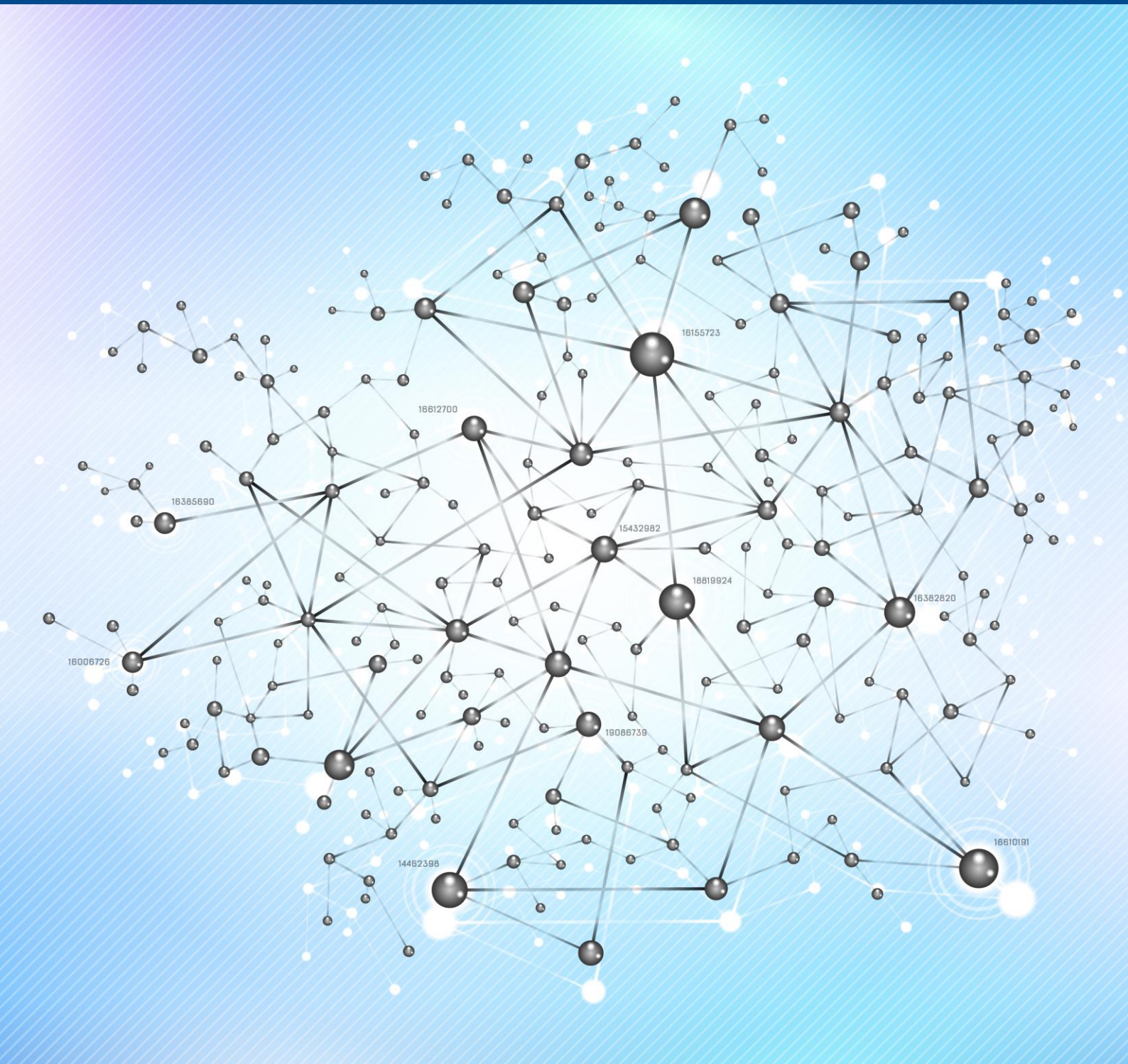
Datatilsynet

Big Data

– privacy principles under pressure

September 2013



Contents

Summary

Big Data is a concept 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.

The use of Big Data challenges key privacy principles. Some individuals therefore claim that the current privacy legislation must be adapted to a new reality. The Data Protection Authority does not share this opinion. At a time when ever-increasing amounts of information are collected about us, it is more important than ever to safeguard fundamental principles of privacy. The principles constitute our guarantee that we will not be subject to extensive profiling in an ever-increasing array of new contexts.

Big Data can be used for many good and socially beneficial purposes. Analytic techniques are used to analyse anonymised data in order to identify and predict trends and correlations. The use of anonymised data does not in principle challenge privacy protection. However, Big Data can also be used such that it affects individuals directly. In this report, we highlight ten key privacy challenges related to Big Data:

1. *Use of data for new purposes:* Big Data largely involves the reuse of data. This entails a challenge to the privacy principle of purpose limitation; i.e. that collected data may not be used for purposes that are incompatible with the original purpose for collection. According to this principle, enterprises that 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. This may entail a considerable challenge for commercial Big Data analysis.
2. *Data maximisation:* Big Data entails a new way of looking at data, where data is assigned value in itself. The value of the data lies in its potential *future* uses. Such a view of data may influence the enterprises' desire and motivation to delete data. Neither private nor public enterprises will want to delete data that at some point in the future, and when compiled with other data sets, may prove to be a source of new insight or income.
3. *Lack of transparency:* Lack of openness and information on how data are compiled and used may entail that we fall prey to decisions that we do not understand and have no control over. The average Internet user, for instance, has very little insight into how personal data is collected and utilised by commercial interests. In addition, many of the actors that operate in this market are unknown to most people.
4. *Compilation of data may uncover sensitive information:* A challenging aspect associated with Big Data analysis is the fact that collected information that is not sensitive in itself may generate a sensitive result through compilation. It is important that enterprises that use Big Data are familiar with this problem and do not develop algorithms that may reveal vital privacy interests.
5. *Risk of re-identification:* One of the really major challenges associated with Big Data analysis is that of re-identification. Through compilation of data from several sources, there is a risk that individuals may become identifiable from data sets that are supposedly anonymous. This renders anonymisation less effective as a method to prevent the privacy problems associated with profiling and other data analysis.

6. *Imbalance between enterprises and individuals* Big Data increases the imbalance between large enterprises on the one hand and individuals on the other hand. The enterprises that *collect* personal data are extracting ever-increasing added value from the analysis and processing of this information, not those of us who *provide* the information. It is more likely that this transaction will be a disadvantage to us in the sense that it may expose us to future vulnerability.
7. *Incorrect data:* It is an important principle of privacy that decisions of consequence to individuals must be based on correct information. One weakness of Big Data analysis is that it often does not take the context into account. Basing decisions on information that is intended for other purposes may yield results that do not correspond to the actual situation.
8. *Data determinism:* Extensive use of automated decisions and prediction analysis may consolidate existing prejudices and reinforce social exclusion and stratification. A development where more and more decisions in society are based on the use of algorithms may result in a "Dictatorship of Data", where we are no longer judged on the basis of our actual actions, but on the basis of what all the data about us indicate our probable actions may be.
9. *Chilling effect:* If all of the tracks we leave behind on the Internet and elsewhere are used for a growing number of new purposes that are unknown to us, this may cause us to show restraint in how we participate in society. Ambiguity and uncertainty connected to the authorities' use of Big Data may threaten our trust in the authorities. If the premises for its use are kept concealed and cannot be verified, extensive use of Big Data may at worst have a chilling effect on freedom of expression.
10. *"Echo chambers":* With increased personalisation of the web, individuals will be increasingly less exposed to opinions deviating from their own. This may have an impact on the framework conditions for public debate and the exchange of ideas. This is not primarily a privacy challenge, but constitutes a challenge for society at large.

Even though Big Data raises a number of privacy challenges, it is possible to make use of this type of analysis and respect the privacy of individuals at the same time. In this report, we will be making the following recommendations, among others:

- As a rule, the processing of personal data should be based on *consent*. If it is not possible or desirable to use consent, the information should be *anonymised*.
- Good routines for the anonymisation and de-identification of information are of major importance. This will contribute to reducing the risk of re-identification.
- Big Data should be used in accordance with the principles of data protection by design.
- Enterprises that use Big Data must be open about how they process the personal data they collect. This means giving the individual access to the *decision-making criteria* (algorithms) used as the basis for developing profiles, and to the *sources from which* the information is retrieved, among other things.
- Individuals should be given an opportunity to have all the data about them that the enterprise possesses disclosed to them in a user-friendly format. Data portability will prevent customers from becoming locked into services that have unacceptable terms and conditions.

1 Introduction

Data is everywhere. Most of these data are generated by us consumers in the form of videos uploaded to YouTube, Twitter messages, training applications, emails, location data from mobile phones, Facebook updates, music streaming, web searches, purchases of books on Amazon, etc. With the emergence of the Internet of Things[1], new data streams will be added. Countless sensors will upload information to the cloud about how we humans interact with the things around us. It is estimated that by 2020, there will be more than 50 billion sensors that can communicate with each other (IBM 2013).

A growing number of commercial enterprises and authorities are discovering that these enormous data streams can be strategically exploited. This is called Big Data, and it is predicted to have a revolutionary effect on society. The purpose of Big Data is to make use of the gigantic volume of data to look for patterns and correlations that were not possible to detect previously. Such knowledge is valuable not just for marketing and sales, but also for the authorities, with a view to combating disease and crime, for example.

Big data can be used for many good purposes, but the phenomenon also contains privacy challenges. Big Data entails a new way of looking at data, where data are assigned value in itself. The value of the data lies in its potential *future* uses. Such a view of data challenges key privacy principles related to purpose limitation and data minimisation. Another key challenge of Big Data is that this type of analysis entails a risk of re-identification, which makes anonymisation less effective as a method for preventing the privacy disadvantages associated with profiling and other data analysis.

The use of Big Data is currently at the starting gate. How we handle the development towards increasingly more extensive use of the enormous data streams that are generated is critical to privacy. Strong forces – both commercial actors and public authorities — embrace Big Data. The exploitation of the data streams – referred to as the new oil – will be important in order to promote competitiveness and innovation in society. Yet this must be done in a manner that does not threaten the privacy of the individual.

1.1 Problems for discussion

The report's overarching problems for discussion are: What potential privacy challenges does the use of Big Data entail, and what does the legal scope look like?

Big Data is a relatively imprecise term that involves a number of different problem areas, activities and actors. A goal of the report is therefore to obtain a better overview of the actors and the specific activities that represent the core of Big Data.

In order to obtain a deeper understanding of the use of Big Data, we have looked at how this type of analytic technique is used in various sectors. We have selected four sectors: internet-based companies, credit rating and insurance industries, the health sector and the justice sector. The sectors have been chosen to illustrate the breadth of the use of Big Data in both the private and

[1] The Internet of Things refers to objects that are equipped with devices that can communicate wirelessly with each other in networks.

public sectors. The selection of sectors is also based on the fact that these are areas in which Big Data is already used extensively, or has been predicted to be used extensively in the future.

Scope

As mentioned, Big Data affects many different issues of importance to privacy. Several of the problems could have individually been the subject of extensive reports, such as challenges related to tracking technology, profiling and the risk of re-identification of anonymous data.

The scope of this report has not been to treat individual problems in detail. Our goal has been to survey the challenges that Big Data entails for privacy at an overall level, and the limitations that current privacy legislation places on large-scale data analysis.

Structure of the report

In Section 2, we discuss what we call the value chain for Big Data. We show what processes, technology and actors are associated with Big Data. In Section 3, we look at how Big Data is used in different sectors and what potential privacy challenges this use entails. In Section 4, we discuss the legal scope for the use of Big Data. In conclusion, we will make recommendations and propose measures we believe to be important in order to ensure that the use of Big Data respects the privacy of individuals.

1.2 Definitions

In the following, we define key terms in the report, including Big Data, personal data, anonymous data and re-identification.

1.2.1 Big Data

There is no single definition that has been agreed on for the term Big Data. Big Data is used to refer to many things, and the meaning of the term is vague. The term refers to both the data in itself and to the activities related to collecting, storing and analysing the data.

The European Commission's advisory body on data protection, the Article 29 Group, defines Big Data as follows:[2]

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 will 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 key and challenging aspect from a privacy perspective.

[2] Opinion 03/2013 on purpose limitation

1.2.2 Personal data concept

The concept of personal data is the cornerstone for an analysis of the privacy challenges associated with the use of Big Data. In order for information to be defined as personal data, it must be possible to link the information directly or indirectly to an individual person.[3] Big Data largely involves the analysis of anonymous or anonymised data. Pursuant to the Personal Data Act, such data are not regarded as personal data. Big Data analysis that encompasses the use of anonymous or anonymised information will therefore normally fall outside the scope of the Personal Data Act.

1.2.3 Anonymous data and re-identification

The Article 29 Group defines anonymous data as data that cannot possibly be linked back to a natural person by means of any reasonable technical aids. Anonymised data is defined as anonymous data that was *previously* linked to a natural person, but where identification is no longer possible.[4]

One of the major challenges of Big Data analysis from a privacy perspective is the risk of re-identification. Re-identification means that data that initially emerges as anonymous is made identifiable again by means of various techniques. Through the compilation of multiple data sets, as is the case in Big Data analysis, a risk may arise that individuals can be identified from data that is initially anonymised.

1.3 Key privacy principles

The European Data Protection Directive stipulates certain core principles for how personal data shall be processed in a reasonable, legal and legitimate manner. [5] The following principles are of particular relevance in relation to Big Data:

Purpose specification. Personal data shall only be collected for specific purposes, and these purposes must be expressly stated and legitimate. In addition, the purposes for the subsequent processing of the information must not be incompatible with the purposes for which the information was originally collected. Personal data must not be disclosed to others without consent or other legal grounds for disclosure.

Relevance and minimisation. Personal data shall only be collected, stored and processed to the extent necessary in order to fulfil the purpose. Collected data that are no longer necessary for the stated purpose must be deleted or anonymised.

Completeness and quality. Personal data must be relevant, correct and complete according to the purposes for which the data are to be used. Information stored in a register is often used as the basis for making decisions about the data subjects. This principle ensures that decisions are not made on an incomplete or incorrect basis.

Information and access. Data subjects are entitled to be informed of the collection and use of their personal data. They also have the right of access to the information that is registered about them. In

---

[3] The personal data concept is discussed in more detail in Section 0.
[4] Opinion 04/2007 on the concept of personal data
[5] 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.

addition, they are entitled to a manual assessment of decisions that are fully based on automated processing of personal data, provided that the decision made is of significant importance to the data subject.

1.4 Big Data at the starting line in Europe and Norway

A report from the McKinsey Global Institute (2011) shows how the use of Big Data will have a transformative effect on entire sectors, ranging from marketing to the health sector and political campaign activities. It is stressed that Big Data will increase competitiveness and the degree of innovation in business. In Europe, public authorities will be able to realise more than EUR 100 billion in efficiency gains by using Big Data, McKinsey maintains. This number does not include the use of Big Data to combat tax and National Insurance fraud.

Previously only very large enterprises were able to engage in advanced data analysis, since this required access to an enormous data storage capacity and computing power. The opportunity to store data in the cloud, as well as access to inexpensive analysis software, makes it now practically possible for both small and large companies to make use of Big Data.

At present, the use of Big Data is not particularly widespread in Europe compared with the USA (NESSI White Paper 2012). Knowledge of Big Data is present to some extent among larger European enterprises, but there is still a low level of knowledge among small and medium-sized enterprises. Only a limited number of enterprises specialising in the delivery of services and products related to Big Data have been established in Europe (NESSI White Paper 2012). Most such companies have been established in the USA.

