■ Chapter 1 ■

Corporate Social Responsibility and the use of Big Data

Bartłomiej Zinczuk
Maria Curie-Sklodowska University in Lublin
Faculty of Economics

Abstract: The aim of this article is to present the key issues related to the use of Big Data sets by enterprises pertaining to socially responsible activities in the modern-day economy. Here the attention is paid to the characteristics of Big Data sets, their importance in the operation of a given enterprise and on showing the possibilities, limitations and dilemmas related to the use of Big Data with regard to Corporate Social Responsibility. The analyses presented is based on domestic and foreign literature and the reports of analytical and consulting companies, with a focus on the positive and negative consequences of the use of information and communication technologies (ICTs), in particular Big Data. Indicated are the selected benefits and limitations of Big Data in the operation of modern enterprises. Referred to also are the selected threats resulting from the complexity of the process of shaping modern technological solutions and their impact both on the functioning of enterprises in the modern economy and the course of socio-cultural changes in the world. An important issue addressed in this article is its indication of the appropriate involvement of enterprises in the identification of mechanisms related to data security, the protection of individuals' privacy, as well as the legal and ethical protection of practices related to the use of data by enterprises.
Keywords: Big Data, ITC, Social Responsibility, New Economy

Introduction

Globalisation and technological progress, stimulated by the emergence of ever faster computing machines and the development of the Internet, have led to the expansion of information and communication technologies (ICTs) in almost all spheres of life. The economy of the 21st century is dominated by

numerous digital technologies whose functioning relies on the rapid transfer and processing of large amounts of information in real time from multiple sources. Large data sets (Big Data), whose socially responsible use on the part of enterprises remains a relevant and significant topic in modern public discourse, are of particular importance when it comes to ICTs. The aim of this article is to present the key issues related to the use of Big Data by enterprises pertaining to socially responsible activities in the modern economy. Here attention is focused on the characteristics of Big Data and the presentation of the possibilities and limitations resulting from the use of Big Data in the operational running of enterprises. The analyses presented in this article are based on domestic and foreign literature and reports of analytical and consulting companies.

Materials

The modern economy, called the "New Economy", is based on modern information and communication technologies, which make it possible to collect, process and transmit large amounts of data in an ever shorter time. Since the beginning of the digital revolution, the phenomenon of faster and faster transfer of information between different entities, creating a socio-economic space, has been observed. The dynamic growth of data has brought about the creation of extremely capacious volumes called "Big Data" which means "large data sets".

The term Big Data was first used by NASA researchers M. Cox and D. Ellsworth in 1997, who paid attention to the problem of Big Data Collections and Big Data Objects which, according to the authors, were too large to be processed using standard algorithms and software on a single computer (Cox et al. 1997). The Gartner analytics and consulting company defined Big Data as "high volume, velocity and/or a variety of information assets that demand new, innovative forms of processing for enhanced decision making, business insights or process optimization" [https://www.gartner.com/it-glossary/big-data]. In 2001, Gartner's analyst created the 3V model describing three main Big Data attributes (Laney, 2001). Volume, indicating the amount of data, Velocity informing about the speed of data generation and processing; and Variety, describing the diversity of data. The IBM IT company, based on the 3V model, made a deeper characterisation of the basic

attributes of large data sets by dividing them into three dimensions (http://www-01.ibm.com/software/in/data/bigdata):

- Volume – an observable dynamic growth of data whose capacity is calculated in petabytes (PB). According to information published by IBM, 2.5 trillion bytes of data from various sources are generated every day around the world, including sensors used to collect buyer information, social media posts, digital photos, videos, purchase transactions, GPS signals from mobile phones [https://www.ibm.com/blogs/insights-on-business/consumer-products].
- Velocity – refers to the speed of data analysis and its inflow. At high speeds of the stream inflow of digital data, the ability to select data from the point of view of its informative value becomes extremely important.
- Variety – refers to various types of data: structural, semi-structural and unstructured (the so-called Dark Data), for which the use of typical methods of their processing and storage becomes insufficient.

