifiable individual.[12] IP-addresses, mobile phone numbers, RFID-tags and UDID-numbers are examples of unique identifiers that are considered to be personal data.[13] Data that reveal information about the habits and interests of uniquely identified individuals are sought after by companies and governments. The industry is therefore relentlessly developing new techniques aimed at this purpose, for instance device fingerprinting. As a result, the list of unique identifiers defined as personal data is constantly expanding.

10. The Big Data "value chain" involves several steps ranging from data collection, to storage and aggregation, analysis and the use of the analysis results (see value chain diagram at the end of the document). We will address the individual steps in turn.

---

[10] The Internet of Things. How the Next Evolution of the Internet Is Changing Everything, Cisco White Paper, 2011, http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf

[11] McKinsey Global Institute (2011),"Big Data: The next frontier for innovation, competition, and productivity" http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

[12] As defined by OECD (2013), "Guidelines governing the protection of privacy and transborder flows of personal data", http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf, and in the European Data Protection Directive, Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

[13] Article 29 Data Protection Working Party, Opinion 4/2007 on the concept of personal data

11. The first step of the value chain is *data collection*. Examples of potential sources of personal data include mobile phone apps, smart grids, toll tag transponders in vehicles, patient records, location data, social websites, air traffic passenger data, public registers, customer loyalty programs, genome sequencing, sales history, etc. Due to the proliferation of sensor technology, it will be possible to collect information from an extensive range of smart devices, such as smart toothbrushes, smart umbrellas, smart refrigerators, smart shoes, smart TVs etc. Data sources such as these may supply information which could potentially reveal a lot about the lifestyle of each individual.

12. For example, personal data may *be collected* in the following manner:

    i. Personal data may be submitted by individuals at their own initiative (e.g., by posting personal information on online social networks).
    ii. Personal data may be collected as a requirement of a service.
    iii. Personal data may be collected on the basis of legal requirements.
    iv. Personal data may be collected *automatically* in connection with the use of specific services (e.g., toll booth transaction data and location data). Such collection of data may also be carried out *unknowingly* (e.g., collection of airport Wi-Fi to track travellers[14]).
    v. Personal data may be *inferred* by the processing and analysis of data collected for previous and other purposes. Personal data may also be derived from various sets of information that appear to be anonymous.
    vi. Personal data (e.g., customer data (CRM)) may *be added* from external sources to enrich the (previously collected) data.
    vii. Personal data (e.g., (detailed) customer records) may *be shared* with external sources to enrich (personal) data from partner companies.

13. In a Big Data context, information collected about web users is very attractive since it may contain detailed information on the interests, networks, habits and behaviours of individuals. Such information may be collected explicitly (e.g., upon registration of a social profile on the web) or in a more covert manner, by use of various types of tracking technologies.[15]

14. The second step is to *aggregate and store* [16] the data after it has been collected. Some organizations aggregate and anonymise the data before it is stored, while others store

[14] http://www.cbc.ca/news/politics/csec-used-airport-wi-fi-to-track-canadian-travellers-edward-snowden-documents-1.2517881

[15] Within the EU, consent is now required for the use of certain cookies in order to make data collection more visible to the users and to ensure they are more in control.

[16] Aggregation in this context is to be understood as obtaining insight regarding a group of individuals, not individual persons. Aggregation entails display of data as sum totals. Data which may be

data containing personal identifiers. The phenomenal growth in storage and analytic power, at lower and lower cost, means that Big Data is no longer the province of a few giant companies. Big Data is now a tool available to both small and big enterprises, in all sectors of the economy. Big Data technology represents a switch from the traditional thinking concerning storage and processing of data using mainframe computers. New technologies make it possible to process and generate value from new and unstructured data sources.

15. The third step of the value chain involves *correlation and analysis* of the collected and stored data. A central element in the creation of value at this step is to merge data from various different sources to generate profiles and use analysis tools to derive information which would not otherwise be available. Big Data users may either compile only their own internal enterprise data, or they may buy data from third parties (or obtain data from open sources), and combine these with their own data. Examples of analysis techniques associated with Big Data include: Data Mining, Machine Learning, Social Network Analysis, Predictive Analytics, ”Sensemaking”, Natural Language Processing and Visualization.

16. The fourth step of the value chain involves the *use* of the results from the analysis. Big Data may be used in many different ways. An increasing number of players, for example banks, insurance companies, credit rating agencies, employers and the police, want to exploit knowledge obtained through analysis of Big Data to facilitate better, more informed decisions.