The impediments that are pointed out to explain why Europe is a Big Data latecomer include, for example, a lack of expertise (not enough students are educated in the field of data science), a fragmented market, and the absence of an adequate number of large market actors who can drive the development into the future. In addition, the current European data protection legislation is pointed out as a significant impediment as to why Europe has not made as much progress as the USA in the field of Big Data (NESSI White Paper 2012).

The EU's Digital Agenda programme points out the importance of making provisions for the use of Big Data (Whelan 2013). [6] Investment in Big Data is linked to the EU's strategy for open data and focus on cloud services. Modernisation of the data protection regulations is regarded as decisive for the EU to have an opportunity to realise gains from the use of Big Data in the future.

The Data Protection Authority has been in contact with various actors in the Norwegian market in connection with this report. Based on these conversations, our impression is that the use of Big Data has not come very far in Norway either. A few actors have implemented this technology, among them enterprises in the fields of telecommunication, media and retail trade. Certain analysis firms have also emerged that specialise in Big Data and offer their expertise and technology to Norwegian

[6] The goal of the Digital Agenda for Europe (DAE) is to help European citizens and companies realise the greatest possible gains from digital technology.

and foreign companies.[7] There are strong Big Data research communities at the Universities of Oslo, Trondheim and Tromsø.

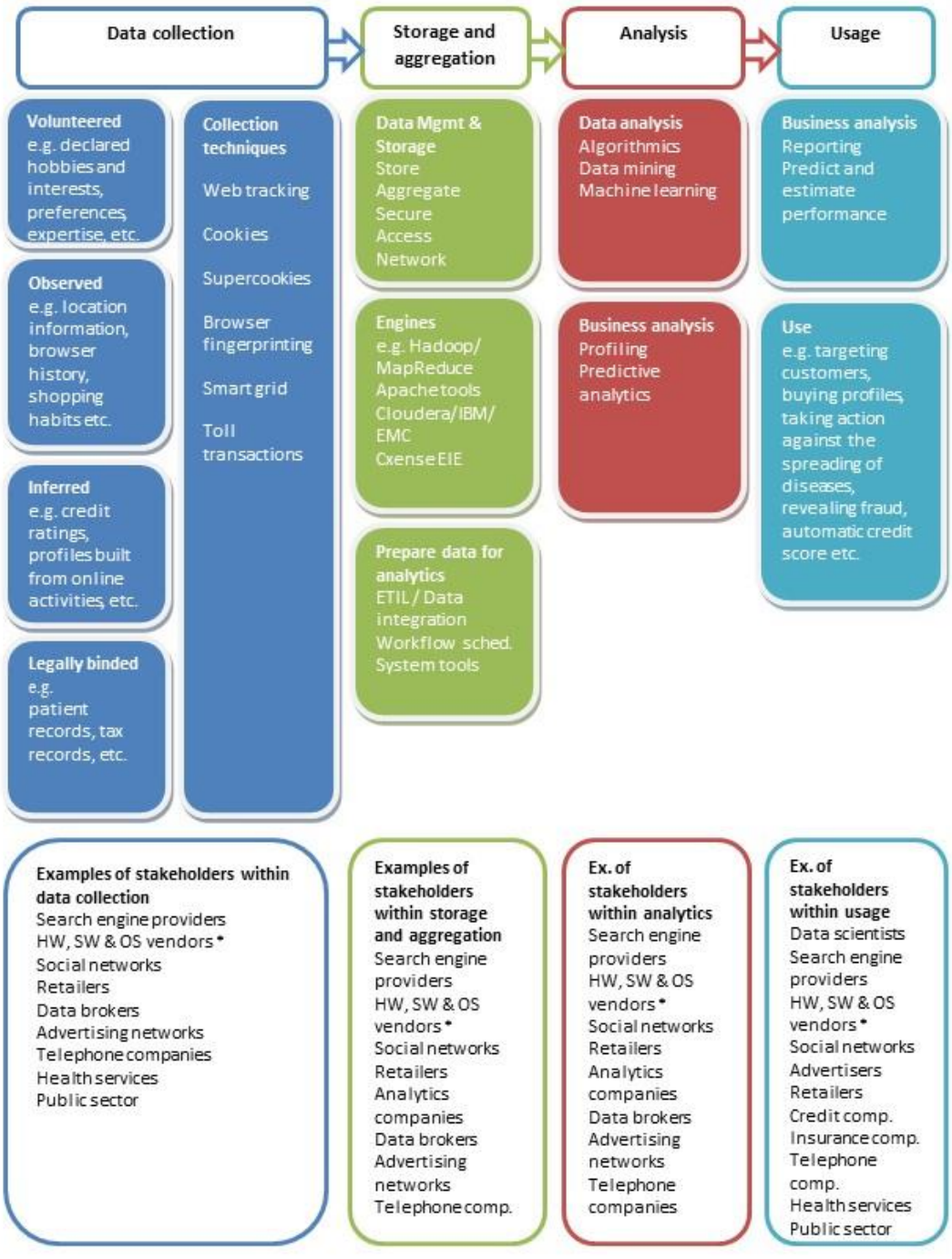
[7] An example is cXense, which supplies expertise and technology to large Norwegian and foreign media companies: http://www.cxense.com/

2 The Big Data Value Chain: actors, processes and technology

In the following, we will review the value chain for Big Data, that is to say the process from the collection of data to storage, aggregation and analysis.

The purpose of this presentation is to provide an overview of the different stages of the analysis process, and of how the various actors are involved at different points in the value chain.

Figure 1, value chain:



* Hardware, Software and Operating System vendors

2.1 Collection of data

The first stage in the value chain for Big Data is the collection of information that is to form the basis for further analysis. One of the characteristics of Big Data is the fact that a multitude of different data sources are used, including both structured and unstructured data[8]. The data sources may contain personal data, or may be sources that do not contain such information – such as weather data and information generated from sensors and various types of production equipment in a factory.

It is the data sources that contain personal data that are of interest from our perspective, and there are many such sources; we leave electronic traces in most of our activities from the time we get up in the morning until we go to bed at night. The day starts by checking Facebook. Toll stations record when we drive to and from work. Customer loyalty cards in stores and credit cards register our purchases. Access control cards at work register when we start and end our day. Mobile phones in our pockets and the use of location-based applications register our movements throughout the entire day. All of these may be relevant and attractive sources in a Big Data context.

The development towards the Internet of Things will contribute to generating new streams of data. The Internet of Things means that a growing number of objects and persons are equipped with devices that can communicate wirelessly with each other in networks. The devices may be sensors that collect data, or RFID chips that are used to identify objects, animals or persons. Such units can be used to monitor and control objects, persons or processes, and we can control these remotely through apps or websites. For example, you can check whether the door is locked at home, and lock it if it is not. Your training shoes can communicate with your smart phone so that all the details of your training activities are available to you, and your alarm clock can talk with the coffee machine and the light switch, and adapt to your needs. However, not everything is controlled over the Internet yet. Many devices can only be controlled within a local area network. The technology is available today, and IPv6[9]will enable every single thing to have its own unique resource identifier (URI) and thus be available over the Internet. The Internet of Things is a disruptive[10] technology, and, when it takes off, Big Data will be really huge.

The privacy challenges associated with the Internet of Things arise when fragments of data that are related to different objects or services are collected and aggregated. A collection of many fragments of data can suddenly become personal data when events are assessed in the context of time, place

---

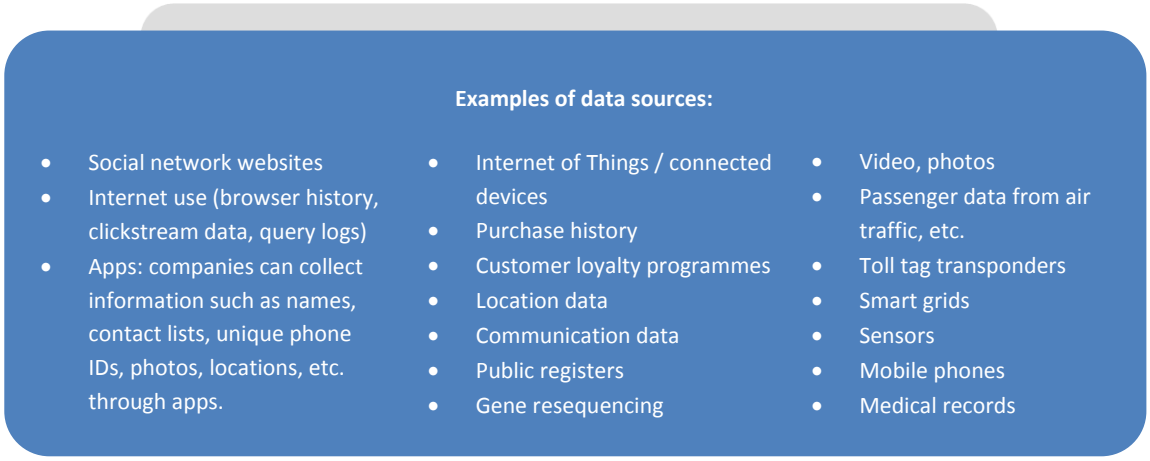
[8] *Structured data* can be described as data that follows a formal structure of data models, such as rows and columns in a relational database, as for example in a customer register. *Unstructured data* refers to data that lacks a recognisable structure. For example, photos, video, email, Word documents and pure text are regarded as unstructured data in a data set or file. *Semi-structured data* is a form of structured data that does not follow the formal structure of data models. Some parts have a fixed format and other parts consist of free text. This is often referred to as a schemaless or self-describing structure. Semi-structured data are often represented by XML files, books, websites, email and EDI – often data collected from sensors and machines.

[9] Internet protocol version 6 (IPv6) is a set of protocols that computers use to exchange information over the Internet and over home and company networks. IPv6 allows the allocation of many more IP addresses than with IPv4.

[10] Disruptive innovation is innovation that disrupts an existing market. This concept is used in business and technical literature to describe innovations that improve a product or service in a way that the market does not expect, normally by lowering the price or designing for a different set of consumers.

and recurrence. The use of sensor technology in grocery stores can, for example, give an indication of religion, or provide grounds for assumptions concerning a person's lifestyle or health by revealing routine purchases of certain food products. Sensor data retrieved from refrigerators, for example, can say something about the owners' daily routines; when they are at home, at what time of day they eat, how often, etc. With new technology and new things that are connected to the Internet, new vulnerabilities that allow applications and systems to be attacked will also arise. TVs can be hacked, and apps that control home security alarms can be hacked or leak information to a third party (ars technica 2012).

Figure 2, data sources:



Personal data can be retrieved in various ways.

- It can be on a voluntary basis, by individuals explicitly making personal data on themselves available. For example, this can occur by creating a profile on a social networking website, providing information to become a member of a chain store loyalty program, downloading an application on a mobile phone or registering personal information to gain access to a service.
- Personal data can be automatically registered by enterprises in connection with the use of a specific service – as opposed to information that is provided on request. Examples of such data include location data, browser history, purchasing habits, history of training centre visits and data on passing toll stations.
- Personal data can be *derived* from the processing and analysis of data collected for prior and other purposes. Personal data can also be derived from different sets of apparently anonymous information.
- The collection of personal data for local and central government authorities may be prescribed by a special act or regulations. This applies to personal data in medical records, in tax lists and in the motor vehicle register, for example.

A great deal of the information that is used in a Big Data context is generated online. This information can be collected explicitly (when a social profile is registered on the Internet), or in a more concealed manner, such as is the case when tracking technology is used. The use of cookies

now requires consent so that the collection is more visible to the users and gives them more control. Here is a description of various techniques that are used to collect personal data using the Internet:

- *Web tracking* can be defined as the collection, analysis and use of data consisting of user activity from a computer or other device while using various online services. The intention of those who collect data is to combine and analyse the information for various purposes. The greatest challenge associated with web tracking is when tracking is performed by a third party. One example of this is when, as a registered user of a website, you click on a banner ad on this website and your email address is forwarded to a dozen other companies (Mayer 2011).
- *Cookies* (information capsules) are small text files that are placed on a user's computer when a website is downloaded. A cookie can contain information that the user has registered on the page, such as his user name and password in an encrypted form, which is transferred to the website on subsequent visits. Accordingly, it is not necessary to register the information more than once.
- Supercookies are a type of cookie that is permanently stored on a user's computer. Supercookies are generally more difficult for the users to find and remove from their devices, since they cannot be removed in the same way as ordinary cookies.
- *Browser fingerprinting* can be used as a tracking technology for people who restrict the use of cookies. This method involves collecting data on the type of browser you have, plug-ins that are installed, system fonts, type of browser, operating system, time zone, screen resolution and colour depth, and whether cookies are blocked. One method of using browser fingerprinting is to combine the method with an IP address in order to find out what devices are hidden behind the individual IP address. Browser fingerprinting is a powerful technique, and such fingerprinting can be equated with cookies, IP addresses and super cookies in regard to web tracking (Eckersley 2010).

2.2 Storage and aggregation

After data have been collected, they can be stored and aggregated. Individual data elements are organised and stored in data sets that can be used for subsequent processing and analysis. Some entities aggregate and anonymise data before they are stored, while others store data with personal information. In this context, aggregation means that the data are merged to form larger volumes, so that data cannot be related to or identify anyone.[11] The challenge associated with this is the fact that it is possible that the aggregated data set may subsequently be linked with other data sets such that natural persons can be re-identified.

In order for Big Data actors to be competitive, the technology they use must be able to handle a large volume of data and process various types of data (structured, unstructured and semi-structured) from various sources, and index the data and make it searchable within a very short period of time. The key prerequisite for Big Data is the growth of cloud services that can offer practically unlimited

[11] An example of aggregation is rounding off numbers. If an entry in the table that has a value of 10,000, which is the number of people who perform a specific activity in the month, changes to 10,001 the next month, an unauthorised party could compare the tables and find the one occurrence that was different. Rounding off this number could prevent this.

storage capacity at an ever-lower price. Enterprises can collect and process far larger volumes of data than previously. Limited storage capacity no longer represents an obstacle.

As a result of the requirements for managing the volume, velocity (speed) and variation of data[12], many new tools, infrastructure and frameworks have been developed. Big Data technology breaks with the traditional line of thought regarding storing and processing data using mainframes. New technology makes it possible to process and extract value from new and unstructured data sources. Challenges have accordingly arisen with regard to information security – challenges that may have consequences for data protection. Security challenges arise for example when multiple infrastructure layers are used to process Big Data, a new type of infrastructure to handle the enormous throughput of data, and in connection with the use of non-scalable encryption of large data sets.

However, some actors believe that the storage of data in connection with Big Data is a thing of the past, and that the analysis of real-time data is the future. A company that is going against the idea of large-scale data storage is Numenta and its founder Jeff Hawkins. Hawkins believes that the only reason one should look at historical data is if one believes that the world will not change in the future (*The New York Times*, 2012a). Numenta has a product called Grok, and it automatically analyses data streams in real time. This is a cloud-based service that regularly retrieves data streams from thermostats, clicks on the Internet, machines, etc. To begin with, Grok observes the flow of data, and it gradually starts to guess what will happen. The more data, the more accurate the predictions will be (more about "sensemaking" in the next section) (*The New York Times* 2012a).

2.3 Analysis

In the third stage of the value chain, the collected and stored data are *combined with other information*. A key part of the creation of value in this stage is the aggregation of data from various sources in order to create profiles, and the use of analysis tools to derive information that would not otherwise be available. The enterprise can either choose to compile only its own internal enterprise data, or buy data from other actors (possibly collect data from open sources) and then compile this data with its own data.

