Online Journal of Applied Knowledge Management
A Publication of the International Institute for Applied Knowledge Management

Volume 2, Issue 2, 2014

Big data: Three-aspect approach

Jędrzej Wieczorkowski, Warsaw School of Economics, jedrzej.wieczorkowski@sgh.waw.pl

Przemysław Polak, Warsaw School of Economics, ppolak@sgh.waw.pl

Abstract

The term big data is a new, shaping up concept. The paper proposes a three-faceted explanation of the term, by distinguishing three basic aspects of big data: technological (including the opportunities offered by IT and modern analytical methods), business (including a variety of applications of the concept) and social (focusing on the consequences of its implementation). The authors also indicate some other aspects, e.g. legal, which can be used to extend the classification. The social aspect, mainly concerning the risks associated with the mass processing of personal data and privacy, is usually overlooked when trying to define the concept. It is dependent in relation to the other aspects, but our findings support the conclusion that it is the closest to the understanding of the concept big data in common opinion. The paper contains the attempt at a comprehensive multi-faceted analysis of the big data phenomenon.

Keywords: big data, data analysis, personal data, privacy.

Introduction - Understanding big data

The concept of big data is currently used in very different contexts. It concerns not only the information technology, quantitative methods and their applications, but also social issues. Although the term is almost new, its popularity is demonstrated by the fact that it is eagerly used and discussed at least at three levels: academic, business, and political. The purpose of this article is to categorize the concept of big data in its various aspects, with particular emphasis on social issues.

The main challenge is an attempt to understand the phenomenon of big data. To what extent is it a new phenomenon and to what extent is it the development of previously known concepts, particularly business intelligence, and to what extent is it maybe just another marketing slogan? Currently, there is no single, universally accepted definition of big data. Perhaps the concept is so new, multi-faceted and expanding, that it cannot be clearly defined (yet?). Traditionally, when trying to describe big data, the concept of "several v" is used. It is based on features describing the term and beginning with the letter 'v'. There are at least three basic 'v's: *volume*, *velocity*, *variety*. However, other features are often also added, most commonly: *value* and *veracity*.

META Group (Laney 2001), a consulting company acquired in subsequent years by Gartner Inc., published a report which characterized the impact of e-commerce, globalization and other trends in the economy in the development of information technology. Projected changes in IT were presented in the 3V model in three dimensions. The term big data was not used then, but the forecast described very well the development of this phenomenon, which was to occur several years later.

According to the 3V model (see Figure 1), big data includes the following characteristics:

- Volume – Very large volumes of processed data. It is associated primarily with the ever-increasing growth of the quantity of digital data available in the form that can potentially be used.
- Velocity – Variability and dynamics of processed data. This concerns in particular the issues of data rapidly changing and generated in a very short period of time, in particular data which should be used shortly after creation or update. This feature is typical for sensory and streaming data, as well as data related to the analysis of behavior in the Internet, including social networks. The analysis of such data is necessary in near real time.
- Variety – A variety of processed data. This feature may be associated with a very large number of various types of structured data attributes in a relational database. Above all, however, it applies to a variety of unstructured data, such as images, video footage, audio, handwritten notes, and other data unsuitable for storage in a classical relational database.

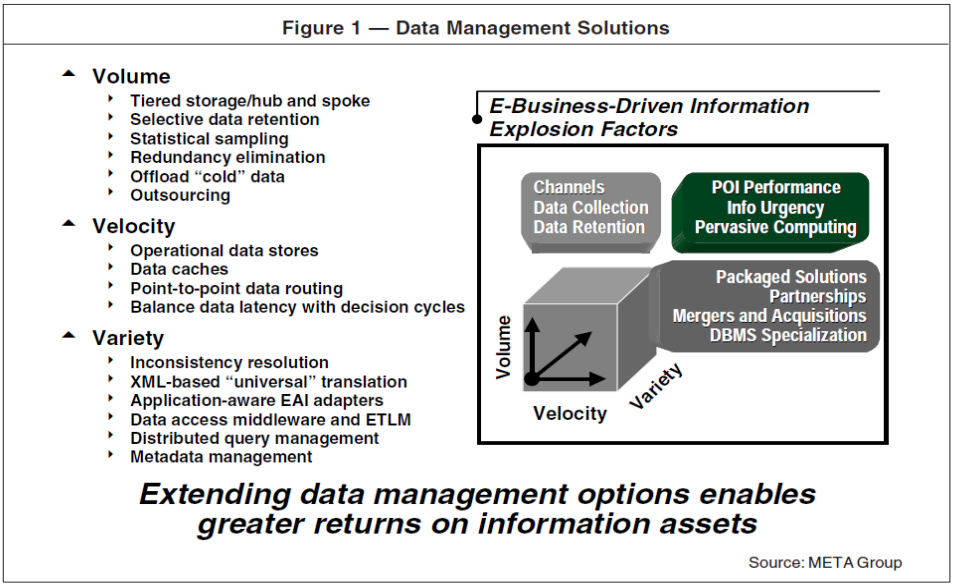

Figure 1. The 3V META Group model (Laney 2001)