17. A wide range of stakeholders are involved throughout the Big Data value chain (see Figure 1 in the enclosure). Some stakeholders are involved only in selected parts of the value chain. For example, data brokers do not typically use the personal data but rather process it and sell it on. Other stakeholders can be involved in all the steps through the value chain. For example, a retailer can collect personal data through a customer-loyalty scheme, then store and aggregate the data, and finally process and use it in its own business model.[17]

18. Personal data has long been an attractive commodity and premise for development of new Internet-based services. Internet users typically gain access to free services by paying for them in the form of personal data. Due to Big Data and the proliferation of the Internet of Things, the market for sale of personal data will increase in volume and possibly expand to new sectors of the economy: Smart shoes with sensors may be offered free of charge in return for the user consenting to collection and analysis of the

---

linked to, or identify, individuals are not displayed. Low values are often hidden by rendering them ”unclear”, or by erasing them. One example of aggregation is use of average values.

[17] OECD (2013), “Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value”, OECD Digital Economy Papers, No. 220, OECD Publishing, http://dx.doi.org/10.1787/5k486qtxldmq-en

data from his or her running trips. A dentist may offer the individual a smart toothbrush (supplied for free by the manufacturer) in return for the user sharing the information collected by the toothbrush with various interested enterprises. New businesses and business models will emerge to extract the added value of the gigantic amounts of personal data generated in an ever expanding range of contexts.

Privacy implications

19. Based on the above review, the following key privacy challenges arise as a result of the use of Big Data.

Use of data for new purposes:

20. To a large extent, Big Data involves reuse of data. This entails a challenge to the privacy principle that collected data may not be used for purposes which are *incompatible* with the original purpose for collection.[18] The potential of Big Data for uncovering valuable knowledge through compilation of bigger and bigger data sets is putting the principle of purpose limitation under pressure. According to this principle, enterprises which use collected personal data as a basis for predictive analysis must ensure that the analysis is compatible with the original purpose for collecting the data. When individuals share data with others, they have a natural expectation about the purposes for which the data will be used. People do not hand over information to a company or the government to do whatever they wish with it. This may entail a considerable challenge for commercial use of Big Data.

Data maximisation:

21. Big Data is about data maximisation. In essence, Big Data is the very antithesis of the privacy principles of relevance and data minimisation.[19] These principles are intended to ensure that no more personal information is collected and stored than what is necessary to fulfil clearly defined purposes. The data must be deleted when it is no longer necessary for the original purpose. Big Data entails a new way of looking at data, where data is assigned value in itself. The value of the data is linked to its potential *future* uses. Such a view of data may challenge the privacy principle that stipulates that the processing of data must be adequate, relevant and not excessive for the purposes that have been defined and stated at the time of collection. It could also influence a data controller's desire and motivation for deleting data. Private enterprises and public bodies may not want to erase data which may at some point in the future prove to be a source of new insights and income. More widespread use of Big Data will make it even more challenging for the data protection authorities to enforce the obligation to erase data.

[18] Cf. article 6(1)b of Directive 95/46/EC.

[19] Article 6(1)c of Directive 95/46/EC.

Lack of transparency:

22. The right of access and the right to information regarding the processing of one's own personal data constitute important privacy principles. Lack of openness and information on how data is compiled and used may entail that we fall prey to decisions which we do not understand and have no control over. The average Internet user, for instance, has very little insight into how the online advertising market operates and how their personal data may be collected and utilised by a wide range of commercial parties.[20] Most people are not familiar with many of the players operating within this market, especially with the data brokers and analysis companies.[21] Thus, the right of the individual to request access to information becomes difficult to exercise.

Compilation of data may uncover sensitive information:

23. A challenging aspect associated with analysis of Big Data is the fact that compilation of collected bits and pieces of information, which may not be sensitive in themselves, may generate a sensitive result.[22] Through the use of Big Data tools, it is possible to identify patterns which may predict people's dispositions, for example related to health, political viewpoints or sexual orientation. This constitutes information subject to special protection. Data controllers must be aware of this risk when compiling and analysing data.[23]

Risk of re-identification:

24. One of the majorrisks associated with analysis of Big Data is that of re-identification. Through compilation of data from several sources, there is a risk that individuals may become identifiable from data sets which appear to be anonymous at first sight. This renders anonymisation less effective as a method to prevent privacy problems

---

[20] Turow, Joseph (2011), "The Daily You. How the New Advertising Industry Is Defining Your Identity and Your Worth, Yale University Press", New Haven & London

[21] Acxiom is one of the big data brokers. It is a US company which collects, analyses and interprets customer and business information for its clients, and it helps them with targeted advertisement campaigns, etc. The client base in the United States consists mostly of enterprises within finance, insurance, direct marketing, media, distributive trades, technology, health, telecommunications as well as the authorities. The company is one of the world's biggest processors of consumer information. It is said that it has 20 billion customer records and information on 96 per cent of the households in the United States.