In the following, we will present an overview of some of the analysis techniques that are used in Big Data:

- *Data mining* involves looking for structure and meaning in large volumes of unstructured data.
- *Machine learning* is a type of artificial intelligence in the field of informatics. It is a scientific discipline that is concerned with the design and development of algorithms, which enable computers to develop behaviour based on empirical data. Algorithms are used to recognise and predict correlations and patterns in the data. An algorithm (in mathematics and informatics) is a precise description of a finite series of operations that is to be performed in order to solve a problem or a set of multiple problems. In other words, algorithms are used to tell a program what is to be done and how it should be carried out. The machine learning

[12] A frequently used definition of Big Data links the concept to the three Vs: volume, variation and velocity (speed): http://www.forbes.com/sites/oreillymedia/2012/01/19/volume-velocity-variety-what-you-need-to-know-about-big-data/

process is quite similar to the data mining process. Both systems search through data to look for patterns. Whereas data mining unpacks data to achieve human comprehension, machine learning uses data to improve the program's own comprehension. Machine learning recognises patterns in data and adjusts the program's actions accordingly. Facebook's news feed is an example of this. The news feed changes based on the user's interactions with other users. If a user often tags a friend in photos, writes on his wall or likes his links, the news feed will show more of this friend's activity in the user's news feed. This is based on the assumed closeness of the relationship and attraction, as well as time and volume. The algorithm that is used is called EdgeRank[13].

- *Social network analysis* looks at social relationships on the basis of network theory. The networks consist of nodes (which represent individual actors in the network) and ties (which represent relationships between individuals, such as friendship, family, position in organisations). These networks are often illustrated in a "social network diagram", in which the nodes are represented as points and ties are represented as lines.
- *Prediction analysis* involves estimating future probabilities and trends. The key element in prediction analysis is "the predictor" – a variable that can be measured for a natural person or other entity in order to predict future behaviour. For example, it is probable that an insurance company will take into account potential predictors for safe driving such as age, gender and driving history when they issue a car insurance policy. A "predictive model" consists of several connected predictors, and is used in analysis to predict future probabilities with an acceptable degree of reliability.
- *Sensemaking* involves integrating new transactions (observations) with earlier transactions, in the same manner that one takes a piece in a puzzle and finds the subsequent pieces on the table. As opposed to other analytic methods, which require that the users question the system, the new systems can operate by means of another principle: data finds data, and relevance finds the user.

Other analysis techniques include A/B Testing, Classification, Cluster Analysis (clustering), Natural Language Processing (NLP), (Social) Neural Networks, Optimisation, Spatial Analysis, Simulation, Time Series Analysis and Visualisation.

2.4 Actors

Many different actors are involved in the value chain for Big Data. We can roughly distinguish between data owners, data brokers, Big Data companies, and Big Data analysis and consulting firms. Certain enterprises are involved in all of the stages, from the collection of data to the use of the finished analysis results (see figure 1, value chain). Other actors are only involved at certain points in the value chain.

Data owners

This is the actual enterprise that makes use of Big Data. This may include small and large enterprises, both public as well as private. It is often data that the enterprise itself possesses that is subject to analysis, even though the use of other data sources is also desired to a greater extent to gain new

[13] A simple explanation of how EdgeRank works: http://blog.hubspot.com/understanding-facebook-edgerank-algorithm-infographic

insight. The enterprise has collected the data as part of performing the enterprise's various activities. It may, for example, be a chain store that collects data through customer loyalty cards, then stores and aggregates these data and analyses them to improve its own business model.

Data brokers

Data brokers collect data with a view to analysing and reselling the information. This is a market that is growing in step with the use of Big Data, since a growing number of enterprises in a growing number of sectors are interested in buying data that they can compile with their internal enterprise data. Acxiom is one of the really large data brokers. It is an American company that collects, analyses and interprets customer and business information for clients, and helps them with targeted advertising campaigns etc. Their client base in the USA consists primarily of enterprises in the fields of finance, insurance, direct marketing, media, retail trade, technology, health and telecommunications, as well as public authorities. The company is one of the world's largest processors of consumer information. They are said to have information on 96% of the households in the USA (*The New York Times* 2012b).

Another company that few people have heard about, but which has grown large without attracting attention, is Flurry. It is a free analysis service that delivers user statistics to the app owner, and in return Flurry gets access to information on the users. Flurry accordingly possesses enormous volumes of information on mobile phone users collected through various applications. This is information that Flurry aggregates and resells as profiles to marketers and others who are looking to reach certain target groups.

In Norway there are no pure data brokerage companies. Nor are there many such companies in Europe, to our knowledge..

Big Data companies

These are companies that build new business models and create new services based on the available data generated on the Internet or open public data. An example of such a company is the credit rating company Kreditech, which will be discussed in Section 3.2.

(Big) Data specialists

Some companies help other companies extract value from the data they possess. Large enterprises, such as Google and Facebook, have the resources to analyse the data they collect in-house. This would be difficult for smaller companies, and they buy the expertise from specialised (Big Data) analysis firms. These are companies with expert knowledge and special software for Big Data analysis. Of the software companies with analysis and visualisation tools, we can mention SAS (Visual Analytics), QlikView, Tableau, Tibco Spotfire and Panopticon. Datasift is an example of an actor that sells real-time analysis from social networks. Palantir supplies analysis software to police and intelligence authorities in the USA, for example. A Norwegian actor involved in analysis and Big Data is cXense.

3 Privacy challenges related to the use of Big Data

The primary purpose of using Big Data is generally the same in all sectors: gaining new insight into correlations that can be used to predict actions or events, among other things. What data is collected, how they are processed and whether the analysis results are intended for use in relation to individuals or at a more general level, varies. In the following, we will look at various application areas for Big Data, and at the possible privacy challenges that arise from the use of Big Data.

3.1 Big Data in use among internet-based companies

Internet-based companies are pioneers in the exploitation of Big Data technology (Bollier 2010). All of the major Internet companies – Google, Facebook, Amazon, eBay, Microsoft, Apple and Yahoo! – use Big Data in one form or another to extract secondary value from the gigantic volumes of data they possess. Google is a good example of this. They do not only use the data they collect for targeted marketing. The data are also used to improve search algorithms and to develop new data-intensive services. An illustrative example is Google's recently launched service *Google Now*, which is referred to as a typical Big Data application. The purpose of this mobile application is to give people help before they themselves realise that they need it. For example, this can be accomplished by notifying them that their bus is late before they leave work for the day. Google Now's algorithm uses data from the users email, calendar and search history to learn about people's habits (*MIT Technology Review* 2013a).

Facebook, the world's largest social network community, possesses enormous amounts of personal data. Not only on their 900 million members, but also on non-members if they visit websites or apps where Facebook's "like buttons" are installed. Through this function, Facebook can track network activity outside the website's own pages (Datatilsynet 2011a). Facebook has its own *Data Science Team*. Their job is to analyse these enormous volumes of data to uncover patterns and trends in people's interactions and activities. This is valuable knowledge for the development of new services and products, and it is something that Facebook can earn money on. It is also knowledge that is very valuable for resale to many other actors that desire to reach special target groups. Facebook maintains that it does not sell information on its members to third parties.

The Internet's fundamental business model has become free services that earn money on personal data generated by Internet users. Personal data is collected either by the users providing the data themselves or by tracking the users' network activities by means of various tracking tools. The tracking tools make use of unique identifiers that can compile individual user behaviour across many different network services over time. The information that is collected is used to build up user profiles. This makes it possible to customise advertising, offers and services to specific customers. These activities are referred to as behavioural advertising. The Article 29 Group has written a statement on the privacy challenges associated with this form of online marketing.[14]

The use of Big Data in behavioural marketing does not represent anything fundamentally new. The technology makes it possible, however, to process and compile an *even greater volume* and to collect

[14] Opinion 02/2010 on online social networking

data from a *broader range of data sources* than previously. The use of Big Data has therefore been called *"data mining on steroids"* (Rubinstein 2012).

With the propagation of the Internet of Things, the market for the sale of personal data will grow. We may see a development in which the Internet's fundamental business model is transferred to other markets. Smart training shoes with sensors can be offered for free in return for the users consenting to data on their jogs being collected and analysed for various purposes. Smart toothbrushes can be given away for free in return for the users sharing information collected by the toothbrush with research institutions, insurance companies, grocery chains, etc. New enterprises and business models will grow in order to extract the added value represented by the gigantic volumes of personal data that are generated in a growing number of contexts. The PRISM case has shown that actors other than commercial actors are also interested in exploiting these data. This case will be discussed in greater detail in Section 3.4.2.

3.2 Big Data in insurance and credit rating

In the insurance and credit rating industries, the use of correlation analysis and profiling is nothing new. Correlation analysis is used to evaluate risk profiles and creditworthiness. Norwegian insurance and credit rating companies, however, cannot extract and use personal data any way they want. The companies have a license from the data protection authority to process personal data under specific conditions. The license specifies what data sources can be used. When insurance companies are to assess the risk associated with a customer, for example, they are banned from using certain types of data, such as gender, ethnicity, religion or payment remarks in their pricing.

The use of new and powerful data mining technology is predicted to be huge in the insurance and credit rating industries internationally. This is a trend that will probably also affect Norwegian actors. Even small improvements in the accuracy of the analyses may result in substantial gains (Dagens IT 2013). Big Data enables a far broader range of data sources to be included in the preparation of credit scores and risk profiles. New credit rating companies that specialise in the use of Big Data have cropped up over time, such as the German Kreditech[15], which maintains that it is the leading company in its field in Europe.

The company says the following about the data sources they use in their credit rating analyses:

> *"Kreditech works like a Big mosaic - Any online data that can be found about an individual will be used for fraud detection, identification, scoring: Location data (GPS, micro-geographical), social graph (likes, friends, locations, posts), behavioural analytics (movement and duration on the webpage), people's e-commerce shopping behaviour and device data (apps installed, operating systems) are just some examples of up to 8,000 data points that are processed in real-time for any single scoring unit."*

Figure 3:

[15] http://www.kreditech.com/#where-we-are

(Source: www.kreditech.com)

Not only will more data sources be used, new data sources may also replace those that are used today. Medical records, for example, have proven to be very costly to analyse. Experts therefore believe that other data sources will be used by the insurance industry to determine the health conditions and risk profiles of people, because this will be less expensive and more efficient (*The Economist* 2012). Information collected from social media may provide information on people's level of activity, and thus the assumed condition of their health. Is the person in question social and have many interests, or does he stay inside a lot? The compilation of large data sets may also uncover correlations concerning people's risk profile and state of health, which may be of interest to the companies. The magazine *The Economist* (2012) refers to an example in which the use of Big Data analysis has identified that people who make frequent withdrawals from ATMs live longer than those who use credit cards and cheques. Over time it has also become possible for insurance and credit rating companies to buy information on people's consumption patterns from data brokers and other companies who possess large databases of such information.

There are also credit rating companies in Norway that say they will invest in Big Data (*Dagens IT* 2013). However, current licensing places restrictions on how Norwegian companies can make use of the technology. The use of information retrieved from social media and the collection of consumer data from external companies, as outlined above, for example, will not be permitted.

3.3 Big Data in the health field

The McKinsey Global Institute (2011) maintains that the public health service may be able to realise substantial efficiency gains by using Big Data, for example by using the technology to reduce the number of cases of incorrect treatment at hospitals. Big Data is further predicted to be important for both individual patient treatment and in preventive health work at the population level.

3.3.1 Health research

At present, we primarily find examples of Big Data in research and preventive health work. Among other things, Big Data has proven to be able to predict the outbreak and spread of epidemics with a

high degree of precision. A research project under the direction of the Harvard School of Public Health has studied the pattern of the spread of malaria by collecting location data from the mobile phones of 15 million Kenyans, and compiling this with the Kenyan authorities' database of information on malaria outbreaks (HSPH News 2012). The compilation of the data sets made it possible to predict how malaria spreads among the different regions of the country. The traditional collection of statistics does not tell us this until after a disease outbreak has taken place. Then it may be too late to react. A computer that instead looks for patterns and communication through social media, or by analysing location and communication logs, for example, may provide early indications of a negative development having started, and where it is, so that measures can be implemented early enough.

Researchers have also been successful in detecting dangerous side effects of various medicines by using Big Data analysis. Researchers at Stanford discovered that two different medications (an antidepressant and a headache pill) could have fatal consequences for the user if they were taken in combination (Tatonetti et al. 2011). They discovered this by compiling aggregated data from medical records with national registers of reported side effects. This information was then compiled with 82 million searches performed on Microsoft's search engine Bing. The results of the analysis showed that people who took *both* preparations searched for words related to side effects such as "headache" and "fatigue" to a greater extent than those who only took one of the preparations. Researchers thus identified a pattern indicating that taking both preparations at the same time could trigger serious side effects.

The use of Big Data in the treatment of patients is not yet very widespread (Bollier 2010 and HealthWorks Collective 2013). This is due to several factors. Firstly, there is little knowledge of Big Data in the health sector. Secondly, the opportunity to process and compile data from medical records is strictly regulated in most countries. And thirdly, linking and compiling data in the health service is challenging because there is no common technical infrastructure that can facilitate such analysis.

3.3.2 Sensors and self-logging

A development that will probably stimulate the use of Big Data in the health field is the increasing use of mobile self-monitoring equipment. People are taking more and more control of their own health. They search for health-related information on the Internet, exchange information on disease symptoms in social network communities such as patientslikeme.com, and use mobile applications to measure their own health condition (Bollier 2010). Smart phones can function both as a stethoscope and as a blood pressure gauge. A survey conducted by the Data Protection Authority and the Board of Technology in 2013 showed that 33% of the respondents had used at least one health or training app. The most extreme users of such equipment are called lifeloggers; this is a movement that encourages the logging of one's own health and all other imaginable tasks. The purpose of collecting such data is to share, compile and analyse the data so that they can bring new insight to individuals and society as a whole (Morozov 2013).[16]

[16] Gary Wolf, a technology journalist, wrote the manifest that launched the "Quantified Self" movement. He points out four factors that explain the growth of life logging: 1) electronic sensors have become smaller and more powerful 2) the sensors have become omnipresent now that they are integrated in mobile phones 3)

A consequence of this trend is that the flow of health information will increase. Health information will be available for analysis in new ways and for more – and other types of – actors than at present.[17] The public health service will have problems receiving all the health information that is generated by the use of mobile and self-monitoring equipment. The users of such technology will therefore probably be dependent on private actors that can analyse the information. Commercial enterprises will thus be able to build up large databanks of health information. This is information that they can use themselves, or sell access to. Health information tells a great deal about us and will therefore be attractive to many actors, such as research institutions, insurance companies, employers and banks.

3.4 Big Data in the police, security and intelligence services

In the following, we will look at how large-scale data analysis is used in the police and intelligence services.

3.4.1 Smart police

Big Data can make the police smarter. Through the use of advanced analysis techniques, the police can uncover previously unknown correlations in crime data and other available data sources. Trends and patterns can be used to define the probability of a future development. This can help the police to predict events, distribute resources and perhaps even to prevent the occurrence of events (Data Protection Authority and Board of Technology 2013). The use of Big Data by the police is called *predictive policing*, and there are many who believe that this will revolutionise the way police work is carried out (Morozov 2013).

The police in Norway have not yet implemented such advanced analysis technology[18]. In the USA on the other hand, there are several examples of the use of Big Data by the police. The police in Los Angeles (LAPD) have for example implemented an analysis tool called *PredPol*, which was originally developed to predict earthquakes and aftershocks. *PredPol* is fed with local crime statistics on car theft, break-ins and other relevant information with a view to combating crime. The LAPD can now predict where and when it is probable that a given criminal offence will take place – and do so within areas as small as 150 m². With the help of mobile digital maps, police patrols can use this information to stay on top of criminal incidents. The police already being present before a crime is committed has of course resulted in a substantial reduction in crime (Data Protection Authority and Board of Technology 2013).