Within a few years after the presentation of the "3V" model, further Big Data attributes such as Veracity and Value were added, creating the "5V" model (Michel, 2018). Big Data is a term used for data sets that are simultaneously characterized by large volume, diversity, real-time stream flow, variability, complexity, and require the use of innovative technologies, tools and IT methods to extract new and useful knowledge from them (Tabakow et al. 2014, p. 141).

The literature review allows us to state that the term Big Data refers to the process of collecting, storing and processing large data sets which require the use of technological tools and solutions for their analysis, taking into account their dynamic growth and structure.

Big Data's composition is characterised by a great variety of sources of their origin, most of which are not correlated with each other. When characterising data in enterprises, two main sources of origin can be identified: internal and external.

Data from internal sources, usually presented as structural, are mainly derived from relational databases such as: financial and accounting system, CRM solutions and transaction systems used in a given organisation (Filip, 2015, p. 43). Structural data are classified as organised, which, in the case of their specific structure, facilitates their processing in a timely fashion.

Data from external sources are referred to as semi-structural or unstructured, and are characterised by an irregular structure. This group includes various types of documents in text form (text files, scans of documents, e-mail content), multimedia, sensory and geolocation data, and from Internet sources (social networking sites, discussion forums, online stores) (Filip, 2015). According to experts, semi-structural and unstructured data constitute up to 85% of all data available on the Internet [https://www.computerworld.pl/news/Goniac-za-nieuchwytna-informacja-w-sieci].

Big Data does not only refer to large volumes of data stored on the server but also to the efficient processing of data sets from various sources. From the perspective of enterprises, Big Data covers a wide range of topics from storage and processing activities to data visualisation (Weinert, 2015, p. 14).

The enterprise's ability to access and analyse Big Data is considered the attribute of a modern enterprise. It is noteworthy that an enterprise's analytical capacity to process information from large datasets represents a competitive factor. Data is a specific resource that enables enterprises to grow, increase their productivity, and satisfy the needs of a wide range of stakeholders. According to P. Płoszajski, the processing of Big Data creates value for enterprises by (Płoszajski, 2013):

- making information transparent and accessible at a higher frequency,
- creating and storing more information on transactions in digital form for better performance measurement,
- creating more precise customer niches and better tailored products and services,
- supporting the development of the next generations of products and services,
- conducting controlled experiments.

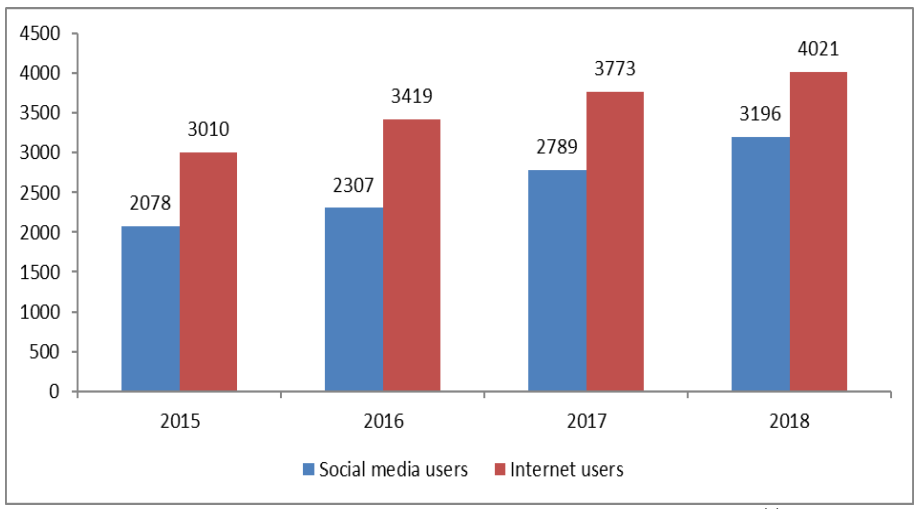
For many years user activity has been responsible for the process of generating data in the global information network. Smart devices connected to the network create and will generate most of the information on the Internet. Graph 1 shows the growth of the number of Internet and social media users globally in 2015-2018 (as of January each year).