[22] http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all

[23] An example frequently used to illustrate this challenge is the US chain Target's so-called "pregnancy algorithm". Target developed an algorithm which could predict which customers were pregnant based on which goods they bought. Target then sent coupons for "pregnancy products" to these customers. In one case, the distribution of such coupons entailed that the father in the household became aware of his daughter's pregnancy before she had an opportunity to tell him about it.

associated with profiling and other data analysis.[24,25] The risk of re-identification may be reduced by ensuring only anonymised data is used in the analysis. However, it is not always easy to determine whether a data set is sufficiently and robustly anonymised. This may prove difficult due to two reasons:

- The first of these is that the term ”to identify” - and thus "to anonymise" - is complicated because individuals may be identified in many different ways.[26] This includes direct identification, in which case a person will be explicitly identifiable from a single data source (for example a list with their full name), and indirect identification, in which case two or more data sources must be combined in order to allow identification.

- The second is that enterprises which use what is assumed to be an anonymised data set will not know for certain whether or not there are other data sets available which will make it possible for a third party to re-identify individuals in the anonymised data set. Even after the identifying information has been erased, it may still be possible to link specific information to individuals on the basis of links existing within different collections of Big Data. An actual example showing this is ”How to break anonymity of the Netflix Prize Dataset”.[27]

Security implications

25. Big Data also entails challenges in terms of information security which may also be of consequence for the protection of privacy. Examples of such security challenges include use of: several infrastructure layers in order to process Big Data, new types of infrastructure to handle the enormous flow of data as well as non-scalable encryption of large data sets. Further, a data breach may have more severe consequences when enormous datasets are stored. Enterprises that acquire and maintain large sets of personal data must be responsible stewards of that information.

Incorrect data:

---

[24] Anonymisation results from processing personal data in order to irreversibly prevent identification, ref Directive 95/46/EC. Anonymisation is also defined in international standards such as the ISO 29100 one – being the “Process by which personally identifiable information (PII) is irreversibly altered in such a way that a PII principal can no longer be identified directly or indirectly, either by the PII controller alone or in collaboration with any other party” (ISO 29100:2011).

[25] In the Working Paper on Web Tracking and Privacy: Respect for context, transparency and control remains essential (15./16. April 2013, Prague (Czech Republic)), the challenges associated with re-identification of data are described as a “game changer”.

[26] The Article 29 Data Protection Working Party, Opinion 05/2014 on “Anonymisation Techniques”

[27] http://arxiv.org/abs/cs/0610105v1

26. It is an important privacy principle that decisions of consequence for individuals must be based on correct information. The use of powerful data mining technology is becoming popular within the insurance and credit rating industries, for instance. Big Data facilitates the use of a far wider range, as well as new types, of data sources in the preparation of credit scores and risk profiles. New credit rating agencies specialising in the use of Big Data have popped up. These agencies prepare profiles for individuals based on information obtained from online sources only.

27. However, basing decisions on information obtained and compiled from social media, for example, does entail a risk that the decisions will be based on inaccurate information. Decisions based on such information will not be as transparent and verifiable as decisions based on information obtained from official registers. A weakness associated with Big Data analytics is that the context is often not taken into account.[28] Even when data is correct, there can be privacy issues related to its use out of context. Basing decisions on information intended for other purposes and generated in other contexts may yield results which do not reflect the actual situation. It is important to stress that the use of information intended for other purposes is, per se, unlawful in a data protection perspective unless the other purposes are compatible with the initial ones or the data have been anonymised.

28. Transparency, for example in the form of the right to familiarise oneself with the content of information processed about oneself, is a precondition for data subjects being able to safeguard their own interests. It is a key privacy principle that individuals may demand that information, assessments and allegations which prove not to be correct be corrected or erased.

Power imbalance:

29. Individuals, as a general rule, have limited power to influence how large corporations behave. Extensive use of Big Data analytics may increase the imbalance between large corporations on the one hand and the consumers on the other.[29] It is the companies that collect personal data that extract the ever-growing value inherent in the analysis and processing of such information, and not the individuals who submit the information. Rather, the transaction may be to the consumer's disadvantage in the sense that it can ex-

[28] danah boyd and Kate Crawford are two researchers who have emphasised the importance of taking into account the context in analysis of Big Data.
boyd, danah & Kate Crawford (2012), “Critical Questions for Big Data”, Information, Communication & Society 15:5, 662-679, http://dx.doi.org/10.1080/1369118X.2012.678878