In Europe, the police in the UK have come the furthest in implementing advanced analysis technology. In connection with arranging the Olympic Games in 2010, the police authorities in

---

social media have normalised and facilitated a culture of sharing 4) the growth of cloud services has made it possible to store and compile health data (and other personal data) in new ways (Morozov, 2013).

[17] Within the so-called self-monitoring movement (lifeloggers' movement) the importance of ownership of your own health data is emphasised; everyone should have the right to obtain a copy of their medical records and an opportunity to use this information as they see fit (Bollier 2010).

[18] The 22 July Commission delivered a devastating judgment of police use of technology. They pointed out in their report that the Norwegian police must become better at exploiting the potential that lies in information and communication technology: http://www.22julikommisjonen.no/Rapport

London used Big Data technology to perform real-time sentiment analysis of words and expressions compiled from social media (Hewlett-Packard 2013).

Prediction analysis is also used in other parts of the justice sector. Several states in the USA are testing a system designed to predict how probable it is that an inmate will kill or be killed while on leave. Based on this system, the authorities decide whether to accept or reject leave applications. (Mayer-Schönberger and Cukier 2013).

3.4.2 No needle without a haystack

The intelligence services in the USA are in the forefront when it comes to the implementation of new and powerful analysis technologies. An article in *The New York Times* (2013), points out that the NSA and the CIA have been testing IBM's Big Data technology for two years. The NSA has recently constructed an enormous data centre in Utah, five times larger than Capitol Hill, dedicated to Big Data analysis.

Big Data has changed the way intelligence work is performed. Formerly, this work started with suspicious individuals and an attempt was made to put them under surveillance with as much detail as possible by phone tapping and other means. We can go the other way with the help of Big Data. The process starts by the collection of enormous amounts of data, and then using this data to look for patterns and correlations that can reveal suspicious events or persons. It is hardly of interest for the intelligence authorities to use anonymised data in this context. Surveying individuals and relations between individuals is one of the key purposes.

In the aftermath of 11 September 2001, the intelligence authorities in many countries have been granted expanded authority to collect and analyse personal data from a long list of sources. In June 2013, the British newspaper *The Guardian*, via an internal whistle-blower in the NSA, leaked top-secret documents concerning the PRISM programme (*The Guardian* 2013). The documents revealed that the NSA had allegedly been granted direct access to emails, chat correspondence, voice calls and documents from Microsoft, Yahoo, Google, Facebook, Paltalk, AOL, Skype, YouTube and Apple. However, representatives from the companies in question stated that they did not know about either the programme or that the authorities allegedly had such direct access to their servers. The companies maintain that they only disclosed information on the basis of specific requests from the authorities, and that these requests were handled in accordance with established rules. The EU has criticised the American authorities in connection with the revelations concerning PRISM, and EU Commissioner Viviane Reding has, for example, stated that:

> *"The concept of national security does not mean that "anything goes": States do not enjoy an unlimited right of secret surveillance. In Europe, even in cases involving national security, every individual – irrespective of their nationality – can go to a Court, national or European, if they believe that their right to privacy or to data protection has been infringed. I have made my point clearly: this is what I want for European citizens also in the US". (Reding 2013).*

All of the facts in this case are still not on the table. Nevertheless, the fact is that the enormous quantities of data containing personal information that the major Internet companies possess are of great interest to many actors. The intelligence services are no exception in this context. To which actors data from the major Internet companies flows is of great importance to the individual citizen. The use of Big Data contains the potential to facilitate a massive surveillance society.

The extent to which Norwegian intelligence services have implemented Big Data technology is not known to the Data Protection Authority. The intelligence services in Norway are not subject to the Personal Data Act, but are regulated by separate special legislation. The EOS Committee, an oversight body appointed by the Norwegian Parliament, oversees the intelligence, surveillance and security services

3.5 Privacy challenges

As we have seen in the prior examples, Big Data can be used for many socially beneficial purposes. Big Data is largely used to identify overarching trends and correlations that do not necessarily lean towards privacy protection. However, Big Data can also be used so that it affects individuals directly.

In some sectors, the use of Big Data is aimed more at individuals. In health research, the purpose of Big Data is primarily to identify patterns and correlations at the population level – not to acquire deeper insight into individuals. For Internet-based companies engaged in social media or e-commerce, as well as for insurance and credit rating enterprises, however, it is of great interest to acquire deeper insight into the behavioural patterns of individuals. For the police and the intelligence services it is of interest to use Big Data both to uncover patterns in the crime picture and to reveal suspicious behaviour among individuals.

Certain forms of the use of Big Data may be in direct conflict with current data protection legislation. Other forms will not be directly unlawful, but may result in pressure on key privacy principles, which may in turn have unfortunate consequences for society as a whole. In the following, we point out key privacy challenges related to the use of Big Data, under both of the two categories mentioned above. How the challenges must be handled within current legislation is discussed in Section 4.

3.5.1 Use of data for new purposes

To a large extent, Big Data involves extracting secondary value from collected data:

> "*Unlike material things – the food we eat, a candle that burns – data's value does not diminish when it is used; it can be processed again and again. (…) Just as data can be used many times for the same purpose, more importantly, it can be harnessed for multiple purposes as well*" (Mayer-Schönberger and Cukier 2013:101)

This challenges the privacy principle that data must only be collected and subsequently used for clearly stated purposes.[19]

Google's collection and use of information from its users in connection with the development of services, such as Google Now, is an example of this. The European data protection authorities are now demanding that Google clarify the purpose of the collection of information from users, and clarify how the information is linked between the company's various services. [20]

[19] Cf. Section 11, first paragraph, letter c) of the Personal Data Act.

[20] In October 2012, the European data protection authorities sent a letter to Director Page at Google demanding that the company implement a number of requirements to comply with the European Data Protection Directive 95/46/EC: http://www.cnil.fr/fileadmin/documents/en/20121016-letter_google-

Public enterprises are also interested in compiling data in new ways to create better and more efficient services. Smart police, for example, will be hungry for data. Data that were originally collected for administrative purposes may be useful in an investigation when they are compiled with information from another register, for example, and are used in that way to uncover correlations and patterns (Data Protection Authority and Board of Technology 2013). The sharing and reuse of data may result in a more effective police force, but this means that information is used for other purposes and in other contexts than originally intended

The opportunity that lies in Big Data to compile ever larger data sets, and the – in many cases very valuable – knowledge that can be extracted from such analysis, may place the principle of purpose specification under pressure. This challenge must be seen in a clear context with the challenge below, the challenge of data maximisation.

3.5.2 Data maximisation

Big Data entails a new way of looking at data, where data is assigned value in itself. The value of the data lies in its potential future uses. Such a view of data may influence the enterprises' desire and motivation to delete data. Neither private nor public enterprises will want to delete data that may at some point in the future, and in compilation with other data sets, prove to be a source of new and valuable insight. Who could have predicted, for example, that the search history of Bing could contribute to uncovering serious side effects from the simultaneous use of two specific medical preparations (see Section 3.3.1)?

Big Data's business model is the antithesis of data minimisation, which is a key principle of privacy protection[21]: More data about persons than necessary to fulfil specifically stated purposes shall not be collected and stored. The data shall be deleted when they are no longer necessary for the purpose. The more extensive use of Big Data may result in it being even more challenging for the Data Protection Authority to monitor that the deletion obligation is fulfilled in accordance with the Personal Data Act.[22]

3.5.3 Lack of transparency – loss of control

The right of access to and the right to information on the processing of one's own personal data are important privacy principles. Lack of transparency and information on how data are compiled and used may entail that we fall prey to decisions that we do not understand and have no control over: What profiles are there on me out there? On the basis of what personal data are they formulated? Am I considered an attractive or valueless customer?

The average Internet user, for example, has little insight into how the advertising market on the Internet and other digital platforms works, and into how his personal data is collected and used by

---

article_29-FINAL.pdf. As a general explanation, there were three demands that were made but have still not been implemented:

- That Google should provide adequate information to its users about what the purpose was for the processing of personal data and about what types of data were processed.
- That Google should clarify what the purpose and means were in connection with linking data among the company's services.
- That Google should provide information on how long they stored personal data for the various services.

[21] Cf. Article 6 (1) c in Directive 95/46/EFC

[22] Cf. Section 11, first paragraph, letter e of the Act, cf. Section 28.

commercial interests (Turow 2011). In a survey conducted by the Data Protection Authority, it was revealed that very few suppliers of mobile applications informed users of what personal data was collected and of how these data were processed (Data Protection Authority 2010). Many of the actors that operate in this market, especially third-party actors such as data brokers and analysis firms, are, moreover, unknown to most people. The right of individuals to access the personal data that has been collected then becomes difficult to practice.

There are also large data brokerage firms that collect personal data via platforms other than the Internet, such as the aforementioned Acxiom. The company has not given the public the right to access the personal data that the company possesses on individuals. After critical media coverage and pressure from the American regulatory authorities, however, Acxiom has signalled that they will open up their databases to access by the end of 2013 (*The New York Times* 2012b, Federal Trade Commission 2012 and Forbes 2013).

3.5.4 Imbalance between enterprises and individuals

Big Data increases the imbalance between large enterprises on the one hand and individuals on the other hand. The enterprises that *collect* personal data are extracting ever-increasing added value from the analysis and processing of this information, not we who *provide* the information. It is more likely that this transaction will be a disadvantage to the consumer, in the sense that it may expose us to future vulnerability.

The OECD (2013) has devoted a great deal of attention to this problem, and has attempted to develop a method for determining the monetary value of personal data. According to the OECD report, a method for establishing the value of personal data will contribute to greater transparency and insight into how the market for the sale of personal data works. Greater awareness in individuals about the value of our personal data may contribute to equalising the relationship between enterprises on the one hand and the individual on the other hand. This can contribute to having higher demands and expectations for how our personal data is handled, among other things.

The OECD report points to a trend that perhaps may contribute to making it easier to establish the value of personal data. Companies have cropped up that offer so-called "data lockers". Services that include *data lockers* give the customers more control over the use of their personal data, for example by allowing them to determine whether their information should be sold to third parties. If the users choose to sell their data, they will receive a commission on the sale.

The company *reputation.com* will make it easier for Internet users to earn money on their personal data. They will launch a service that will enable Internet users to share personal data with various enterprises, in exchange for discounts and other benefits (*MIT Technology Review* 2013b).

3.5.5 Compilation of data may uncover sensitive information

Another challenging aspect associated with Big Data analysis is the fact that information that is not sensitive in itself may generate a sensitive result through compilation. Big Data tools can identify patterns that can indicate something about people's health, political convictions or sexual orientation. Such information is entitled to special protection. Enterprises that use Big Data must therefore be aware of this in the development of algorithms, such that they avoid data being compiled in a manner that reveals vital privacy interests. An example that is often used to illustrate this challenge in the use of Big Data is the American food chain Target's so-called "pregnancy

algorithm" (*The New York Times* 2012c). Target developed an algorithm that can predict which customers are pregnant based on what products they purchase. Target sent special offer coupons for "pregnancy products" to these customers. In one instance, the distribution of such a coupon resulted in the father in the house acquiring knowledge of the daughter's pregnancy before she herself had had an opportunity to inform him. Such use of collected information is contrary to the individual's expectations of the purpose for which the information can be used.

3.5.6 Farewell to anonymity

One of the major challenges associated with Big Data analysis is that of re-identification. Through compilation of data from several sources, there is a risk that individuals may become identifiable from data sets that are supposedly anonymous.

Big Data may consist of a combination of identifiable and non-identifiable information. Even if the data sets used are anonymised, there is a risk that the compilation of data sets may entail the re-identification of natural persons (or even the assumed re-identification of incorrectly assumed natural persons, so-called false positives).

The re-identification can take place by someone taking personal data they already have about others and searching for hits in an anonymised data set, or by a hit from an anonymised data set being used to search for hits in publicly available information.

The risk of re-identification can be reduced by ensuring that only anonymised data is included in the analysis. It is, however, not always easy to obtain an overview of whether a data set is completely anonymised, or if it still contains personal data. This may be difficult for two reasons:

The concept *to identify* – and thus *to anonymise* – is complicated because natural persons can be identified in a number of different ways. This includes direct identification, when someone is explicitly identifiable from an individual data source (for example, a list of full names), and indirect identification, when two or more data sources must be combined in order for identification to take place.

Organisations that have plans to release an anonymised data set may be satisfied that the data in themselves will not identify anyone. However, what they do not know is that in some instances it may be possible that other data are available that make it possible for a third party to re-identify persons in the anonymised data set.

Even after identifying information has been deleted, it is still possible to link specific information to an individual based on links that are available in various Big Data collections. An example of this is "*How to break anonymity of the Netflix Prize Dataset*" (Narayanan and Shmatikov 2008). Netflix announced a competition for developers with a prize of USD 1 million. The goal was for someone to develop a solution that resulted in a 10% improvement to their recommendation module. In this connection, Netflix released a "practice data set" to the competing developers that they could use to test their systems The data set contained a disclaimer, which stated "to protect the privacy of our customers, all personal information that identifies individual customers has been removed and all customer IDs have been replaced by randomly assigned IDs." There are several movie review portals on the Internet, among these is IMDb. Individuals can register on the IMDb website and rate movies, and they appear with their full name. Researchers Narayanan and Shmatikov linked Netflix's

anonymised practice database with IMDb's database (based on the date of the review by a user) and managed thus in part to re-identify the users in the Netflix practice database. There is also a long list of other such examples. [23]

Just as a person's fingerprint can identify a natural person at a location where a criminal act has been committed, data fingerprints can do this as well. It is a combination of data values that are unlike any other combination of data values in the table. It has been much easier for researchers to find such data fingerprints in anonymised data sets than most people would believe possible. As soon as someone has found a unique data fingerprint, they can link the relevant data to additional information. Many anonymisation techniques would have been perfect if unauthorised parties did not know anything at all about the world's entire population. In reality, the opposite is the case, and new databases are created every day with information on people. The Netflix study shows that it is very simple to identify individuals in anonymised data.

3.5.7 Incorrect data

It is an important principle of privacy protection that decisions of consequence to individuals must be based on correct information. Basing decisions on information retrieved and compiled from social media for example, entails a risk of making decisions on an incorrect factual basis. Decisions based on the compilation of such information will not be transparent and verifiable to the same degree as decisions based on information retrieved from official registers. It is also important to bear in mind that data collected from social media, for example, do not necessarily give a correct picture of a person. A weakness of Big Data analysis is that it often does not take context into account[24]. Basing decisions on information that is intended for other purposes and arising from another context may yield results that do not correspond to the actual situation.

3.5.8 Data determinism:

The Big Data line of thought rests on the assumption that the more data one collects and has access to, the better, more justified and accurate the decisions made. However, more data harvesting does not necessarily mean more knowledge. More data can also mean more confusion and more false positives: "*Big Data is driven more by storage capabilities than by superior ways to ascertain useful knowledge*" (Bollier 2010).

More extensive use of Big Data, and the associated use of automated decisions and prediction analysis may have unfortunate consequences for the individual. Big Data can consolidate existing prejudices and stereotypes, and reinforce social exclusion and stratification. We can obtain a society with an A and a B class, in which only those with the "right" profile will be given priority. Profiles at the aggregate level may also be completely wrong for natural persons. A development where more and more decisions in society are based on the use of algorithms may result in a "Dictatorship of Data" (Mayer-Schönberger and Cukier 2013); where we are no longer judged on the basis of our actual actions, but on the basis of what the data indicate our probable actions will be. Individuals are more than the sum of the digital tracks they leave behind.