Other features, outside of the classical 3V model include:

- Value – A large value of the processed data. This feature refers to data usability and the practical possibilities of their use in decision-making and, consequently, to the possibility of generating significant value for an organization.
- Veracity – Reliability of the data being processed. It covers data quality issues, in particular issues such as their incompleteness, errors, noise and other characteristics affecting the quality, which are especially important when processing large amounts of data from diverse sources.

The definition based on the choice of words beginning with a single letter of the alphabet is probably largely a marketing trick, but in practice this is how the importance of big data is explained today. Despite the problems with many aspects of the new concept, there are various definitions of big data formulated by scientists, analytical companies and online encyclopedia authors.

For example, the McKinsey Global Institute (2011) referred to the concept of big data as data sets whose sizes exceed the capacity of conventional database tools for gathering, storing, managing and analyzing data. It should be noted that this is a typical definition, often repeated in similar forms.

According to PcMag encyclopedia (2013) "Big Data refers to the massive amounts of data that collect over time that are difficult to analyze and handle using common database management tools. Big Data includes business transactions, e-mail messages, photos, surveillance videos and activity logs (see machine-generated data). Scientific data from sensors can reach mammoth proportions over time, and Big Data also includes unstructured text posted on the Web, such as blogs and social media."

Boyd and Crawford (2012) went beyond the typical approach and defined big data as a cultural, technological, and scholarly phenomenon that rests on the interaction of technology, analysis and mythology. While the aspects of technology and analysis are understandable, the distinction of mythology aspect is original and interesting. It is defined as "the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy".

Due to the commercial value of the concept, companies offering solutions to support big data try to clarify the term to potential customers. Since the creation of the concept is related in a large extent to the development of technological capabilities, it seems particularly important how leading IT companies understand this concept. Their approach may show the likely big data development directions in the near future. In explaining the phenomenon, they usually rely on the "several v" model, often adding other important comments, for example:

- Big data indicates a tendency to explore and use the business value dormant in the available growing volumes of data (SAS 2013a). It describes the exponential growth and availability of data, both structured and unstructured (SAS 2013b).
- Big data refers to the world where 2.5 quintillion bytes of data are created every day, and as a result, 90% of the data available today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few (IBM 2013).
- Big data is the term increasingly used to describe the process of applying serious computing power – the latest in machine learning and artificial intelligence – to seriously massive and often highly complex sets of information (Microsoft 2013).
- The concept of big data refers to the number of basic groups of data, such as typical enterprise data (coming, for example, from ERP and CRM systems), data collected automatically (e.g. sensory data), data from the internet and social media (Oracle 2013).

The analysis of the above listed definitions suggests that the concept of big data is primarily associated with information technology and analytical methods. This approach is fully justified in the case of big data solutions providers who emphasize the power of modern technology and analytical methods as well as extensive possibilities of application in business. However, this point of view is partly inconsistent with another approach which stresses "soft" aspects of big data, in particular the social aspect of big data. A specific feature of this concept is that it found practical applications in many different areas of life very quickly, as well as it appeared extremely swiftly in colloquial language. Since the concept has become very popular in the media, the question arises whether the common understanding of the term is consistent with the formal description and definitions.

Three aspects of big data

The analysis of the big data phenomenon and both the professional and general public discussions of the topic, led the authors to the distinction of three basic aspects of big data:

- Technological (information technology and analytical methods),
- Business,
- Social.

Purely technology issues were combined with quantitative analytical methods in one aspect of the technology, as they are in practice very closely related to each other. Perhaps such a connection is controversial, because theoretically it is possible to separate those issues without problems. However, the authors are convinced that it is not advisable due to further unnecessary complications of studying relationships between different aspects. The technological aspect should be considered as primary. In a way, the technological aspect arises in a response to demand (mainly business), but the needs which cannot be met are purely theoretical issues.

The business aspect can be treated as an actual application of the concept of big data. The application is possible in various fields, but at the moment the business area seems to provide the most opportunities.

The social aspect is dependent on the other aspects. It covers issues arising from the consequences of big data applications, mainly threats to privacy. It seems the closest to the understanding of the concept by common opinion and therefore it is so important.