[29] Article 29 Data Protection Working Party, Opinion 03/2013 on purpose limitation

pose them to potential future vulnerabilities (for example, with regard to employment opportunities, bank loans, or health insurance options).[30]

Data determinism and discrimination:

30. The "Big data-mindset" is based on the assumption that the more data you collect and have access to, the better, more reasoned and accurate decisions you will be able to make. But collection of more data may not necessarily entail more knowledge. More data may also result in more confusion and more false positives.[31] Extensive use of automated decisions and prediction analyses may have adverse consequences for individuals. Algorithms are not neutral, but reflect choices, among others, about data, connections, inferences, interpretations, and thresholds for inclusion that advances a specific purpose.[32] Big Data may hence consolidate existing prejudices and stereotyping, as well as reinforce social exclusion and stratification. Use of correlation analysis may also yield completely incorrect results for individuals. Correlation is often mistaken for causality. If the analyses show that individuals who like X have an eighty per cent probability rating of being exposed to Y, it is impossible to conclude that this will occur in 100 per cent of the cases. Thus, discrimination on the basis of statistical analysis may become a privacy issue. A development where more and more decisions in society are based on use of algorithms may result in a "Dictatorship of Data"[33], where we are no longer judged on the basis of our actual actions, but on the basis of what the data indicate will be our probable actions.

The Chilling effect:

31. If there is a development where credit scores and insurance premiums are based solely or primarily on the information we leave behind in various contexts on the Internet and in other arenas in our daily life, this may be of consequence for the protection of privacy

---

[30] The OECD has devoted attention to this issue and has released a report where they take a look at methodologies to estimate the monetary value of personal data. According to this report, methodologies to determine the value of personal data may help to provide greater transparency and insight into how the market for trading in personal data works. Further, the report argues that increased awareness among consumers about the value of their personal data may help to even out the economic imbalance between corporations on the one hand and the consumers on the other. It may also help consumers to place higher demands and expectations on how their personal data is being handled. OECD (2013), "Exploring the Economics of Personal Data: A survey of methodologies for measuring monetary value", OECD Digital Economy Papers, No. 220, OECD Publishing. http://dx.doi.org/10.1787/5k486qtxldmq-en

[31] Google Flu Trends has come under some scrutiny in recent days: http://bits.blogs.nytimes.com/2014/03/28/google-flu-trends-the-limits-of-big-data/

[32] Dwork, Cynthia and Deirdre K. Mulligan (2013), "It's not privacy, and it's not fair", 66 Stanford Law Review, Online 35, September 3, 2013, http://www.stanfordlawreview.org/online/privacy-and-big-data/its-not-privacy-and-its-not-fair

[33] Mayer-Schönberger, Viktor & Kenneth Cukier (2013), "Big Data. A Revolution That Will Transform How We Live, Work and Think", John Murray, London

and how we behave. In ten years, our children may not be able to obtain insurance coverage because we disclosed in a social network that we are predisposed for a genetic disorder, for example. This may result in us exercising restraint when we participate in society at large, or that we actively adapt our behaviour – both online and elsewhere. We may fear that the tracks we leave behind in various contexts may have an impact on future decisions, such as the possibility of finding work, obtaining loans, insurance, etc. It may even deter users from seeking out alternative points of view online for fear of being identified, profiled or discovered. With regard to the authorities' use of Big Data, uncertainty concerning which data sources are used for collecting information and how they are utilised may threaten our confidence in the authorities. This in turn may have a negative impact on the very foundation for an open and healthy democracy. Poor protection of our privacy may weaken democracy as citizens limit their participation in open exchanges of viewpoints. In a worst case scenario, extensive use of Big Data may have a chilling effect on freedom of expression if the premises for such use are not revealed and cannot be independently verified.[34]

Echo chambers:

32. Personalisation of the web, with customised media and news services based on the individual's web behaviour, will also have an impact on the framework conditions for public debates and exchanges of ideas – important premises for a healthy democracy. This is not primarily a privacy challenge, but constitutes a challenge for society at large. 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.

Recommendations

33. In spite of the fact that Big Data entails several privacy challenges, it is possible to make use of this type of analysis without infringing on key privacy principles. The Working Group makes the following recommendations regarding how Big Data may be used in ways that will respect the privacy of each individual.