---

[23] Another example is a case in which American researchers managed to re-identify DNA material that was stored in a de-identified form (Gymrek et al. 2013).

[24] Danah Boyd and Kate Crawford are two researchers that have pointed out the importance of taking the context into account in Big Data analysis (2012).

It is also important to bear in mind that algorithms are not objective: *"The prejudices of a society are reflected in the algorithms that are searched"*, the author of the book The Numerati, Stephen Baker, stated (Bollier 2010). Algorithms are probability models based on historical data, and they are born in a social context (Morozov 2013). More use of profiling techniques and predictive analysis in police work, for example, may result in people feeling incorrectly criminalised on the basis of ethnicity, place of residence, etc.

Is it possible that in the future one may be arrested based on an algorithm? Can such probability forecasts be adequate to give the police "probable cause for suspicion" and the opportunity to search persons who would not otherwise have been stopped, or to knock on doors that would not otherwise have been knocked on (Data Protection Authority and Board of Technology 2013)? Casting suspicion on people, not because of what they have done, but of what they may do in the future, challenges the very principle on which states governed by the rule of law are founded; that one is innocent until proven guilty. Such use of algorithms will also undermine the view that people have free will and can themselves make moral choices.

In Norway and many other countries, laws have been passed in recent years, which entail that it doesn't take much to be arrested based on suspicious conduct. The use of Big Data in light of such laws may make individuals extra vulnerable relative to inadvertently ending up under scrutiny by the police.

3.5.9 Chilling effect

If we see a development in which credit scores and insurance premiums are based on almost all the information we leave in various contexts on the Internet and otherwise in our daily life, this may have consequences for privacy and for how we behave. In 10 years perhaps, your children may not be eligible for insurance because you have shared on social networks that you have a predisposition for a hereditary disease. This may result in us placing restrictions on how we participate in society, or in actively adapting our actions – both online and otherwise. We fear the consequences with regard to whether we will be granted loans, obtain car insurance, become a tenant, etc.

With regard to the use of Big Data by intelligence authorities, secrecy concerning the data sources they collect information from, and of how the data are used, threatens the privacy interests of more than just individuals. It also challenges trust in the authorities, and in the last instance – the actual foundation for an open and well-functioning democracy.

What are the consequences when we are uncertain as to how the information we provide in various contexts on the Internet and via mobile phones is analysed and potentially used for new, and for us unknown, purposes? Will we then dare to express ourselves as freely? Humans who know they are being watched change their behaviour because the context is different – trust in their environment changes. Poorly safeguarded privacy protection may weaken democracy if citizens restrict their participation in the open exchange of ideas. Extensive use of Big Data in intelligence contexts, and by the police in general, may at worst have a chilling effect on the freedom of expression, if the premises for its use are kept concealed and cannot be verified. There are also other public enterprises that may be interested in retrieving information on citizens for control purposes. The customs and tax authorities and the Norwegian Labour and Welfare Service may be better equipped to combat smuggling, tax evasion and benefit abuse through the use of Big Data. It is the uncertainty

connected to what information is considered relevant and of interest to the authorities, not necessarily what the reality is, that may affect the candouor of the citizens.

3.5.10 Echo chambers

Personalisation of the Internet, with custom offers based on the behaviour of individuals on the Internet, also affects the framework conditions for the exchange of ideas – an important foundation for a well-functioning democracy. This is not primarily a privacy protection challenge, but constitutes a challenge for society at large. Eli Pariser (2011) and Joseph Turow (2011) point out the risk of "echo chambers" or "filter bubbles" – i.e. personalisation of the Internet results in one being only exposed to content that corresponds to one's own profile. The Internet, and thus society, is divided into various "boxes" that do not have any contact with each other. This will place a damper on the exchange of ideas, because one will now be exposed to a lesser extent to opinions that deviate from one's own.

4. Legal questions

The analysis of Big Data can lead to valuable results in several sectors. However, Big Data raises certain legal questions. The answers to these questions are not apparent. There are few Norwegian and European sources of law that directly address this subject.[25] The problems must consequently be assessed based on the general, technology-neutral and discretionary rules of law that do exist in the area. For our part, the legal point of departure will first and foremost be the *Personal Data Act*.[26]

Below we will look at some of the legal questions that arise from the use or processing of Big Data in light of the Personal Data Act. We will also examine what the personal data concept in the Act entails, since this term is of such key importance.

4.1 Big Data and the law

The concept of Big Data does not encompass any specific form of data processing. As mentioned, the term lacks a precise definition, and it may encompass both the collection and the analysis of information. The different stages in the analysis process may also vary, as we have seen in Section 2. That the data processing involves information on physical persons is not obvious either – the purposes of the analyses can often be achieved by means of anonymous or anonymised information. In other cases, it is precisely information about individuals that is of interest.

The processing of Big Data may trigger obligations and rights pursuant to the Norwegian Personal Data Act. The prerequisite here is that the processing concern personal data. The Act namely applies to the "processing of personal data wholly or partly by automatic means", pursuant to Section 3 of the Act. That the processing of Big Data takes place by automatic means lies in the nature of such data.

Thus the question of what is encompassed by the concept of personal data is of decisive importance for the application of the Act. This question, however, is not always so easy to answer, as we shall see below.

---

[25] We do not know of any decisions handed down in Norwegian courts or by the central government that refer to Big Data as a phenomenon. A search for the term in Lovdata resulted in no hits on 30 May 2013. The search included all the courts, acts, regulations and decisions from the Privacy Appeals Board, in addition to legal literature.
A search for "Big Data", in combination with terms such as *privacy* and *data protection* in international legal article databases, results in only a handful of hits, and the articles are often based on American law. A search in *LexisNexis* in June 2013 resulted in 13 hits, few of which were relevant.

[26] The Personal Data Act, Act no. 31 of 14 April 2000 relating to the processing of personal data builds on 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).

4.2 Personal data concept

According to the definition of the concept of personal data in the Act, this is a matter of *"information and assessments that can be linked to a natural person"*.[27] The definition[28] contains three main components:[29]

- any form of information
- that can be linked to
- an identifiable or identified natural person

Even if the components of the definition are closely related, and they affect each other reciprocally, we will review each of them separately below.

4.2.1 Any form of information

The first element can be interpreted very broadly. In brief, it covers all imaginable information, regardless of the nature, content or format.

Firstly, the formulation encompasses objective information, such as information on a person's age, place of residence or income. The same applies to subjective information, such as one person's assessment or characterisation of another person. If Peter tweets that John cannot be trusted, this is an example of such subjective information.

In this context, the content of the information is of no importance. Whether the information appears to be important to those it concerns or others is of no importance. That the information is not credible or possible to prove is also immaterial. Moreover, it is of no consequence whether the information directly concerns matters that are traditionally associated with privacy. Information that originates from the public domain, or from the workplace, is also encompassed by the concept of information in the Act. The same applies to both sensitive and general information.

The concept encompasses all information, regardless of the format. Information may be expressed through writing, numbers, drawings, photographs, sound or biometric characteristics. With regard to the sources of information, they may be anything from emails to Post-it notes, public case documents, messages on social media networks, text messages, digital or analogue photographs or sound recordings, and so forth.

4.2.2 Linking element

It must, however, be possible to link the information to a natural person. This means that the information can say something about – or that the information concerns – an individual. Sometimes this link is quite apparent, such as in medical records, personnel files at work or in the registers of a credit information company.

At other times the link is less clear. It is reasonable to assume that information on the condition of a house or a car will primarily be associated with the object itself. The same information may,

---

[27] Section 2, no. 1 of the Act.
[28] Opinion 04/2007 on the concept of personal data
[29] In certain presentations, the linking element and the identification requirement are bundled together (Schartum and Bygrave 2011)

however, also reveal other circumstances, for example concerning persons who have been involved with the object. Information on the value of a house or a car's repair history may also say something about the owner of the house or the person who has used the car. In certain cases, information on one person may also be information on another person at the same time, as may be the case in medicine or genetics.

In other words, it is sufficient that the link between information and person is indirect. The link will also be indirect when the information primarily concerns a group, but can say something about the individuals in the group at the same time. This requires that the individuals are *identifiable*. We will look at what this term means below.

4.2.3 Identifiable natural persons

It must be possible to link the information to an *identified or an identifiable natural person*.[30] This means first of all that information on legal persons falls beyond the substantive scope of the Personal Data Act.[31] Secondly, the natural person must be *identifiable*.

Natural persons

It may be difficult to determine in practice where the line should be drawn between information concerning legal persons and information concerning natural persons. Information on a sole proprietorship, for example, will also be able to say something about the proprietor, especially if the company does not have any employees. Other information that primarily emerges as company information may also reveal circumstances that can be linked to an individual. The processing of this information will then be subject to the provisions of the Personal Data Act.

Identifiability criterion

That it must be possible to identify the natural person is not expressed directly in the legal definition in the Personal Data Act.[32] This is evident, however, from the corresponding definition in the Directive, which makes reference to the fact that it must be possible to link the information to an *identified* or *identifiable* person.[33]

That person P has been *identified* means that P can be distinguished from a group of persons. P is thus *identifiable* when it is *possible* to identify him, even though the identification has not yet taken place. That P has been identified, will normally be easy to ascertain. Therefore it is the identifiability criterion that defines the outer limit for the concept of personal data.

This means that the Personal Data Act will apply if it is possible to distinguish one individual from another. The fact that it is conceivable that the identification will take place at some point in time in

[30] Cf. wording of Article 2 of Directive 95/46/EC.
[31] The Personal Data Act applies, however, also to the processing of credit information on persons other than natural persons, cf. Section 4-1 of the Personal Data Regulations.
[32] Section 2, no. 1 of the Act.
[33] In the legislative background to the Personal Data Act, it is also expressed that the identifiability criterion lies implicit in the concept of a "natural person", see Proposition no. 92 (1998-1999) to the Odelsting, comments on Section 2 in Chapter 16.

the future is thus sufficient. It is not necessary either that it be the data controller himself who has the opportunity to link the information that makes the identification possible.[34]

Information that appears to be anonymous may prove to be personal data in the sense of the Act. The explanation is that it is possible to identify one or more persons *indirectly*. For example, IP addresses are considered personal data in certain contexts.[35] Internet providers possess lists of subscribers and their assigned IP addresses. These can be aligned, so that the identity of the subscriber is revealed.[36] Other information that emerges together with such addresses, will thus also be considered personal data.

4.2.4 Summary – personal data concept and Big Data

The combination of the aforementioned elements make the personal data concept quite broad. The conclusion is that the Personal Data Act, which defines its own scope by means of this concept, has a correspondingly broad area of impact. The legislation essentially applies to the processing of all forms of personal information, as long as it is possible to link it to an identifiable individual in the future. Whoever processes such information in connection with some form of Big Data, thus has a set of statutory restrictions to observe.

If, on the other hand, the information is *anonymous* or *anonymised*, the rules will not place any restrictions on processing. In other words, Big Data analysts have more leeway when dealing with such information. Anonymous information can be defined as information that it is not possible to link to an identifiable individual, when all of the reasonable means that can conceivably be used to identify the individual are taken into account, either by the data controller or by any other person. The term "anonymised information" refers to information that has previously been linked to an identifiable person, but where the link to identifiable persons has been made impossible.[37]

The legal requirements that we will discuss below, thus apply only if the information can be linked to identifiable persons, and not to the processing of anonymous information.

4.3 Legal requirements for Big Data processing

Whoever is responsible for the processing of Big Data – the data controller[38] – must therefore observe the provisions of the Personal Data Act. The basic requirements in Section 11 of the Act are

---

[34] Cf. Paragraph 26 in the Directive's preamble: *"to determine whether a person is identifiable, account should be taken of all the means likely to be reasonably used either by the controller or by any other person ".*

[35] Cf. practice of the Data Protection Authority and Privacy Appeals Board (see, for example, the Board's case 2011-10), the Swedish Data Inspection Board (D no. 1402-2007) and the French Commission Nationale de l'Information et des Libertés (CNIL), see http://www.cnil.fr/linstitution/actualite/article/article/ladresse-ip-est-une-donnee-a-caractere-personnel-pour-lensemble-des-cnil-europeennes/

[36] The subscriber can of course be a legal person, such as the proprietor of a café. In this case, the café's IP addresses will not automatically be linked to an identifiable natural person. If the café keeps lists of Internet users etc., for example, it may make identification possible, and thus the IP addresses will be considered personal data.

[37] Opinion 04/2007 on the concept of personal data

[38] According to the legal definition in Section 2 no. 4 of the Personal Data Act, the data controller is the person who determines the objective of the processing of personal data and the tools that are to be used. The English version of the Directive uses the term "data controller".

of key importance in this connection. If the data processing does not satisfy all the conditions that are stipulated in this section, then the processing is unlawful. Whoever plans to process personal data must in other words make sure in advance that the data processing is not in contravention to Section 11 of the Personal Data Act.[39]

Below we will take a closer look at what these requirements involve. What their importance will be in connection with various forms of Big Data processing are of particular interest. In this context, we refer to the examples that are presented in the sections above.

4.3.1 Legal grounds for the processing of personal data

Section 11 of the Personal Data Act is introduced by making reference to the fact that no one can process personal data without there being legal grounds for data processing. This point of departure is in other words a general ban on such data processing as is encompassed by the Act, but the Act is also equipped with specific legal grounds that nevertheless make data processing lawful. The Act specifies several different grounds, such as consent, statutory authority and various grounds of necessity.

Statutory authority

The statutory processing of personal data means that rules have been prescribed in or pursuant to a formal law that requires or permits specific processing of personal data. Most often this concerns laws that impose duties and obligations on the authorities. Examples of this include the Immigration Act[40] and the Police Register Act.[41] Both of these acts prescribe specific rules concerning the opportunity of the immigration authorities and the police to process certain categories of personal data. The implementation of the Data Retention Directive in Norwegian law will also entail a statutory storage obligation in the private sector, namely for the providers of telecommunication and Internet services.

Whether a statutory provision provides legal grounds for a specific data processing must be determined by an interpretation of the individual provision. The requirement of legal authority is relative. The greater the possible impact on privacy by the processing, the clearer the wording of the provision must be.[42] Processing regulated by other acts must also satisfy the requirements stipulated for lawful processing in Article 7 of the EU Directive, provided that the act in question comes under the areas that are covered by the EEA Agreement.

We will not delve further into the legal authority grounds in this report; we will review the alternatives that are most relevant in a Big Data context below.[43]

---

[39] Unless the general exceptions in Section 3, second paragraph of the Act (private purposes) or Section 7 (freedom of expression) apply, or the general conditions in the introduction to the Act have not been satisfied, such as the provisions concerning the geographic scope in Section 4, etc.

[40] Act no. 35 of 15 May 2008 on the entry of foreign nationals into the Kingdom of Norway and their stay in the realm.

[41] Act no. 16 of 28 May 2010 on the processing of information by the police and the prosecuting authority; the Act enters into force as determined by the King.

[42] This follows from the *principle of legality*, and is emphasised in Proposition no. 92 (1998-1999) to the Odelsting, see the comments on Section 8 of the Act in Chapter 16 of the proposition.

[43] We assume that the other grounds of necessity in the Act – fulfilling legal obligations, protection of vital interests and exercising public authority – are of limited interest in this context.