According to the information presented in Graph 1, a dynamic, annual growth of Internet and social media users can be observed. This phenomenon is facilitated by the development of telecommunications infrastructure and the

availability of equipment enabling active use of the Internet. The information presented in Graph 1 shows that the growing number of network users will be the source of subsequent information streams created by the current and new Internet and social media users that can be used by enterprises.

The possibilities offered by Big Data are almost limitless. Development and use of modern technologies of data collection, processing, transfer and analysis are applied in such sectors of the economy as industry, banking, telecommunications, insurance, tourism, health care, public administration, power industry, transport, construction and science.

Graph 1. Global number of Internet and social media users in the years 2015-2018 in million



Source: Own calculations based on [https://wearesocial.com]

Discussion

Along with the development of digital technologies, the way the new technological solutions influence the social and economic sphere has changed. Big Data and its dynamic development raises a number of new, unexplored issues concerning the management of big data sets and their social consequences.

Big Data with the intensive development of ICT technology has enabled the acquisition, archiving, transfer and analysis of data and become an integral part of the enterprise management system. A key role in the process of enterprise management is played by data analytics, thanks to which it is

possible to discover new, unknown relationships among data and phenomena useful in cognitive, research and decision-making processes (Jurczyk-Bunkowska et al. 2017, p. 96-97). The knowledge gained in the process of data analysis can be used in business activity as well as in the publicsphere. Data analysis is used in such areas as: health care, environmental protection, public safety, banking and finance, production, trade, services, transport. According to a report by the IDC consulting firm, the global market for business analytics and Big Data is growing at a double-digit rate every year and should reach USD 210 billion by 2020. IDC analysts indicate that the largest customers of Big Data technology in the next few years will include the banking and industrial sectors, which confirms the value and number of realized and planned investments in these sectors (https://www.idc.com).

The development of Big Data technology has also contributed to the development of predictive analytics, which anticipate the behaviour of specific market entities (e.g. customers) and entire sectors of the economy on the basis of historical data analysis. Predictive analytics in the area of consumer behaviour is currently an extremely popular topic. Analysing the history of customer purchases and the information provided by customers through a variety of information channels (e.g. social media) allows enterprises not only to create a competitive offer but also to anticipate their future needs before they arise (Mróz, 2017, p. 145-146). The creation of added value for customers is now possible through the integration of digital technologies with expert knowledge of customer requirements, continuously modified by available data analyses provided by Internet of Things (IoT) technologies and supported by artificial intelligence (AI) (Nowak et al. 2018, p. 67). Apart from supporting strictly marketing activities, predictive analytics is also used in many other spheres of everyday life. It is applied in socially responsible projects devoted, among other things, to the development of specific diseases among the population, the possibility of natural disasters, climate change and many other predictable issues.

One of the problems with Big Data is the quality of data coming from many different sources, relating to their reliability and precision. It is reasonable to seek ever more perfect ways of separating valuable from information from useless information in the shortest possible time. Advances in computational technology offer new opportunities to use statistical methods in order to explore the relationships among variables, determine the

probability of their occurrence and to eliminate statistical errors (Kosior, 2016, p. 99-100).

The challenge for the future in the context of the development of Big Data will be to effectively manage expanding digital data streams. Enterprises with limited capacity in terms of using their own computing capacities will increasingly use data analytics in the form of external services, such as Cloud computing). and data storage,, which, according to Gartner, an analytic and advisory company, are defined as a transformation of the existing business model based on stationary solutions into the so-called "cloud shift", which can be described as a shift of enterprises (https://www.erp-view.pl/business intelligence).

The possibilities of using Big Data in the modern economy are extremely wide, bringing a together an amalgamation of benefits in economic, socio-cultural and environmental terms; but the dissemination of information technologies supporting the management of large data sets also reveals a number of problems which have arisen at the interface between the functioning of society and digital technologies. Large data sets and their misuse have contributed to socially irresponsible behaviour.