---

[34] The Norwegian Data Protection Authority conducted a survey of Norwegians on privacy-related issues in 2013. One of the issues explored in the survey was whether there is a general tendency towards a chilling effect in Norway. In the survey, people were asked whether they have decided not to do something because they were not sure of how the information may be used in the future. The results from the survey indicate a tendency towards a general chilling effect. The results show that a significant share of the population has avoided certain activities because they have been uncertain regarding the potential future uses of the information. It is worth noting that as many as 26 per cent have decided not to sign a petition and that 16 per cent have decided not to do certain web searches. Norwegian Data Protection Authority (2014), “The Chilling Effect in Norway”, January, 2014.
http://www.datatilsynet.no/Global/04_planer_rapporter/Nedkj%C3%B8ling%20i%20norge_eng_.pdf

Consent:

34. It has been argued that consent as a legal basis for the processing of personal information will not function well in the age of Big Data.[35] Some claim that the constant demand for consent on the Internet paradoxically may result in poorer protection for the individuals. We can already see a development where enterprises ask for wide-ranging consent from their customers, perhaps speculating that consent statements are often not studied in detail, thus allowing them ”elbow room” to make use of the information for future and other purposes. Such use of consent is illegitimate.

35. Though there are undoubtedly challenges associated with obtaining meaningful consent, consent remains the cornerstone of modern privacy laws. Diminishing consent threatens to diminish an individual's control of the use that is being made of his/her data.[36] Consent is only one of several legal grounds to process personal data. It has an important role, but this does not exclude the possibility, depending on the context, of other legal grounds for processing perhaps being more appropriate from both the controller's and from the data subject's perspective.[37]

36. Valid consent should be obtained from the data subjects in connection with use of personal data for analysis and profiling purposes.[38]

37. In cases where it is not viable to request consent, processing of the data could be possible within carefully balanced limits[39]. For instance, the data controller may process the data if the processing is necessary for the purposes of the data controller's legitimate interests, as long as these interests are not overridden by the interests of the individual. The data controller must balance the two opposing interests – the legitimate interests and the interests of the individual – against each other. The result of the balancing of interests will differ from case to case, depending on which privacy related interests of the

---

[35] For instance by Cate, Fred H,and Viktor Mayer-Schönberger (2013), “Data Protection Principles for the 21st Century: Revising the 1980 OECD Guidelines”, December 2013, http://op.bna.com/pl.nsf/id/dapn-9gyjvw/$File/Data-Protection-Principles-for-the-21st-Century.pdf

[36] “Removing consent from the equation risks undermining fundamental individual rights, protections and freedoms far beyond the “notice and choice” systems. Instead of doing away with consent, we should work on improving transparency and individual control mechanisms — addressing the challenges head-on”
(Cavoukian et. al (2014) “The Unintended Consequences of Privacy Paternalism”)

[37] Article 29 Data Protection Working Party, Opinion 15/2011 on the definition of consent

[38] A valid consent shall be freely given, specific and informed, ref Article 2 (h) of Directive 95/46/EC.

[39] The Article 29 Working Party has provided guidance on how to carry out such balancing test (Cf. WP217, pp. 55-56). In addition, the US Whitehouse Big Data Report emphasizes the importance of context when considering how to address situations when consent may not be practicable (Executive Office of the President, White House (2014) ”Big Data: Seizing opportunities, preserving values”).

individual are at stake and the legitimate interests of the controller.[40] The more significant the impact on the data subjects, the more attention should be given to relevant safeguards.[41]

38. Data controllers wanting to use the collected data for a purpose different than the original one must assess the compatibility between the original and the new purposes on a case-by-case basis.[42] As long as the compatibility test is not satisfied, personally identifiable data may not be processed.

Procedures for robust anonymisation:

39. The data controller must decide whether the personal data to be utilised in the Big Data analysis is to be anonymised, pseudonymised or remain identifiable. This choice will determine how the legislation relating to data protection will affect the enterprise's further processing of the information. Anonymised data fall out of the scope of data protection legislation.

40. Anonymisation may help in alleviating or eliminating the privacy risks associated with big data analysis, but only if the anonymisation is engineered appropriately.[43]

---

[40] Although we cannot exclude that the weighing of interests can justify processing of personal data in some cases, this must always be considered on a case by case basis. This particular alternative can therefore hardly be used as a legal basis for the collection and analysis of Big Data in general. In any event, it is the controller that has the burden of proof that the conditions of the balancing of interests test are met. The provision, and its numerous discretionary components and uncertainties, may represent a challenge for the controller in this regard.

[41] The Article 29 Data Protection Working Party, in its Opinion 06/2014 on the "Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC", provide guidance on factors to be considered when carrying out a balancing of interest test. Special attention is devoted to the role that safeguards may play in reducing the undue impact on the data subjects, and thereby changing the balance of rights and interests to the extent that the data controller's legitimate interests will not be overridden. Safeguards may include, among others, strict limitations on how much data are collected, immediate deletion of data after use, technical and organisational measures to ensure functional separation, appropriate use of anonymisation techniques, aggregation of data, and privacy-enhancing technologies but also increased transparency, accountability, and the possibility to opt-out of the processing.