In the proposed model, the aspects of big data can also be described as follows:

- Technological aspect – oriented on big data opportunities,
- Business aspect – oriented on big data applications,
- Social aspect – oriented on the consequences of big data.

Aspects of big data in popular media

The authors conducted a study on the usage of the term big data in popular media. In March 2014, the authors searched for the 10 most popular articles (according to the search algorithm) containing the phrase "big data" in the most popular Polish internet news portal (www.gazeta.pl) belonging to the publishing group Agora. The portal contains news and articles from all sources belonging to the Agora. Search results included articles that had appeared mostly in a printed form from the previous few months. The majority had been published in the newspaper Gazeta Wyborcza, belonging to the quality press category. Gazeta Wyborcza has extensive general news and business sections as well as quite significant science, technology and reportage sections.

During the study, the portal search engine was used to find the highest-ranked results for a phrase "big data". Some findings were discarded, if the phrase appeared accidentally or if the article was essentially about different topics, and if the term was only mentioned in passing (e.g. reports from IT exhibitions, labor market analyses). The content of the remaining ten highest-ranked items was thoroughly examined. Four texts were from the technology section, three from the business section, two were reportages and one from the general news section.

Within the content of every text, one or more aspects of big data was distinguished. One article covered all three aspects. Two texts included the coverage of two nearly equally important aspects. In remaining seven articles, one aspect was dominant (see Table 1).

No	Section	Date of publication	Technological aspect	Business aspect	Social aspect
1	business	Dec 2013		X	X
2	technology	Jun 2013		X	
3	reportage	Apr 2012		X	X
4	business	May 2013			X
5	business	May 2013			X
6	technology	Apr 2013	X	X	X
7	news	May 2013			X
8	technology	Mar 2014			X
9	technology	Feb 2014		X	
10	reportage	Jan 2014			X

Table 1. Big data aspects covered by articles in gazeta.pl portal.

Only in one case, was the article associated with the technological aspect, but even then it was also attributed to the other two aspects, because the article contained a general review of the topic. In five cases the business aspect was assigned, but in three of these it was in combination with the social aspect. Two articles contained only the business aspect. They were focused on specific possible application of big data: predicting the future based on historical data analysis and medical applications. Definitely the most assignments, a total of eight, concerned the social aspect. In the sample, particularly popular was the topic of collecting personal data by financial

institutions, which was covered by 4 texts. It probably resulted from the widely discussed TV interview with the vice-president of Alior Bank, one of medium-sized banks in Poland (the TVN-CNBC 2013). He talked frankly about the possibility of using detailed customer data. Generally, the analyzed texts were focused mainly on social aspects: the danger of surveillance and the threats to privacy resulting from the analysis of web content, including social networks, by the private sector.

It should be taken into account that the portal is universal, therefore scientific and technical issues represent a relatively small part of the information content, but, on the other hand, the business section is a good place to discuss potential applications and is quite extensive. However, the survey results indicate that the professional definition of big data is significantly different from the colloquial understanding of the concept. Furthermore, the topics related to big data and the interest from journalists and are the public consider mostly social issues. It can be assumed that these results stem from the fact that many journalists and even scientists try to follow the current topics, but, at the same time, are not competent to deal with technological issues. Thus, they discuss the impact of technological and economic changes on society. This fits well into very popular mainstream issues of the information society. Nevertheless, even if this is the cause of social interest in issues arising from big data, they play an important role in the current understanding of the concept.

Technological aspect of big data

The primary issue of the technological aspect is what new opportunities are offered by big data, compared to the previously used solutions. Analyzing the differences between big data and earlier concepts and methods, for example, business intelligence and data mining, according to the authors, special attention should be paid to new real-time (or near real-time) processing capabilities, as well as the capabilities of processing poorly structured data. The mere fact of increasing the amounts of processed data can be considered as an evolutionary trend as well as the effect of technological progress and increasing data set size. Therefore, there is no point in indicating any specific size limit of data sets when defining the concept of big data.

Very large volumes of data have so far been also used in both On-Line Transaction Processing (OLTP) and On-Line Analytical Processing (OLAP). In the latter case, closer to the concept of big data, basic and popular approach is to build a well-structured data warehouse, typically containing aggregated data, separated from transaction processing, periodically updated in ETL (extraction, transformation, loading) process. Such an approach, however, in practice, prevents real-time analysis, and often the use of detailed source data. In addition, a potential source of data in big data processing can also be archived using data that has not been entered into the data warehouse because of the high costs.