Consent

One of the questions that the processing of Big Data raises is whether whoever collects and analyses the information must obtain consent from the persons the information concerns. Valid consent in the sense of the Act shall be *voluntary*, *informed* and *express*.[44]

Consent is sometimes presented as the main rule of the Act – which is to say that the processing of personal data in general shall be based on consent from whomever the information concerns.[45] If we take private autonomy as our basis, as well as the idea that individuals should be able to have the right of control over their own personal data – as in the theory of *integrity-focused* privacy protection – this appears to be both sensible and fair. [46]

However, there is no legal cover for such a point of departure. The alternatives in the Act are equivalent – this is evident from the wording of the Act, in which the various grounds are separated by the conjunction "or".[47] Consent must, however, be obtained if none of the other legal grounds in the Act apply.[48] Consent may thus – at least in some contexts – appear as the most appropriate legal ground in practice. The reason for this is the fact that the other processing grounds stipulate thresholds that may be difficult to exceed.[49]

Alternative to consent – contracts

The collection of personal data over the Internet can in certain cases be based on the alternative in Section 8 (a) of the Personal Data Act. This provision states that personal data can be processed when *necessary in order to fulfil an agreement with the data subject*.[50]

The exchange of contact and payment information for the purchase of goods and services is an example of the processing of personal data that can be based on this legal ground. A company that sells goods over the Internet can therefore freely require the information on a buyer that is necessary in order to complete the sale. This alternative may also be applicable to certain information categories in other circumstances. One example of this is location information in connection with downloading a map application for a smart phone.[51]

It is, however, important to bear in mind that the Act stipulates a *requirement of necessity* here. The information can in other words not be processed if the purpose of the processing can be achieved by

---

[44] Pursuant to Section 2, no. 7 of the Personal Data Act.
[45] It is stated on page 108 of Proposition no. 92 (1998-1999) to the Odelsting that *"the processing of personal data should be based to the greatest possible extent on the consent of the data subject, even if it can also be authorised in the grounds that are stipulated in letters a to f. Firstly, this would strengthen the data subject's opportunities to control his own information. Secondly, by basing the processing on consent, doubt concerning whether the more discretionary conditions in letters a to f have been met will be avoided."*
[46] See Official Norwegian Report (NOU) 1997:19 (Ministry of Justice and the Police 1997)
[47] The Privacy Appeals Board goes in the same direction in the decision PVN-2012-1.
[48] Cf. Section 8 of the Personal Data Act, and Section 9, which applies if sensitive personal data is processed.
[49] Cf. PVN-2012-1.
[50] It is a prerequisite that the civil law requirements that apply to a valid formation of contract are satisfied (see Opinion 15/2011 on the definition of consent p. 6).
[51] See, for example, the Data Protection Authority's case 12/00276 (downloading applications for smart phones).

other means. This alternative thus has a fairly narrow area of application, and it should be used cautiously.

Alternative to consent – balancing of interests

Sometimes personal data is collected over the Internet without the users being aware of it, for profiling purposes for example. The data processing must nevertheless be based on at least one of the legal grounds in the Act. The contract alternative will obviously have limitations in these cases. Consent may be an alternative, but there may be many reasons why the data collectors do not want to give the users a genuine choice – an *opt in* solution will, for example, reduce the source data considerably. The only relevant legal ground that is left then is the Act's *balancing of interests alternative*.[52]

Pursuant to this alternative, personal data can only be processed if the processing is necessary in order to enable the data controller, or third parties to whom the information is disclosed, to protect a legitimate interest, *and* the data subject's privacy interests do not override this interest.[53]

This is thus a case in which the pros and cons of two conflicting interests must be weighed. The outcome depends on what interests are found on the scales, interests that must be weighed against each other on a case-by-case basis.

This may involve very discretionary assessments on both sides of the scale. The legitimate interests that are referred to in the provision have been interpreted quite broadly in Norwegian administrative practice. Economic gain is, for example, undoubtedly such an interest in this context. The genuine obstacle on this side of the scale is the *criterion of necessity*. In order to fulfil this criterion, the data controller must demonstrate that the stated and legitimate processing purposes cannot be achieved other than precisely through the specific processing that the weighing of pros and cons is performed in light of.

At the same time, the outcome of the weighing of pros and cons will depend on what privacy interests are at stake. It is thus not adequate that the processing as such is necessary – the interests served by the processing must also weigh more heavily than the conflicting privacy interests. It is difficult to say anything in general about what importance should be attached to the various privacy interests in themselves, but a risk-based approach can be used as a point of departure. The greater the need to protect information on the basis of confidentiality considerations, the greater the importance that must be attached to the privacy interests. In the preparatory work for the Norwegian Act, it is stated, for example, that "*in general, substantial importance must be attached to*

[52] Section 8, letter f of the Personal Data Act.
[53] This provision has a counterpart in Article 7 of the Directive – the alternative is often referred to as the *"balancing of interests test"* or simply *"legitimate interests"*.

privacy considerations when weighing against commercial interests."[54] Privacy interests must also be assessed in light of the fact that privacy is protected by the human rights.[55]

Compensatory measures in the form of initiatives related to information security or the *pseudonymisation* of personal data may also play a role. They can compensate for the privacy disadvantages that the processing would have otherwise entailed.[56]

The Data Protection Authority's recommendation

Even if we cannot rule out that the balancing of interests may legitimise the processing of personal data in certain cases, a specific assessment of this must always be made. It is therefore difficult to use this alternative as a legal ground for the collection and analysis of Big Data in general. In any case, it is the data controller who has the burden of proof that the conditions for the balancing of interests alternative have been met. Since there are many discretionary components and elements of uncertainty in the provision, this may present challenges.[57]

At the same time, in many cases it will be unnecessary to analyse data on identifiable persons in order to achieve the socially beneficial purposes that are often referred to. A large-scale analysis of people's movement patterns in a major city in order to enhance the efficiency of the public transport offerings may entail both environmental and economic benefits. These benefits may, however, be achieved without knowing the identity of those who move around the city.[58]

With regard to the legal basis stipulated in the Act, the Data Protection Authority recommends that the data controller primarily use obtaining a valid consent from the data subjects in advance – this is probably the best way to avoid violation of the provisions of the Act. For instances where there exists a direct relationship between the data controller and the data subjects, for example a customer relationship, such an approach will be practical and manageable. For example, a website that sells books may give the users an opportunity to click to accept that information on what they look at and buy be analysed and used for profiling purposes.

Constant requests for consent on the Internet will result in a form of consent apathy among Internet users, some maintain (Hordern 2013). The argument is that this may paradoxically result in reduced protection for individuals (Hordern 2013, Hildebrandt 2009, Tene and Polonetsky 2012). The data controllers will then obtain consents that are as broad as possible. They will assume that the consent declarations in general will not be read, and they will accordingly process personal data as they see fit. Consent is therefore not suitable as a legal ground for the processing of Big Data, it is maintained.

---

[54] See also "*Who Owns The Future?"* (Lanier 2013) in which the author argues that it is immoral that a few people in control of the largest servers should be able to get rich from collecting and analysing data, without us receiving anything in return for it.

[55] Cf. for example, Article 1 of the Data Protection Directive, Article 8 of the Charter of Fundamental Rights of the EU, and Article 8 of the European Convention on Human Rights (ECHR). ECHR is incorporated into Norwegian law by Act no. 30 of 21 May 1999 on strengthening the status of human rights in Norwegian law.

[56] Opinion 04/2007 on the concept of personal data

[57] The Article 29 Group has announced that they will publish a statement on *"legitimate interests"* by the end of 2013, which is to ensure homogeneity in the interpretation of the criterion across Europe.

[58] Cf. Data Protection Authority's case 09/00163 ("Ruter AS").

The requirement that the consent shall be informed does not entail anything other than the data controller providing the data subjects with an overview of what the data processing entails.[59] The terms and conditions do not have to be 60 pages long, which is something that is made fun of at times.[60] Short and concise information, which only covers the relevant aspects of data processing, could in other words reduce the risk of such apathy.

Secondly, it is not always necessary to obtain consent, for example in cases where the information must be provided to fulfil a purchase agreement. If the privacy threat is low, at the same time that the conflicting interests are weighty, this can also form the basis for the processing of Big Data. And thirdly, the data controller must in any case fulfil his obligation to inform the data subjects. Consent sceptics sometimes appear to overlook these factors.

4.3.2 Purpose limitation principle

The reuse of information is a distinguishing feature of Big Data.[61] The basic requirements mentioned above must also be fulfilled for such reuse.[62] This applies regardless of whether the personal data are processed on the basis of consent, or if the processing is based on another legal ground.

The *purpose limitation principle* is particularly important in this connection.[63] The principle means that collected personal data cannot be used for purposes that are *incompatible* with the purpose of the collection, unless the data subject consents to such use. This principle enables the data subjects to make an informed choice about entrusting their data to the data controller. At the same time, the data subjects are assured of a certain degree of predictability, since the principle gives assurances that the information will not subsequently be processed for completely different purposes without the data subjects' knowledge.

It may be difficult to determine what is considered to be incompatible.[64] However, it is certain that the point of reference for the assessment is the purpose that the data controller is obliged to make explicit when the data are collected.[65] Regardless, it is a prerequisite for the discussion that the other purpose is distinct from the initial purpose. The data controller may, however, not claim any purpose whatsoever in order to circumvent this restriction. The purposes must be expressly stated, and they must be legitimate, in the sense that there is an objective link between the purpose of the processing and the data controller's activities.[66] The data controller should thus think carefully through how the purposes of the processing are formulated before starting to collect the data.

---

[59] See, for example, Section 19 of the Personal Data Act

[60] Moreover, these terms and conditions are not meant to inform the data subject in accordance with the requirements for valid consent, rather they are to give the service providers contractual cover, often based on principles of Anglo-American law, cf. the so-called *"four corners"* principle, see for example http://en.wikipedia.org/wiki/Four_corners_(law)

[61] Cf. quote in the introduction to section 3.5.1.

[62] Section 11 of the Personal Data Act, which corresponds to Article 6 of the Directive.

[63] Section 11, first paragraph, letter c of the Act; this principle is also referred to as the *purpose limitation principle* or the *finality principle*.

[64] What is regarded as a single processing is in turn dependant on the purpose of the processing, see Proposition no. 92 (1998-1999) to the Odelsting, Chapter 16, comments on Section 2.

[65] The purpose must, as mentioned, also be legitimate.

[66] Section 11, first paragraph, letters b and c of the Personal Data Act.

What more is meant by the incompatibility concept must be determined on the basis of a *compatibility assessment*.[67] This does not mean that all other use is excluded. Key elements in the assessment include, for example, whether subsequent use of the information entails disadvantages to the data subjects, and whether the use violates the data subjects' legitimate expectations or if it differs greatly from the grounds on which the collection was based.[68]

If data is collected by public authorities, pursuant, for example, to statutory provisions under the threat of criminal liability, the subsequent processing of such data for the prediction of individual behaviour for insurance purposes would most likely be illegal.[69] For the predictive analysis of collected data, the data controller must ensure that the prediction analysis is not incompatible with the original purpose of the collection.

The principle thus stipulates some limits for the reuse of personal data that have already been collected. This may entail a considerable challenge for the commercial analysis of Big Data. Obtaining consent in advance that encompasses the analysis purposes may be a solution. Often, however, the analyses will be conducted by an actor other than the one that collected the information – as mentioned in Section 3.5.2, Big Data often involves subsequent use of the collected data for new purposes, in order to extract the inherent secondary value of the data.

The safest measure that can be taken to not violate this principle will therefore be to ensure that the information is anonymised. In this case, the subsequent processing will not be regulated by legislation, regardless of whether the original collector or a new actor is responsible for the processing. The anonymisation must, however, be *genuine* – see more about this in Section 3.5.6.

4.3.3 Relevance principle and data minimisation

Another basic requirement in the Act stipulates that personal data can only be processed if they are relevant for the purpose of the processing. The relevance requirement must be understood first and foremost as a requirement that the data controller shall limit himself to only processing personal data that are *necessary* for achieving the purpose of the processing.[70] The relevance requirement is sometimes referred to as *data minimisation*, and it may be an advantage to view it in context with the purpose limitation principle.

We see immediately that the data maximisation line of thought that we described in Section 3.5.2 is in direct conflict with the relevance requirement. The requirement for data minimisation may in other words curtail the large-scale collection and accumulation of personal data that is motivated by the future potential value of the data sets.

Once again, we see that anonymisation of personal information can give data collectors greater leeway, provided that they ensure genuine anonymisation of the information.

---

[67] See Opinion 03/2013 on purpose limitation, cf. the term "incompatible" above.

[68] Proposition no. 92 (1998-1999) to the Odelsting, Chapter 16, comments on Section 11 of the legislative proposal.
l.c. See also the Supreme Court decision in Norwegian Supreme Court Reports 2013 p. 143.

[70] Schartum and Bygrave (2004) point out that the relevance requirement may express several factors, such as *logical*, *legal* and *cognitive* relevance. We will not delve further into what distinguishes these factors from each other here.

4.3.4 Obligation to ensure that the data are correct

Another basic requirement in the Personal Data Act stipulates that the data controller shall ensure that the information that is processed is *correct*.[71] The requirement should be interpreted in this context based on the fact that it applies to all forms of processing of personal data, and not just at the collection stage.[72] The data controller has thus an obligation to ensure that the conclusions drawn about individuals are based on Big Data that are correct. If the analyses show that there is an 80% probability that persons who like X will be exposed to Y, it is not possible to conclude that the causality will occur in 100% of the cases. Discrimination based on statistical analysis can thus also be a question of privacy. This question is perhaps particularly relevant in connection with various forms of profiling or predictive analysis at the individual level.

If personal data are openly available, for example on the Internet, and are subsequently used for new purposes by another data controller, this responsibility will apply. If Peter tweets that Mary is a swindler, company S, which provides consumer loans at a high interest rate, can only process the information that this Mary is a swindler if the company can ascertain that the claim is correct.[73]

Transparency, for example in the form of the right of access to information that is processed on oneself, is also a prerequisite to ensure that the data subject can safeguard his interests. Pursuant to the regulations, a demand may be submitted for the correction or deletion of information, assessments or claims that prove not to be correct.[74]

4.4 Rights of the individual

4.4.1 Transparency – information and access

The Personal Data Act stipulates that data controllers must provide information on the data processing to those who are affected by it.[75] The obligation to provide information arises when the information is collected, and it applies regardless of whether the personal data are collected based on consent or on another legal foundation. If the data are collected from the data subject himself, the information shall be disclosed in advance. If the data are collected from individuals other than the data subject, the information shall be provided as soon as the information has been collected. The information is also to be provided without solicitation. In addition, the data subject has the *right to access* the information, that is a right to request further information on the data processing.[76]

The intention of the provisions is to give the data subject an opportunity to obtain an overview of the information being processed about him. How the information is processed, what the purpose of the processing is, whether the information is to be disclosed and where it has been obtained, are also matters that must be disclosed. This information is a prerequisite for the data subject's right to

---

[71] Section 11, first paragraph, letter e of the Personal Data Act, cf. Article 6 (d) of the Directive, which makes reference to the fact that the data must be *"accurate"*.
[72] This point is illustrated by reference to the text of the Directive, which adds that the information "*must be (…), where necessary, kept up to date*".
[73] Since the requirements in Section 11 of the Personal Data Act are cumulative, it is a prerequisite for this discussion that the other basic requirements have been met.
[74] See Section 27 of the Personal Data Act.
[75] Sections 19 and 20.
[76] Section 18

control that incorrect information is not being processed, cf. Section 4.3.4, and for the right to demand that incorrect or incomplete personal information be corrected or deleted.

4.4.2 Personal profiles and automated decisions

An expanded obligation to provide information arises under two circumstances. The first circumstance is if the data controller makes decisions that are completely based on the automated processing of personal data, and the decision is of significant importance to the data subject. If so, the data subject is entitled to require that the data controller provide an account of the rule content – in other words the *logic* or *algorithm* – in the software that "makes" the decision.