[42] The Article 29 Data Protection Working Party, in its Opinion 15/2011 on purpose limitation, suggests an assessment of all relevant circumstances and, in particular, of the following key factors:
- the relationship between the purposes for which the personal data have been collected and the purposes of further processing;
- the context in which the personal data have been collected and the reasonable expectations of the data subjects as to their further use;
- the nature of the personal data and the impact of the further processing on the data subjects;
- the safeguards adopted by the controller to ensure fair processing and to prevent any undue impact on the data subjects.

[43] The Article 29 Data Protection Working Party, in its Opinion 05/2014 on "Anonymisation Techniques" stresses that anonymisation techniques can provide privacy guarantees, but only if their application is engineered appropriately – which means that the prerequisites (context) and the objec-

Anonymisation results from the processing of personal data in order to prevent identification irreversibly. In doing so, several elements should be taken into account by data controllers, having regard to all the means "likely reasonably" to be used for identification (either by the controller or by any third party).[44] It is important to test anonymised data in terms of acceptable risk level. This should be documented, for example as part of a Privacy Impact Assessment.

41. The optimal solution for anonymising the data should be decided on a case-by-case basis, possibly using a combination of techniques. Several anonymisation techniques may be envisaged, mainly consisting of data randomization and generalisation.[45] Knowing the main strengths and weaknesses of each technique may help in determining how to design an adequate anonymisation process. The robustness of each technique should be based on three criteria:[46]

    i. is it still possible to single out an individual
    ii. is it still possible to link records relating to an individual, and
    iii. can information be inferred concerning an individual?

42. Pseudonymised data is not equivalent to anonymised data. Data controllers who choose to pseudonymise the information, rather than to anonymise it, must be aware that the information will still be defined as personal data and thus must be protected.

43. Great care must be exercised before sharing or publishing pseudonymised, or otherwise identifiable data sets. If the data is detailed, may be linked to other data sets,[47] and contains personal data, access should be limited and carefully controlled. If the data has been aggregated and there is less risk of linking it to other data sets, it is more likely that the data may be made accessible without any significant risks.

---

tive(s) of the anonymisation process must be clearly set out in order to achieve the targeted anonymisation level.

[44] The Article 29 Data Protection Working Party, in its Opinion 05/2014 on "Anonymisation Techniques" emphasis that it is not possible nor useful to provide an exhaustive enumeration of circumstances when identification is no longer possible. However, the document provides some general guidance on the approach to assessing the identifiability potential of a given dataset that undergoes anonymisation according to the different available techniques.

[45] Randomization and generalisation are two families of anonymisation techniques covering for instance noise addition, permutation, differential privacy, aggregation, k-anonymity, l-diversity and t-closeness.

[46] The Article 29 Data Protection Working Party, Opinion 05/2014 on "Anonymisation Techniques" provides an overview of the strengths and weaknesses of the techniques considered in terms of the three basic criteria.

[47] It may be possible to link pseudonymised information in a data set with information in another data set, for example by using the same unique ID for each individual.

44. If a data controller makes pseudonymised or otherwise identifiable data available to other organisations, it should contractually prohibit such entities from attempting to re-identify the data.[48] This should also include open data.[49]

45. The Working Group recommends that a network or body be established where anyone who needs to anonymise or pseudonymise data may discuss challenges associated with anonymisation as well as exchange lessons learned. There is such a network in UK (the UK Anonymisation Network (UKAN)) which is coordinated by the universities in Manchester and Southampton, the Open Data Institute, and the Office for National Statistics.[50]

Greater transparency and control from collection to use of data:

46. Each individual should be informed of which data is collected, how the data is handled, for which purposes it will be used and whether or not the data will be distributed to third parties.[51]

47. Each individual should have access to their profile and all the information the data controller holds about them. Each individual should also be informed of the sources of the various personal data. They should further, subject to applicable law[52], be able to correct their information and to opt-out of collection schemes used for (behavioural) profiling purposes, or opt-in.[53]

48. Classification systems may have adverse consequences for the individual. Each individual should therefore have access to information on which algorithms have been used as a basis for profiling or decision-making. The information should be presented in a clear and understandable format. This is important to prevent unfair discrimination and

---

[48] This recommendation is also put forward by the FTC in the report "Protecting Consumer Privacy in an Era of Rapid Change", FTC Report, Federal Trade Commission, March 2012

[49] Article 29 Data Protection Working Party, Opinion 6/2013 on Open Data, p. 14

[50] http://www.ukanon.net/

[51] For instance, some companies offer their customers so-called personal data dashboards, providing the customers with an overview of how their personal data is being processed.