In terms of quantitative analytical methods, new tools for advanced data analysis have been developed primarily on the basis of statistics, artificial intelligence, machine learning and the concept of data mining. These methods also process large volumes of data, however, in comparison with big data, such analyzes are often based on the sample data, what in many applications is not sufficient. However, according to the authors, the methods of data analysis have not been developed in recent times as fast as information technology. In contrast, primarily

due to higher calculation speed it became possible, in a satisfactory time, to perform analysis of large volumes of less structured data, for example, text data from the Internet (text mining).

In terms of information technology practical implementation, big data requires the use of computers with more computing power than before, in particular the so-called supercomputers, High-Performance Computing (HPC) approach, in-memory processing and processing directly by the database engine (in-database). Big data involves also distributed processing, including: parallel processing performed by a network of computers (grid computing), cloud computing, and MapReduce paradigm frequently with Apache Hadoop solution. Also new approaches to databases are involved, e.g. NoSQL and column-oriented database management systems.

Based on technological capabilities and analytical methods, software tool developers offer an increasing number of tools for data analysis. The introduction to this article presents the examples of big data definitions and descriptions by leading software producers. They utilize increasing hardware capabilities and develop new concepts, such as in-memory processing, in-database processing and often implement Apache Hadoop solution. The market analysis of big data predictive solutions was developed by Gualteri (2013) from the Forrester consulting company. The classification was based on the current offer and development strategy (see Figure 2). It is worth noting that the leading providers of these tools belong to the leading companies in IT sector, for example: IBM, SAS, SAP and Oracle. It suggests that big data has become the part of the mainstream IT development.

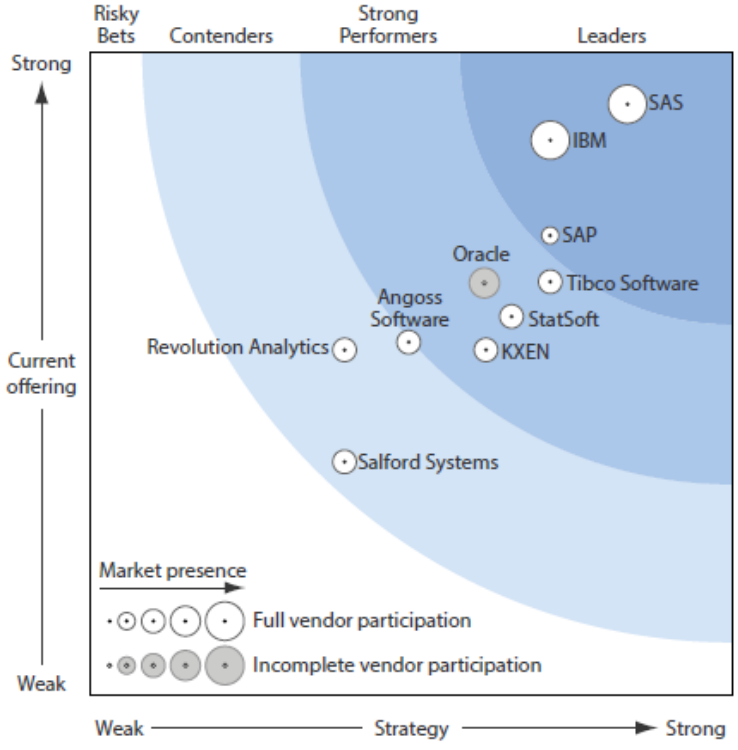

Figure 2. The market analysis of big data predictive solutions (Gualteri 2013)

In summary, according to the authors, the term big data in the aspect of technology should be regarded as the description of new trends in IT and analysis methods, with particular emphasis on analyzes conducted in real time, directly on the data source, often involving unstructured data.

Business aspect of big data

New technological capabilities, as well as a sharp decline in prices of access to advanced IT infrastructure, led to new applications of data analysis, particularly in business. However, some of those applications were developed before the emergence of the big data concept. For example, in the field of commercial applications, data mining enabled conducting many similar analyses. This method has been and is still widely used, however, new data analysis capabilities in real time and processing all available data sets, rather than their samples, significantly expand the scope of potential applications. The new method does not require at the beginning constructing theses and a model, but allows to search directly source data for previously unknown correlations. So far, organizations designing data structures, in particular implementing data warehouses, had to evaluate in advance which data were important. The excess of processed data could lead to the lack of processing power, or at least to the significant increase of cost and processing time. The model used in big data is based on the assumption of nearly costless data collection and processing. Dramatically falling cost of collecting and processing data means that it is possible to solve computational tasks that, not so long ago, were not economically justified, and even seemed impossible.

The development of big data can be considered a solution to the problem of information overload in analytical processing described by Wieczorkowski and Dałek (2013). Information overload appears when an organization has access to sufficient quantity of essential data, but has no possibility of processing and, consequently, concluding. Thus, an organization may have an excess of accurate information, but useless in practice due to the lack of processing capabilities.