The other circumstance is when the data controller contacts the data subject or makes decisions aimed at him, based on personal profiles.[77] The content of this information does not differ significantly from what should be disclosed in accordance with the general rules in Sections 19 and 20, but here it is no longer the collection of data that triggers the requirement. It is evident from the provision that the disclosure requirement arises when the decision is made.

If the specific processing of Big Data entails such contact or decisions as mentioned above, the controller will have an expanded obligation to inform the persons that the information concerns. There are few examples of these provisions being invoked in practice, but it is not hard to imagine that they will gain renewed relevance in connection with the processing of Big Data. Personalised marketing based on profiles that are prepared by means of various tracking techniques on the Internet is one of the examples described in Chapter 2. The right to demand an account of the content of algorithms on which decisions of significant importance are based may also prove to be an important guarantee of the legal protection of individuals.

4.4.3 Correction and deletion

The right to access and the duty to inform are prerequisites for the data subject's right to require the correction of incorrect or incomplete personal information. As we touched on in Section 3.5.2, different variations of the Big Data phenomenon may lead to an aversion to the deletion of data, since the data have a value in themselves.

The deletion rules can be divided into two categories. Firstly, the data controller is obliged to delete personal data that are no longer necessary to process in order to fulfil the original purpose of the processing.[78] In this situation, we see that the information has been relevant, but that it has "played out its role" after a while. One example of this is the traffic information that telecommunication companies store for invoicing purposes. When the invoicing has been completed, the information should be deleted.[79] If the telecommunication companies want to subject these data to Big Data analysis, the companies must then either obtain consent from their customers, or ensure that the information is anonymised.[80]

---

[77] Sections 21 and 22.
[78] Section 28 of the Personal Data Act, cf. Section 11, first paragraph, letter e.
[79] When the amendment act for the Data Storage Directive enters into force, it will be possible to store the same information longer, out of consideration for crime-fighting purposes, see Act no. 11 of 15 April 2011 on amendments to the Electronic Communications Act and Criminal Procedure Act, etc. (implementation of the EU Data Storage Directive in Norwegian law).
[80] Unless the analyses are compatible with the invoicing purpose.

Secondly, information for which processing is not permitted must also be deleted.[81] We can envision that there will be a desire to anonymise unlawfully collected personal data, so that the anonymised information can be analysed. It is difficult to say whether the Act prevents such a practice, anonymisation is often equated with the aforementioned deletion according to existing legislation. In a Big Data context, anonymisation has also become a more unpredictable affair than it was just a few years ago. Under any circumstances, it would appear to be unreasonable if the data collector could achieve economic gains based on an act or practice that was essentially unlawful.

4.5 Some international questions

4.5.1 Choice of law

What country's laws and regulations regulate the processing of Big Data? The answer is the rules of law that apply in the data controller's country of establishment.[82] In other words, in principle it is not possible to derive obligations based on the Norwegian Personal Data Act unless the data controller is established in Norway.

In practice, this criterion presents a number of challenges that we will not delve into here. It is sufficient to mention that it essentially involves the actual performance of activities within a relatively fixed structure, and that this structure's legal status is not of decisive importance. It is, however, necessary to mention the *equipment criterion*. If the data controller is not established in the EEA area, the Act can nevertheless apply if the data controller makes use of equipment in Norway.[83] The Norwegian Personal Data Act may thus have an extraterritorial scope, for example in relation to an enterprise in the USA.

It is, however, very unclear what the equipment concept entails. For example, there have been suggestions that computer programs and cookies are a form of equipment in the sense of the directive. The question is interesting, but it has not yet been clarified.[84] If it does not take more to trigger the application of Norwegian rules than the use of cookies, this would at least possibly involve a dramatic expansion of the geographic scope of the Norwegian Personal Data Act. The provision concerning the geographic scope of the regulations, however, may possibly be changed dramatically, see below.

4.5.2 Export of data to third countries

If the data processing entails the transfer of personal data to a third country, the data controller must overcome certain legal impediments.[85] A (fictive) Norwegian chain of stores has, for example, collected and stored information on all the transactions with the chain's customers for the last 10

---

[81] Section 27 of the Personal Data Act stipulates that information that is *incorrect or incomplete, or for which processing is not permitted* shall be corrected if the data subject so demands. If there is no opportunity to process the data at all, such correction may be made by means of deletion, as is evident from Proposition no. 92 (1998-1999) to the Odelsting, see the comments on Section 27 in Chapter 16 of the proposition.

[82] Section 4, first paragraph of the Personal Data Act.

[83] Second paragraph of the provision.

[84] A similar question was raised before the European Court of Justice – the pending statement may be of decisive importance in this context. Case C-131/12: Reference for a preliminary ruling from Audiencia Nacional (Spain) lodged on 9 March 2012 — Google Spain, S.L., Google Inc. v Agencia Española de Protección de Datos, Mario Costeja González

[85] Sections 29 and 30 of the Personal Data Act.

years. The chain would like to analyse the information now. Since the chain does not have the necessary expertise itself to complete these analyses, the information will be sent to three different service providers in the USA, Germany and Israel, respectively.[86]

The information can be transferred freely to Germany, since there is free flow of personal data in the EEA area.[87] The information can also be transferred to Israel, since the country has been recognised as a safe data importer by the European Commission.[88] With regard to the USA, it must be investigated whether the supplier is Safe Harbour certified.[89] If so, the information can also be sent there without restriction.[90] If the data processor in the USA is not listed on the official list of Safe Harbour certified companies, the chain of stores must issue *guarantees*, which are to be approved by the Data Protection Authority, before the data can be transferred.[91] The surest way of obtaining such approval is to use the EU model contracts for the transfer of personal data.

4.6 New data protection rules

The Data Protection Directive, on which the Norwegian Act is based, is from 1995. Even though the Directive's objective and its general principles are still valid, globalisation and technological developments have changed the world. The European Commission therefore stated in 2010 that it was time to revise the data protection legislation in Europe.[92]

The Commission has prepared a draft general regulation on the processing of personal data.[93] The European Parliament has in turn published comments and an amendment proposal for the draft.[94] The draft regulation will become applicable law in Norway, since legislation in the data protection area is EEA relevant.[95] If it indeed does end up in the form of a regulation, the rules will be incorporated directly into Norwegian law.[96] The rules will also take precedence over other rules in Norwegian law that are not EEA relevant.[97]

There are four elements in particular in the proposed regulation that are of interest in connection with Big Data. We will take a closer look at these below.

---

[86] The three partners in the example are *data controllers*, cf. Section 15 of the Act.
[87] Cf. the Act's main rule in Section 29, first paragraph.
[88] Section 6-1, of the Personal Data Regulations, cf. Commission Decision 2011/61/EU.
[89] Commission Decision 2000/520/EC.
[90] https://safeharbor.export.gov/list.aspx
[91] Section 30, second paragraph of the Personal Data Act.
[92] COM(2010) 609 final.
[93] The current draft is dated 25 January 2012.
[94] The so-called Albrecht report: COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)
[95] Cf. Annex XI of the EEA Agreement.
[96] A number of member countries believe that the regulations should be introduced as a directive. The differences between regulations and directives are explained in Arnesen, Finn (2009).
[97] Cf. Section 2 of the EEA Act.

4.6.1 Consent and balancing of interests

The European Commission stated already in 2010 that it was a goal to strengthen and clarify the rules on consent.[98] Strengthening consent has also been a priority for Parliament, which is clearly evident from the proposed amendments to the wording of Article 7 of the proposed regulation.

If Parliament gains acceptance for its views, it will be more difficult to collect and use data without the consent of the data subjects. In practice, this will then entail that the scope of the balancing of interests is correspondingly narrowed. Parliament is thus also opposed to the European Commission's proposal that the Commission itself, through so-called "*delegated acts*",[99] should be able to establish decisive guidelines for the balancing of interests. Parliament proposes in the same breath the introduction of an extensive list of what interests can be regarded as legitimate in the text of the regulation. Whether one or more Big Data related interests are found on such a list remains to be seen.

4.6.2 Purpose limitation

The Commission has proposed to *narrow* the scope of purpose limitation. The data controllers' leeway will thus be *expanded* correspondingly. The Commission has proposed that it should no longer be necessary to obtain consent from data subjects in order to process personal data already collected for new and incompatible purposes. In the opinion of the Commission, it should instead be adequate for the information to be processed in accordance with at least one of the other legal grounds.[100] This means that in practice it will be *simpler* for those who process Big Data to overcome this threshold.

Parliament would for its part like to maintain the principle as it is today.[101] The Article 29 Group also wants this, and it points out that the proposal will entail such extensive exceptions to purpose limitation that the principle will lose its meaning in practice.[102]

If an adjustment of the purpose limitation principle as proposed by the Commission becomes a reality, this may make the daily life of Big Data users and analysts easier. On the other hand, if Parliament and the Article 29 Group gain acceptance for their objections, the prevailing law will remain unchanged.

4.6.3 Privacy by design and default settings

In Article 23, the Commission has proposed to codify the principles of data protection by design and privacy-friendly default settings under the heading "*Privacy by design and by default*". The general rule in the first paragraph of the proposed article is formulated as follows:

> *"Having regard to the state of the art and the cost of implementation, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures and*

---

[98] Cf. COM (2010) 609 final

[99] A form of legal instrument introduced by the Treaty of Lisbon; authority has been delegated to the Commission to prescribe rules concerning "*non-essential elements*" in pre-existing legislation, see Article 290 of the Treaty.

[100] The same legal grounds that are mentioned in Section 4.3.1 are involved here.

[101] COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)

[102] Opinion 03/2013 on purpose limitation.

procedures in such a way that the processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject."

The provision stipulates that the data controller is obliged to implement appropriate technical measures to safeguard the requirements prescribed by the regulation. This is not really anything new, at least not when seen with Norwegian eyes. Corresponding obligations can be derived from current regulations, from Section 14 of the Personal Data Act and the supplementary provisions in the third chapter of the Personal Data Regulations, among others.

However, it may possibly be considered an innovation that this obligation is made effective already from a point in time before the actual data processing has started ("*at the time of the determination of the means for processing")*. The latter thus entails a legal obligation to follow the *privacy by design* line of thought already at the preparation stage, before the processing of data starts.[103] It is an advantage to read the article in light of the basic requirements for the processing of personal data and the risk assessment rules (*Privacy Impact Assessment*) in Article 33 of the proposal.

The second paragraph of the article proposes the inclusion of a provision on *data protection by default,* or privacy-friendly default settings, which are to ensure that information is not processed beyond what is strictly necessary for the specified and legitimate processing purposes to which the data controller can make reference:

"The controller shall implement mechanisms for ensuring that, by default, only those personal data are processed which are necessary for each specific purpose of the processing and are especially not collected or retained beyond the minimum necessary for those purposes, both in terms of the amount of the data and the time of their storage. In particular, those mechanisms shall ensure that by default personal data are not made accessible to an indefinite number of individuals."

The collection and other processing of personal data should in other words be limited to a minimum through default settings, and the standard for this consists of what can be regarded as necessary in relation to the legitimate purposes on which the data processing in question is based. The aim of this provision is to make it easier to comply with the requirements stipulated by the regulations, in order to ensure protection of the individual's rights.[104]

4.6.4 Expansion of the data protection zone

An amendment proposal from the European Commission that is attracting a great deal of attention is Article 3 of the proposal. The Commission proposes to expand the geographic scope of the regulation in the second paragraph of the article at the same time as the article strengthens the country of establishment principle as the general rule with regard to the territorial scope[105] of the regulation. The Commission desires to give the regulation extraterritorial effect in connection with certain activities.

---

[103] Albrecht points out, moreover, that the point of intersection is the point in time for "determination of *the purposes* and means for processing".
[104] See the Albrecht Report's *Amendment 178*. COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)
[105] Cf. first paragraph of the article.

If companies outside the EU offer goods and services to individuals in the EU,[106] or collect personal data to monitor the behaviour of these same individuals, the companies must respect the European rules in this area. American companies that do not have any link to Europe in a company law sense may thus have to respect future European standards for data and privacy protection. Social network communities and other services that are offered for free on the Internet, which Norwegian citizens currently use at their own responsibility, may also be subject to common European data protection rules in the future.

5. Summary and recommendations

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.[107] 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. We are of the opinion, however, that protection of the privacy principles is more important than ever at a time when increasing amounts of information is collected about us. The principles constitute our guarantee that we will not be 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.

In spite of the fact that the use of Big Data raises several privacy challenges, this does not mean that the use of this form of analysis is not possible within current data protection legislation. We will briefly outline below the key prerequisites that should be present, and the measures that should be implemented in order for the use of Big Data to take place within prevailing law, and respect the privacy of individuals.

5.1 Consent still the point of departure

Obtaining valid consent from data subjects in connection with the use of personal data for analysis and profiling purposes is the best insurance against violating data protection legislation. The new European Data Protection Regulation also proposes restricting the opportunities for the processing of personal data on legal grounds other than consent.

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.[108] Some claim that the constant demand for consent on the

---

[106] Since the regulation is EEA-relevant, citizens in Norway and other EEA states will necessarily also trigger such legal consequences as mentioned in the article.

107 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

108 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

Internet paradoxically may result in poorer protection for the individuals. We can already see a development where enterprises ask for wide-ranging consents 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, however, is illegitimate. There is no opportunity to collect all types of information and process the information without restrictions, even if this is based on consent. The other mandatory basic requirements that are stipulated by the regulations – such as the relevance principle and the purpose limitation principle – must be respected.[109]

Enterprises that would like to use the collected data for purposes that are incompatible with the original purpose must request the consent of the data subjects. If it is not possible or desirable to request such consent, anonymisation of the data that are to be compiled and analysed is a practical alternative. The information will then not be regarded as personal data in the legal sense, and the processing will fall beyond the scope of the Act.

In some cases, the anonymisation of data may not be practically feasible or meaningful seen in the light of the purpose of the analysis. Various techniques for the de-identification of information may then possibly be used in order to limit the privacy disadvantages associated with (re)use of the information. If techniques for de-identification are used, the main requirements for the processing of information in the Act will, however, still apply. De-identification may nevertheless work as a compensatory measure that may have influence on the balancing of interests that are integrated in the principle of purpose limitation. Through de-identification, there may thus be a possibility that it becomes unnecessary to request new consent for the reuse of collected information for new purposes.

5.2 Procedures for robust anonymisation

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.

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

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

[109] The principles are expressed in Section 11, first paragraph, letter d of the Act and Article 6 (c) of the Directive, and Section 11, first paragraph, letter c and Article 6 (b), respectively.

consisting in data randomization and generalisation.[110] 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:

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

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.

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,[111] 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.

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.[112] This should also include open data.[113]

The Data Protection Authority 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.[114]

5.3 Access to profiles and algorithms

The right to information on, and access to, the processing of one's own personal data are important principles of privacy protection. Enterprises that use Big Data in an open and transparent manner will benefit from this in the form of increased trust among customers, users and society in general. Enterprises that use collected data for profiling and prediction analysis in a manner that is not very transparent risk offending people. This applies in particular if they violate the customer's expectations of how the data is used.

Consent must be informed in order for it to be valid. This entails that individuals should be given information on what data are collected, how they are processed, for what purposes they will be used

110 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.

111 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.

112 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

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

114 http://www.ukanon.net/

and whether the data will possibly be disclosed to third parties. In a Big Data context this also entails a right of access to one's own profile. To ensure the greatest possible transparency related to the use of Big Data, access should also be granted to the decision-making criteria (algorithms) the development of the profile is based upon. Individuals should also be given information on the sources from which the various personal data are obtained. In accordance with Norwegian data protection legislation, individuals currently have a right to disclosure of the rule content of algorithms on which automated decisions that are of significant importance to individuals are based[115]. This is to avoid unlawful discrimination and decisions of importance to individuals being made on an incorrect basis, among other things.