[52] Note that different regulations may apply to the public and the private sector.

[53] FTC`s Commissioner Julie Brill has voiced similar recommendations in her "Reclaim Your Name" initiative. This initiative is aimed at empowering the consumer so they can find out how data brokers are collecting and using data. Reclaim Your Name will give the consumer access to information that data brokers have amassed about them, allow consumers to opt-out if they learn a data broker is selling the information for marketing purposes and provide consumers with the opportunity to correct errors in information used for substantive decisions. (Reclaim Your Name, 23rd Computers Freedom and Privacy Conference, Keynote Address by Commissioner Julie Brill, FTC, Washington, DC, June 26, 2013)

avoid decisions of significance for individuals being based on an incorrect factual basis.[54]

49. Upon request each individual should receive all data about themselves in the possession of the controller in a user-friendly, portable and machine-readable format where appropriate. This will make it easier to switch to the provider with the best terms and conditions, including in terms of protection of privacy. Data portability will prevent customers from being locked into services with unacceptable terms and conditions. Over time, such a requirement may result in the development of more privacy-friendly services. It may also help data subjects to improve their understanding about the data relating to them.

Privacy by Design and Accountability:

50. More robust anonymisation techniques will not, by themselves, solve the challenges Big Data presents to privacy. There is a need for additional solutions. Privacy by Design and accountability are also important to help alleviate the privacy challenges.

51. Use of Big Data technologies should be based on the seven principles of Privacy by Design.[55] Privacy by Design entails taking into account protection of privacy at all stages of system development, in procedures and in business practices.

52. In order to retain the confidence of those whose personal data is collected, processed and analysed, it is important to assess the challenges in terms of protection of privacy as early as possible, and in any case prior to the processing of Big Data. This may be done in the form of a Privacy Impact Assessment (PIA). A PIA should include an evaluation of any legal basis for distribution and reuse of personal data, evaluate the principles of purpose limitation, proportionality and data minimisation, as well as evaluate privacy and security safeguards. Such an assessment should also carefully evaluate any potential consequences for the data subjects.[56,57]

---

[54] Providing transparency into the algorithms used for profiling is important. However, given the complexity of algorithms, it is unreasonable to expect transparency alone to remedy potentially inherent biases. Automated decision-making systems should also be subjected to ethical and accountable oversight, as highlighted in paragraph 53.

[55] The seven principles for Privacy by Design are: 1. Proactive not Reactive; Preventative not Remedial, 2. Privacy as the Default Setting, 3. Privacy Embedded into Design, 4. Full Functionality — Positive-Sum, not Zero-Sum, 5. End-to-End Security — Full Lifecycle Protection, 6. Visibility and Transparency — Keep it Open, 7. Respect for User Privacy — Keep it User-Centric, http://www.ipc.on.ca/images/resources/7foundationalprinciples.pdf

[56] Article 29 Data Protection Working Party, Opinion 6/2013 on Open Data and Article 29 Data Protection Working Party, Opinion 06/2014 on the "Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC"

[57] The EU has established a PIA framework for RFID applications to help identify the consequences of use of RFIDs in terms of privacy. This framework is also interesting for enterprises utilising Big Data in light of the emergence of the Internet of Things. The framework has been established by the

53. Accountability is an important privacy principle. Accountability builds trust between data subjects and data controllers. Data controllers need to demonstrate that they are being accountable and can make responsible and ethical decisions around their use of big data. For instance, data controllers should be aware that an anonymised dataset may still have an impact on individuals. Anonymised datasets may be used to enrich existing profiles of individuals, thus creating new data protection issues. Both profiles and the underlying algorithms require continuous assessment. This necessitates regular controls to ensure that decisions resulting from the profiling are responsible, fair, ethical and compatible with the purpose for which the profiles are being used. Injustice for individuals due to fully automated false positive or false negative results should be avoided.[58]

Enhancement of knowledge and awareness:

54. Knowledge and awareness of privacy challenges linked to Big Data is important among data controllers utilising this technology. The industry must put the challenges on the agenda and provide training in how they may be resolved, for example through use of Privacy by Design.