There is a common opinion that the concept of big data is an extension of the idea of business intelligence. Therefore, it seems to be natural to have the applications of big data in the same commercial sectors like business intelligence, e.g. telecommunication, banking, insurance, retail, advertising. A good example is the application data mining methods in solving problems such as risk modeling, fraud detection, prediction of customer churn, load forecasting of power systems, technical analysis of securities, optimization of logistics chains (Szupiluk 2013). Davenport and Dyche (2013) distinguished future general objectives for big data in the commercial sector: cost reduction, reducing the cycle time for analytical calculations, developing new product and service offerings based on data, supporting internal business decisions.

Of particular interest is the use of big data in other areas, such as the public sector, which is so far slightly off the beaten track of basic business intelligence applications. Public administration almost always is associated with the use of unusually large data sets, in particular this applies to official statistics and public records. For example, the tasks of national statistical offices clearly involve processing very large amounts of data used, inter alia, for the purpose of decision support at the level of state governance. It is difficult at the moment to name the current methods used by the public statistics as big data, primarily due to the long delay in distributing the output data in relation to the periods to which they refer, and also because of the well-structured nature of the data. However, some elements of the evolution of public statistics in the direction of real-

time processing can be observed, hence the big data seems to be the quite natural direction for its further evolution.

Public statistics use public records, for example, a register of citizens. In Poland, this role is played by Universal Electronic System for Registration of Population (PESEL, in Polish: Powszechny Elektroniczny System Ewidencji Ludności). Its aim is to support information processes in the sphere of civil registration. Other public registers in Poland include: National Official Register of Economic Entities (REGON), Land and Property Register, the National Court Register (KRS), Central Register of Entities for National Taxpayers Records (CRP KEP). The question arises whether the current use of such reference identification systems can be considered as big data processing. The size of data sets seems to confirm that, but data set structures and processing methods are not typical for big data. Typical processing within those systems belongs to traditional transaction processing, especially batch processing and OLTP. These registers are usually well structured and stored in databases in tables. Their design and methods of use are specific to the registration systems that are not suitable for advanced analysis. However, new storage and processing methods should be developed in the near future, in order to enable various analyzes by information systems of public institutions.

Based on reports by TechAmerica Foundation (2011) and McKinsey Global Institute (2012), which showed examples of specific applications, as well as our own observations, according to the authors, some potential general group of big data applications can be distinguished in the public sector:

- Detection of various types of abuse, including financial. This applies in particular to solutions aimed at improving the efficiency of the tax and customs services (e.g. the control of the shadow economy, in particular through the analysis of e-commerce, discovery of assets not reflected in reported income), restricting fraud benefits (e.g. false sickness and unemployment), detection of unauthorized attempts to influence the securities market.
- Public Safety. This applies especially to new opportunities for police and special services in the infiltration of environments that may pose a threat to the state and public order. In particular, this is related to the Internet tracking, as well as the use of unstructured sensory data, for example, images from monitoring systems.
- Providing information for state management: the whole economy, as well as solving specific, detailed, often local problems. This includes statistical data used in near real-time.
- Management of various public services. For example, in the field of health care, there are various possible applications: diagnosing using a broad base of historical data, monitoring epidemiological risks, management of the entire health care system for the efficient use of public funds. Data analysis of communication streams (including monitoring and geolocation data) supports the planning of road networks and public transport management. Other possibilities include management of education, social welfare and vocational activation, all of them using mostly demographic data originating, inter alia, from public civil records.

Potential applications of big data in the public sector can be based partly on the processing of the described above public records. For example, all analyzes of business organizations (e.g., tax avoidance or hidden economy analyses) require access to REGON and CRP KEP registers. Whereas, all analyses concerning citizens (e.g. fraud detection and ensuring public safety) require access to PESEL register. However, data from the traditional public records and official statistics are not the only source of information for big data analysis, they require, as well, special repositories collecting data gathered purposely for those analyses, for example from surveillance systems.

Accessing data, particularly personal data, both by public and commercial organizations arises also legal questions. For example, in Poland, even the access by the public authorities to detailed confidential personal data is a subject to complex formal regulations. In practice, the technology allows for a lot more than the legal system. Regulations, in particular complex procedures to access data, cannot keep up with the changing capabilities of modern IT. However, they can be universal enough in order to limit to some extent the potential, but not yet known possibility of future abuse. Less restrictive rules apply to the processing of data that does not contain personal information, such as anonymous detailed data (used for example to monitor epidemiological risks) and aggregated data (for example used for state management).