Sets of data as a valuable resource can be temptation for any organization. The analysis of the behaviour of individuals or entire social groups on the basis of digital records of their activities can be used against them in projects the participants have not onlynot given consent, but have also no knowledge about their purpose, course and effects of such analysis. The exchange and fee-based sharing of customer personal data (often sensitive) by enterprises to other market participants are always controversial, and in terms of social responsibility, such practices can be described as unethical; and often illegal. An example of such actions is the situation in which Facebook users found themselves in March 2018. Users received a signal that the information they had posted on Facebook had been used for purposes for which they had not given their consent. Cambridge Analytica, which had access to the Facebook platform, obtained and used data on 50 million users of the portal. Cambridge Analytica's data analysis was used for political purposes such as predicting and influencing the election decisions of Facebook users in the U.S. [https://www.spidersweb.pl/2018/03/facebook-cambridge-analytica]. The situation proved to be the starting point for a broad and lively debate on information security on the Internet, and on the importance to counteract the

use of data by different market players without the knowledge and awareness of social networking users.

Another example of the risk associated with large data sets is is the challenge of protecting them against theft. In 2018, 1000 GB of data per second was produced worldwide [https://ubezpieczenia-cyber.pl/statystyki-atakow-wycieki-danych]. Such a large amount of information appearing in such a short period of time represents a challenge not only for information storage security, but above all for the protection of infrastructure:its transfer and processing. The development of effective data protection systems and preventing data from falling into the hands of third parties should be a priority. According to the Check Point report, 2017 was an exceptional year in terms of the number and specificity of hacker attacks. According to specialists, every day hackers attack companies around the world more than 6.5 million times. Poland, ranked 7th in Europe in terms of network security, was a frequent target of cybercrime attacks in 2017 (http://businessjournal.pl/raport-check-point-software-technologies-polska-celowniku). Experts dealing with the issues of network security emphasise that the blame for hacker attacks can be attributed to imperfect IT security systems, but also to the human factor, which is often manifested by a lack of awareness concerning the threats associated with cyber attacks; and also a lack of knowledge about the mechanisms behind such attacks.

One of the risks associated with Big Data, and extremely important from the perspective of social responsibility, is the protection of privacy. Data created with the use of electronic devices communicating with each other without human participation (Internet of Things - IoT) may record and transmit, among others, confidential and sensitive information. Different institutions justify individual or group surveillance practices by explaining that it is for reasons of public security, which may not always be true. The supervision of activity together with the use of data collected in this way, without the consent and knowledge of those under observation, is an intrusion on the private sphere, and such activities are often both illegal and ethical. Another problem of Big Data from the legal point of view are the legislative delays in relation to the use of ICT in everyday life. The global nature of the network also raises the issue of the legal protection of IT infrastructure elements, which are often located in different countries and subject to different jurisdictions. This situation makes it all the more necessary for legislators to regulate of provisions with respect of of their social perception and impact.

Conclusions

Information and communication technologies (ICT) have become an essential part of everyday life and their increasing use has fundamentally changed the functioning not only of enterprises and institutions but also, and above all, of society as a whole. Continuous and intensified data generation is a major challenge for enterprises, who must use such data in a socially responsible way. Ignorance of the long-term effects of digital technological solutions should be an impulse for the conducting of research on the impact of ICT on various spheres of human life, and its socially responsible application. The effect of this socially important discourse would be to show the complexities associated with absorbing modern information and communication technologies into business practices. It will also be important to pursue andan impact assessment on the course of economic and socio-cultural changes. The context of the use of information contained in Big Data, which is incompatible with binding legal and ethical principles, should constitute a starting point for actions aimed at the effective elimination of such practices among enterprises and public institutions, as well as the active prevention of their occurrence. Legislative actions must keep pace with technological changes, the effective protection of personal data and the fight against digital crime. The socially responsible use of ICT moust also be to the fore, with actions addressing both the users and beneficiaries of digital solutions. Only such steps can effectively solve the problems related to the misuse of Big Data.