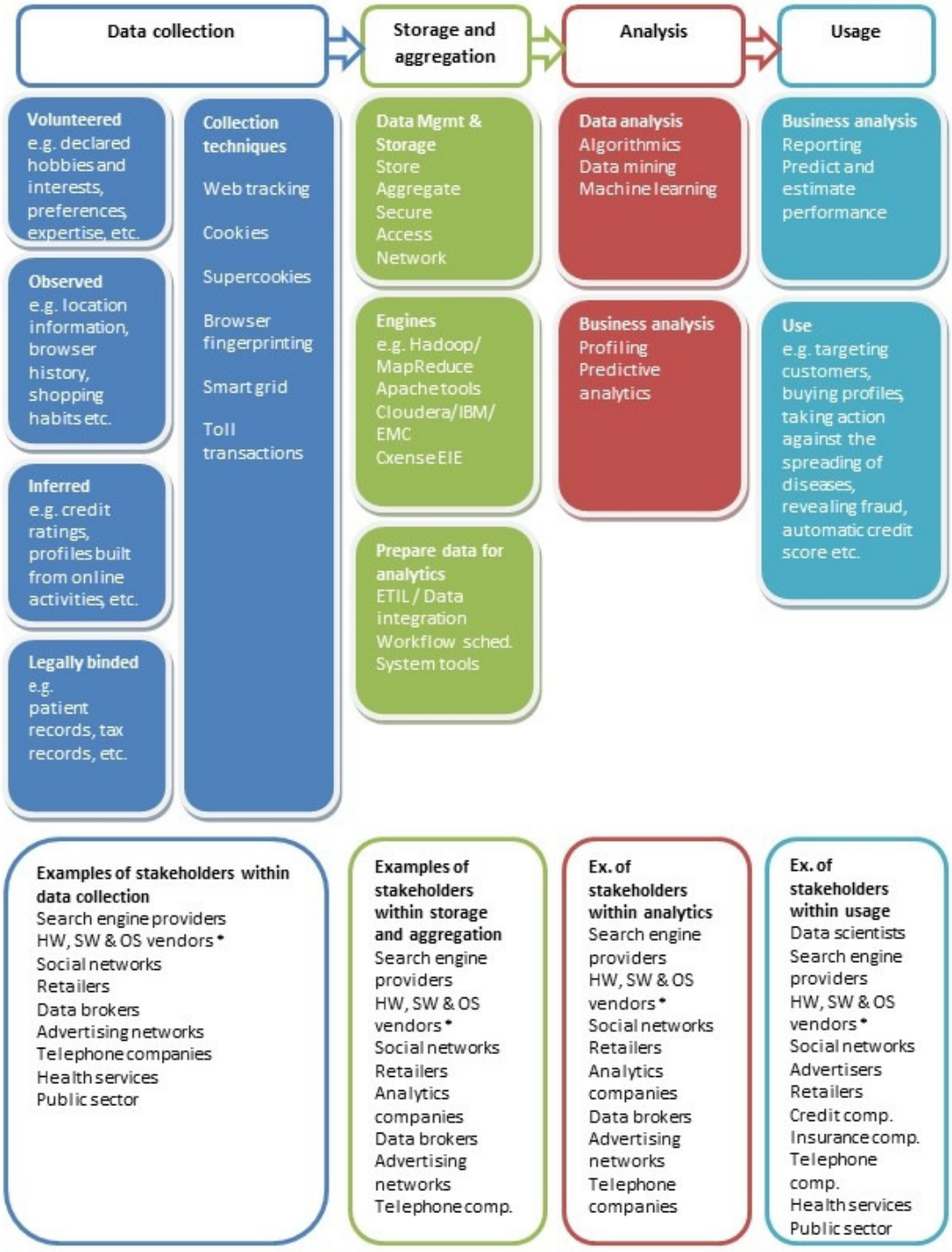
55. The subjects of protection of privacy and privacy challenges linked to the use of Big Data should be taught at universities and colleges teaching data or information science.

56. Public authorities should have the necessary knowledge and awareness regarding the potential of Big Data. This is especially important in connection with stipulation of new acts and regulations. Awareness of the challenges is important in order for public authority to be able to safeguard their function as protectors of some of society's key values.

RFID industry and has been recognised by the data protection authorities in the EU as being in compliance with the legislation for protection of privacy. The Article 29 Working Party encourages the establishment of a similar framework for use of Big Data technology by Big Data professionals. (The European Commission (2011), "Privacy and Data Protection Impact Assessment Framework for RFID Applications", 12 January 2011, http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp180_annex_en.pdf)

[58] Uruguay Declaration on profiling (2012), 34th International Conference of Data Protection and Privacy Commissioners, 25. – 26. October 2012, http://privacyconference2012.org/wps/wcm/connect/7b10b0804d5dc38db944fbfd6066fd91/Uruguay_Declaration_final.pdf?MOD=AJPERES

The Big Data Value Chain



* Hardware, Software and Operating System vendors