When considering the business aspect of big data, apart from the issue of improving business methods, economic issues at the macro level should be taken into account. New methods of processing mass data provide opportunities for industries involved in providing access services to IT infrastructure (e.g. the cloud model), and consulting services based on the analysis of large data volumes. Analyses may relate to, among others, personal data (including depersonalized data), but this issue, like the other previously mentioned legal issues, is at the crossroads of the business aspect and the social aspect discussed below.

In summary, according to the authors, the term big data in business aspect should be described as possible applications of technology and quantitative methods in the processing of bulk data for analytical purposes, both by business organizations and public administration. These capabilities are closely associated with a legal system.

Social aspect of big data

The social aspect focuses on issues of processing and using personal data. That is, for example, according to Polish Act on the Protection of Personal Data (Ustawa o ochronie danych osobowych 1997), data related to an identified or identifiable individual. The person identified is an individual who can be identified, directly or indirectly, in particular by reference to an identification number or one or more attribute specific to his physical, physiological, mental, economic, cultural or social characteristics.

A specific type of personal data constitutes data derived from the internet, which can be an almost inexhaustible source of users' data, for example information about potential consumers. Banks and loan companies try to analyze social networking sites in order to create client profiles and to better assess their creditworthiness. Insurance companies can, potentially, by knowing more about the customer, better estimate individual risk. The ads displayed on web pages can be based on the data profiled in real time.

There are also other sources of personal data from various, also commercial systems, for example: data from the telecoms (in the case of a mobile systems also location data), data on private financial transactions (credit card payment, bank account transactions, loans), data on purchases made as the part of a loyalty program, video data from monitoring, medical data, education data from schools and universities.

The main problem of personal data processing is ensuring privacy. The above-mentioned data sources, thanks to big data technologies, may have great value for the analyzes carried out for different purposes, both commercial and public. However, there is a risk of breaching the barrier of privacy. Hence, the legal protection of personal data, including sensitive data in particular. Laws concerning the processing of such data may regulate the principles of storing them on servers outside the country. In the case of distributed processing, typical for the concept of big data, this can be a major obstacle. Thus, the social aspect of big data is an issue at the crossroads of human rights, economic freedom, public safety, and the legal system.

In the case of numerous analyzes anonymization methods are used. However, there are various publications which, symbolically, but not very seriously, compare big data to Orwell's Big Brother, for example:

- According to the dictionary Urbandictionary (2013) big data is "a modern day version of Big Brother. Online searches, store purchases, Facebook posts, Tweets or Foursquare check-ins, cell phone usage, etc. is creating a flood of data that, when organized and categorized and analyzed, reveals trends and habits about ourselves and society at large".
- Simon (2013) noted that Big Brother didn't just punish innocent citizens for independent thinking as thought-crimes, it also caught legitimate criminals as well. The same story is possible with big data. On the one hand, it creates a threat of total surveillance, but may, as well, counteract the actual crimes, e.g. extortion in the insurance sector.
- In 2007, the British media noted that within 200 yards of the London flat where G. Orwell wrote "1984" were more than 30 surveillance cameras (Mayer-Schonberger and Cukier 2013).

On the one hand, the law should provide protection of privacy, on the other hand, the excessive protection of data blocks the development and usage of modern technology, what can bring a negative impact on economic development. In the U.S., there are ongoing debates whether information that does not violate privacy and security and has been collected by public institutions should be widely available and considered as a common good. Whether and to what extent, the law should allow re-use of what was created for public money (the Economist 2013). Generally, the desired situation is when individuals can protect their personal information while not completely restricting data analysis.

The problem is seen similarly by the public administration in Poland. In October 2013, the minister of administration and digitization spoke in the context of big data about maintaining a balance between business opportunities and ways of expressing consent to data processing. He claimed that the ministry is not opposed to profiling, but certain parameters or conditions must be met under which profiling can take place (MAC 2013). In another statement, in the same period of time, he argued that for some time a completely new phenomenon is observed: new

awareness of privacy issues is increasing, as fast as the importance of big data is growing and creating added value for the whole economy (MAC 2013b).

Another situation involves the use of big data by the police, special forces and intelligence agencies. It is applied to the infiltration of entire groups which may present danger, and to detecting violations of law by the individuals. These opportunities include, for example, tracking the internet, and collecting monitoring data. The analysis of the monitoring conducted by Stępniak (2013) emphasizes the importance of not only the direct observation of the people but also the vehicle based on identification number plates.