The author Evgeny Morozov (2013), argues in his book *"To save everything, click here",* that police algorithms, among others, should be subjected to public auditing. External actors should regularly audit police algorithms, so that the public is aware of what variables are included in the Big Data analysis. This will contribute to making decisions made on the basis of Big Data technology more transparent and thus more democratic. Such a right currently exists in Norwegian data protection legislation. However, cases that are decided or investigated pursuant to the so-called administration of justice acts are not directly encompassed by the Personal Data Act.[116]

Transparency and access to how the police use Big Data are important in light of the fact that the threshold for suspecting anyone of planning terrorist acts has been lowered. If the conditions surrounding such use are kept concealed and are not verifiable, it may at worst have a chilling effect on the freedom of expression.

With regard to the intelligence services' use of Big Data, this lies outside the scope of the Personal Data Act. In order to alleviate negative privacy consequences from the use of Big Data in these services, however, it is important that the decision-making criteria on which their data analyses are based, as well as the data sources that are included in them, are subject to democratic control. We would particularly like to point out the importance of the oversight body for the services, the EOS Committee, possessing the necessary expertise and insight into the consequences the use of large-scale data analysis may have on privacy and other important social values.

5.4 "Right of ownership" to one's own personal data

Big Data increases the economic imbalance between individuals on the one hand and large enterprises on the other hand. Personal data have become a very valuable commodity and component in the development of new services. It is the industry alone that extracts value from our personal data, not those of us who have provided the data.

There are different ways of remedying this imbalance. One solution is requiring the enterprises to give the data subject or customer access to all the data that the enterprise possesses in a user-friendly, portable and machine-readable format. This is referred to as data portability, and it will contribute to strengthening the control that individuals have over their personal data. It will make it easier to change from one service provider to another, so that one can choose the service that offers

[115] See Section 22 of the Personal Data Act
[116] This has been stipulated in Section 1-3 of the Personal Data Regulations.

the best terms and conditions, also for privacy concerns. Data portability will prevent customers from becoming locked into services that have unacceptable terms and conditions. In the long-term, such a requirement may contribute to forcing the provision of more privacy-friendly services. The proposed new European Data Protection Regulation includes data portability as a right.

Companies that offer so-called "data lockers", as mentioned earlier in this report, are another means of giving the user greater ownership of his own personal data. "Data lockers" give the customer an opportunity to benefit themselves from the resale and secondary use of their own personal data. The growth of such solutions shows that consideration for the users' privacy can be exploited as a business model and competitive advantage.

5.5 Privacy by Design and Accountability

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.

Use of Big Data technologies should be based on the seven principles of Privacy by Design.[117] Privacy by Design entails taking into account protection of privacy at all stages of system development, in procedures and in business practices.

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 technical access and security. Such an assessment should also carefully evaluate any potential consequences for the data subjects. [118,119]

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 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

117 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

118 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"

119 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 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)

regular controls to verify if the results from the profiling are responsible, fair and 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.[120]

Privacy impact assessments in connection with legislative work

The reuse of personal data may be desirable from a social perspective (Ministry of Government Administration, Reform and Church Affairs 2013). The potential that lies in Big Data for the realisation of efficiency gains in the public sector may result in a greater desire for the reuse of collected data. As mentioned earlier, reuse requires a new ground for processing. For the public sector, an act will often be the most appropriate ground for processing. Obtaining consent for new processing may be demanding, and in many cases it may also not be very appropriate. In establishing the reuse of personal data by law, it is important that the legislator make a thorough assessment of the privacy consequences in the legislation process. This is affirmed by including privacy as a separate item in the guidelines for the Official Studies and Reports Instructions .

A separate guide concerning the assessment of privacy consequences has also been prepared for the Instructions for Official Studies and Reports (Ministry of Government Administration, Reform and Church Affairs 2013). The Data Protection Authority has repeatedly pointed out the lack of privacy impact assessments prior to the distribution of draft acts and regulations for comments. In the Big Data era, with the opportunity to compile and analyse ever-larger data sets with personal data, it is more important than ever that this guide be followed.

5.6 Raising knowledge and awareness

Knowledge and awareness of the privacy challenges associated with Big Data are important among the enterprises that implement the technology. The Data Protection Authority urges 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 data protection 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.

120 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

List of references

Article 29 Data Protection Working Party – uttalelser og arbeidsdokumenter:

- Working document on a common interpretation of Article 26(1) of Directive 95/46/EC (WP 114)
- Opinion 4/2007 on the concept of personal data (WP 136)
- Opinion 5/2009 on online social networking (WP 163)
- Opinion 2/2010 on online behavioural advertising (WP 171)
- Opinion 8/2010 on applicable law (WP 179)
- Opinion 9/2011 on the revised Industry Proposal for a Privacy and Data Protection Impact Assessment Framework for RFID Applications (WP 180)
- Opinion 15/2011 on the definition of consent (WP 187)
- Opinion 04/2012 on Cookie Consent Exemption (WP 194)
- Opinion 05/2012 on Cloud Computing (WP 196)
- Opinion 03/2013 on purpose limitation (WP 203)
- Opinion 06/2013 on Open data and public sector information ('PSI') reuse (WP 207)

Blixrud, K. B. & Ottesen, C. A. (2010), "Personvern i finanssektoren", Gyldendal Akademisk, Oslo

Bollier, D. (2010), "The promise and perils of Big Data", The Aspen Institute, Washington DC, http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf

boyd, d. & Crawford, K. (2012), "Critical Questions for Big Data", *Information, Communication & Society 15:5, 662-679*, http://dx.doi.org/10.1080/1369118X.2012.678878

COM (2010) 609 final, "Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee og the Regions. A comprehensive approach on personal data protection in the European Union", *European Commission*

COM (2012) 11 final 2012/0011 (COD), "Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)", *European Commission*

COM (2012)0011 – C7-0025/2012 – 2012/0011(COD), "Draft report on the proposal for a regulation of the European Parliament and of the Council on the protection of individual with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)", *Committee on Civil Liberties, Justice and Home Affairs*, *European Parliament*

Commission Nationale de l'Information et des Libertés (2012), "Vie privée à l'horizon 2020", *Cahiers IP Innovation & Prospective N°01*, Paris

Dagens IT, (15.06.2013), "Facebook-data kan gi deg lån – eller avslag", http://www.dagensit.no/article2629493.ece, [downloaded: 10.09.2013]

Datatilsynet (2011a), "Social Network Services and Privacy – A case study of Facebook", www.datatilsynet.no/Global/english/11_00643_5_PartI_Rapport_Facebook_2011.pdf

Datatilsynet (2011b), "Hva vet appen om deg", http://www.datatilsynet.no/Global/04_veiledere/app_rapport_DT2011.pdf

Datatilsynet (2013), "7 steg til innebygd personvern", http://www.datatilsynet.no/Teknologi/Innebygd-personvern/

Datatilsynet & Teknologirådet (2013), "Personvern. Tilstand og trender 2013", http://www.datatilsynet.no/Global/04_veiledere/personvernrapport_tilstand_trender2013.pdf

Eckersley, P. (2010), "How Unique Is Your Web Browser?", *Electronic Frontier Foundation*, https://panopticlick.eff.org/browser-uniqueness.pdf, [downloaded: 10.09.2013]

The Economist, (02.06.2012), "Very personal finance, Marketing information offers insurers another way to analyze risk", http://www.economist.com/node/21556263, [downloaded: 15.08.2013]

Federal Trade Commission, (18.12.2012), "FTC to Study Data Broker Industry's Collection and Use of Consumer Data", http://www.ftc.gov/opa/2012/12/databrokers.shtm, [downloaded: 10.09.2013]

Fornyings-, administrasjons- og kirkedepartementet (2008), "Vurdering av personvernkonsekvenser. Veileder til utredningsinstruksen", Oslo

Fornyings-, administrasjon og kyrkjedepartementet (2013), "Personvern – utsikter og utfordringar", St.meld. nr. 11 (2012-2013), Oslo, Fornyings-, administrasjon og kyrkjedepartementet

Forbes, (25.06.2013), "Finally You'll Get To See The Secret Consumer Dossier They Have On You", http://www.forbes.com/sites/adamtanner/2013/06/25/finally-youll-get-to-see-the-secret-consumer-dossier-they-have-on-you/, [downloaded: 10.09.2013]

*The Guar*dian, (07.06.2013), "PRISM scandal: tech giants flatly deny allowing NSA direct access to servers", http://www.guardian.co.uk/world/2013/jun/07/prism-tech-giants-shock-nsa-data-mining, [downloaded: 15.08.2013]

Gymrek, M. , McGuire, A. L., Golan, D., Halperin, E. og Erlich, Y. (2013), "Identifying Personal Genomes by Surname Inference", *Science 18 January 2013: 339 (6117), 321-324*, [DOI:10.1126/science.1229566]

Hewlett-Packard (2013), "Metropolitan Police leverage social media to engage in local community", http://www8.hp.com/h20195/v2/GetDocument.aspx?docname=4AA4-5393EEW,[downloaded: 10.09.2013]

HealthWorks Collective, (26.02.2013), "Big Data in Healthcare – Hype or Reality?", http://healthworkscollective.com/shahidshah/85441/guest-article-try-not-fall-Big-data-healthcare-hype-focus-actionable-data, [downloaded: 10.09.2013]

Hildebrandt, M. (2009), "Who is profiling who? Invisible Visibility", i *Reinventing Data Protection?*, red: Gutwirth, S., Poullet, Y., De Hert, P., de Terwangne, C. og Nouwt, S., Springer

HSPS News, (11.10.2012), "Using cell phone data to curb the spread of malaria", *Harvard school of public health*, http://www.hsph.harvard.edu/news/press-releases/cell-phone-data-malaria/, [downloaded: 10.09.2013]

Hordern, V. (2013), "Consent – the silver bullet?", *Data Protection Ireland* (DPI 6 1 (13)

IBM (2013), "Connect to millions of devices and sensors with event-driven, near-real-time communications", *IBM MessageSight*, http://public.dhe.ibm.com/common/ssi/ecm/en/wsd14115usen/WSD14115USEN.PDF, [11.09.2013]

Information Commissioner's Office (2012), "Anonymisation: managing data protection risk code of practice", http://www.ico.org.uk/Global/~/media/documents/library/Data_Protection/Practical_application/anonymisation_code.ashx

Justis- og politidepartementet (1997), "Et bedre personvern – forslag til lov om behandling av personopplysninger", NOU 1997:19, Oslo, Statens forvaltningstjeneste

Lanier, J. (2013), "Who owns the future?", Simon & Schuster, New York

Mayer, J. (11.10.2011), "Tracking the Trackers: Where everybody knows your username", *The Center for Internet and Society, Stanford Law School*, http://cyberlaw.stanford.edu/node/6740, [downloaded: 10.09.2013]

Mayer-Schönberger, V. (2009), "Delete: The Virtue of Forgetting in the Digital Age", Princeton University Press

Mayer-Schönberger, V. & Cukier, K. (2013), "Big Data. A Revolution That Will Transform How We Live, Work and Think", John Murray, London

McKinsey Globale 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

MIT Technology Review (2013a),"Big Data Gets Personal", *Business Report*

MIT Technology Review, (30.07.2013b),"If Facebook Can Profit from Your Data, Why Can't You?", <http://www.technologyreview.com/news/517356/if-facebook-can-profit-from-your-data-why-cant-you/>, [downloaded: 10.09.2013].

Morozov, E. (2013), "To save everything click here. Technology, solutionism and the urge to fix problems that don't exist", Penguine Books, London

Narayanan, A. og Shmatikov, V. (2008), "Robust De-anonymization of Large Datasets. (How to Break Anonymity of the Netflix Prize Dataset)", http://arxiv.org/pdf/cs/0610105v2.pdf, [downloaded: 10.09.2013]

The New York Times, (28.11.2012a), "Jeff Hawkins Develops a Brainy Big Data Company", http://bits.blogs.nytimes.com/2012/11/28/jeff-hawkins-develops-a-brainy-Big-data-company/, [downloaded: 10.09.2013]

The New York Times, (16.12.2012b), "You for Sale. Mapping, and Sharing, the Consumer Genome", http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html?pagewanted=all, [10.09.2013]

The New York Times (16.12.2012c), "How Companies Learn Your Secret", http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all [downloaded: 10.09.2013]

The New York Times (08.06.2013a), "How the U.S. Uses Technology to Mine More Data More Quickly", http://www.nytimes.com/2013/06/09/us/revelations-give-look-at-spy-agencys-wider-reach.html?_r=0 [downloaded: 10.09.2013]

Nikulainen, T. (2013), "Big Data Revolution – What is it?", *ETLA Brief No 10.* http://pub.etla.fi/ETLA-Muistio-Brief-10.pdf

Nordbeck, P. & Lundqvist, D. S. (2012), "Data som skapas i molnet – hur långt sträcker sig personuppgiftsansvaret?", *Lov&Data nr. 110*

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

Ot.prp. nr. 92 (1998-1999) om lov om behandling av personopplysninger

Pariser, E. (2011), "The Filter Bubble. What the Internet is Hiding from You", Penguin Books, London.

Personvernnemnda, avgjørelser:
- PVN-2004-1
- PVN-2011-10
- PVN-2012-1

Reding, V. (14.06.2013), "PRISM scandal: Vice-President Reding makes it clear the data protection rights of EU citizens are non-negotiable", *European Commission*, http://ec.europa.eu/commission_2010-2014/reding/multimedia/news/2013/06/20130612_en.htm [downloaded: 10.09.2013].

Rubinstein, I. S. (2012), "Big Data: The End of Privacy or a New Beginning?", *Public law & legal theory research paper series, Working paper NO. 12-56*, New York University School of Law

Schartum, D. W. & Bygrave, L. A. (2006), "Utredning av behov for endringer i personopplysningsloven", Rapport 2006, Justisdepartementet og Moderniseringsdepartementet.

Tatonetti N. P.,Denny J. C., Murphy, S. N.,Fernald G. H., Krishnan, G., Castro, V., Yue, P., Tsau P. S., Kohane, I., Roden, D.M. & Altman, R. B. (2011), "Detecting Drug Interactions From Adverse-Event Reports: Interaction Between Paroxetine and Pravastatin Increases Blood Glucose Levels", *Clinical Pharmacology & Therapeutics*, 90:1, 133–142, doi:10.1038/clpt.2011.83, http://www.nature.com/clpt/journal/v90/n1/abs/clpt201183a.html

Tene, O. & Polonetsky, J. (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

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

VINT research report (2013), “Creating clarity with Big Data”, http://www.sogeti.se/upload/SV/Kalendarium/Dokument/Big-data1.pdf

Whelan, A. (3.11.2012), “Big Data – The Digital Agenda for Europe and Challenges for 2012”, *The Institute of International European Affairs*, http://www.iiea.com/events/Big-data--the-digital-agenda-for-europe-and-challenges-for-2012 [downloaded: 10.09.2013]

World Economic Forum (2013), “Unlocking the value of personal data: From collection to usage”, Prepared in collaboration with The Boston Consulting Group

Data Protection Authority

Street address: Tollbugata 3, Oslo

Postal address: P.O. Box 8177 Dep, 0034 Oslo

Email: postkasse@datatilsynet.no

Telephone: +47 22 39 69 00

Fax: +47 22 42 23 50

www.datatilsynet.no
www.personvernbloggen.no