The use of big data is aimed at improving safety, but there is an open question of ethical issues and the extent of surveillance allowed by law. The disclosure of PRISM program showed that the concept of big data had been used extensively in the U.S. in the field of security for several years. The authors searched for sources to answer the question about analogous situation in other countries, particularly in Poland. The report of the Panoptykon Foundation (2013) analyzed, on the basis of four major Polish companies providing services electronically, the scale of official requests for user data sent from different state organs. There is no information available about the detailed nature of these inquiries. The number of such inquiries in the period from January 2012 to June 2013 grew steadily, but did not exceed a few thousands a year (Jan-Jun 2012: 2479 requests, Jul-Dec 2012: 4375 requests, Jan-Jun 2013: 4505 requests). Absolute numbers are in this case meaningless because they concern only a small number of surveyed companies, while the statistics of authorities requesting queries is worthy of note (see Table 2). It shows the dominance of requests from the public prosecutor's office (62.7%) and the police (33.2%), but a significantly smaller share of various special services (total 2.3%). This suggests that the queries mainly concerned specific pending criminal cases. This is confirmed by public information obtained directly from the forces. Assuming the reliability of the data, and assuming that the state does not use illegal methods for accessing data, it can be believed that in Poland the queries from authorities have nothing in common with mass surveillance and the concept of big data.

Public authority	**Percentage of queries**
Prosecutor's office (incl. military persecutor)	62.7%
Police	33.2%
Other special services and agencies	2.3%
Courts in criminal proceedings	1.3%
Other entities, including courts in civil proceedings	0.5%

Table 2. The structure of queries for companies providing services electronically by inquiring public authorities (Panoptykon Foundation 2013).

The report discussed above was focused on controlling data obtained from the Internet, whereas, the highest supervisory authority in Poland, The Supreme Audit Office (NIK, in Polish: Najwyższa Izba Kontroli) published a report (NIK 2013) on the acquisition and processing

telecommunication data (including phone call records and location data) by the authorized state bodies between January 2011 to June 2012. Despite the generally positive assessment of compliance with the law, in the opinion of the NIK current regulations on obtaining telecommunication data by authorized entities does not sufficiently protect the rights and freedoms of citizens against undue interference from the state. Heterogeneity and generality of the regulations authorizing the acquisition of telecommunication data may raise doubts about the adequacy of restriction of rights in the sphere of freedom of communication with the principles laid down in the Constitution of the Republic of Poland. In March 2014, the Polish Constitutional Court, at the request of, inter alia, the Prosecutor General, the Polish Ombudsman, MPs and NGOs, took up the case concerning the powers of the police and special services in the field of permissible surveillance. In particular, the case concerns the collection of telecommunications data, including billing and location data. The Court examines, inter alia, whether the law regulates precisely enough the aim of data collection and useless data destruction rules.

On the basis of these reports, it can be concluded that the excessive surveillance of the Internet and telecommunication services by the state has not yet been a major problem in Poland. However, at the advent of big data methods, it is necessary to clarify the relevant legal principles. On the basis of press reports on the program PRISM, it can be assumed that the surveillance of society using big data is significantly broader in many well developed countries, particularly in the U.S. due to the potential danger of terrorism. This confirms the stereotypical concerns about the future consequences of big data applications in other countries.

In summary, according to the authors, the social aspect of big data should be regarded as the influence of universal mass data processing, particularly personal data, on the daily life of individuals and the society as a whole. The social aspect involves, in particular, the problem of threats to privacy.

Conclusions

The presented considerations above show, multifaceted characteristics of the emerging concept of big data. They also discuss whether the phenomenon of big data is a problem in the area of information technology and quantitative analytical methods, or rather it is a business or social issue. Naturally, scientists are interested in novelties and they explore the impact of technological developments on economic transformation and social environment. Historically, great inventions and technological advances often led to significant economic and social changes. Are new opportunities concerning data processing known as big data equally important? It is difficult to answer this question now.

The social aspect of big data is currently associated with the problem of privacy and personal data protection. On the one hand, the capabilities of modern technology and quantitative methods in the processing of large volumes of rapidly changing and unstructured data have much broader applications than those that are raised in the social aspect of big data (for example in quantum physics, and climate observation). On the other hand, the social aspect of big data is not just about technological and algorithmic processing capabilities of personal data, but also about the existing legal systems.

The authors have identified three basic aspects of big data: technological, business and social. However, it is arguable whether the technological aspect should be restricted to information technology by separating the aspect of analytical methods. Also, the authors suggested the possibility to isolate the legal aspect, because the legal system links, to some degree, the other aspects, setting legal limits to the use of big data. On the one hand, the law protects privacy, on the other hand, it allows or blocks business development. It is necessary to update the legal system to prevent the negative effects of technological changes, primarily on social relations, but without restricting economic development. Keeping up the legal system with the development of IT, particularly big data, should be the subject of further research.

References

Boyd D., Crawford K. (2012). Critical questions for big data. *Information, Communication & Society, 15:5*, 662-679.

Davenport T.H., Dyché J. (2013), *Big Data in Big Companies*, Retrieved from International Institute for Analytics: http://www.sas.com/resources/asset/Big-Data-in-Big-Companies-Executive-Summary.pdf.

Economist (2013, May 18). A new goldmine. Making official data public could spur lots of innovation. *Economist*. Retrieved from http://www.economist.com/news/business/21578084-making-official-data-public-could-spur-lots-innovation-new-goldmine.

Gazeta.pl. (n.d.). Retrieved January 10, 2014, from http://www.gazeta.pl (2014)

Gualteri M., *The Forrester Wave: Big data predictive analysis solution , Q1 2013*, Retrieved from Forrester Research Inc.

IBM. (n.d.). Retrieved January 10, 2014, from http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

Laney D. (2001). *Application delivery strategies*, Retrieved from META Group: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

MAC. (n.d.). Retrieved January 10, 2014, from https://mac.gov.pl/dzialania/michal-boni-o-big-data-na-warsaw-international-media-summit/

MAC. (n.d.). Retrieved January 10, 2014, from https://mac.gov.pl/dzialania/prywatnosc-zaufanie-wolnosc-i-bezpieczenstwo-michal-boni-przedstawia-polska-wizje-rozwoju-sieci-na-swiatowej-konferencji-o-cyberprzestrzeni-w-seulu/

Mayer-Schonberger V., Cukier K. (2013). *Big data – A revolution that will transform how we live, work, and think*. Boston, MA: An Eamon Dolan Book / Houghton Mifflin Harcourt.

Microsoft. (n.d.). Retrieved January 10, 2014, from http://blogs.msdn.com/b/microsoftenterpriseinsight/archive/2013/04/15/the-big-bang-how-the-big-data-explosion-is-changing-the-world.aspx (2013).

McKinsey Global Institute. (2011). *Big data: The next frontier for innovation, competition, and productivity*. Retrieved from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Naczelna Izba Kontroli. (2013). *Informacja o wynikach kontroli „Uzyskiwanie i przetwarzanie przez uprawnione podmioty danych z bilingów, informacji o lokalizacji oraz innych danych, o których mowa w art. 180 c i d ustawy Prawo telekomunikacyjne”*.

Oracle. (2013). *Oracle White Paper—Big Data for the Enterprise*. Retrieved from http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf (2013).

PcMag. (n.d.). Retrieved January 10, 2014, from http://www.pcmag.com/encyclopedia/term/62849/big-data

SAS. (n.d.). Retrieved January 10, 2014, from http://www.sas.com/offices/europe/poland/actual/press/news2_01_13.html

SAS. (n.d.). Retrieved January 10, 2014, from http://www.sas.com/big-data

Simon P. (2013). *Too big to ignore – The business case for big data*, Hoboken NJ: John Wiley & Sons.

Stępniak C. (2013). Kierunki wykorzystania systemów monitoringu miejskiego w zarządzaniu rozwojem miast. *Roczniki Kolegium Analiz Ekonomicznych SGH, 29*, 295-307.

Szupiluk R. (2013). *Dekompozycje wielowymiarowe w agregacji predykcyjnych modeli data mining*. Warszawa: Oficyna Wydawnicza SGH.

Szymielewicz K., Szumańska M. (2013). *Dostęp państwa do danych użytkowników usług internetowych, Siedem problemów i kilka hipotez*. Retrieved from Fundacja Panoptykon: http://panoptykon.org/files/panoptykon_dostep_panstwa_do_danych_internet_16.12.2013.pdf.

TVN-CNBC (2013, February 14). TV interview with Hucał M., the vice-president of Alior Bank. *TVN-CNBC*. Retrieved from http://tvn24bis.pl/wideo-archiwum,1/zmiany-w-zarzadzie-alior-banku,306356.html

Urbandictionary. (n.d.). Retrieved January 10, 2014, from http://www.urbandictionary.com/define.php?term=Big%20Data

Ustawa z dnia 29.08.1997 r. o ochronie danych osobowych, Dz.U. 1997 Nr 133 poz. 883 ze zm. Retrieved from Sejm RP: http://isip.sejm.gov.pl/DetailsServlet?id=WDU19971330883

Wieczorkowski J., Dałek M. (2013). Problem przeciążenia informacyjnego a integracja systemów informatycznych. *Zeszyty Naukowe nr 762 Ekonomiczne Problemy Usług nr 104, Uniwersytet Szczeciński*, 439-448.