BIG DATA: CHANGES IN DATA MANAGEMENT

Šebalj, Dario; Živković, Ana; Hodak, Kristina

Download date / Datum preuzimanja: **2024-12-28**

Dario Šebalj
*Josip Juraj Strossmayer
University of Osijek
Faculty of Economics in Osijek
Trg Ljudevita Gaja 7,
31000 Osijek, Croatia
dsebalj@efos.hr
Phone: +385912244147*

Ana Živković
*Josip Juraj Strossmayer
University of Osijek
Faculty of Economics in Osijek
Trg Ljudevita Gaja 7,
31000 Osijek, Croatia
azivkovi@efos.hr
Phone: +385912244167*

Kristina Hodak
*Drinska 13,
31207 Tenja, Croatia
krhodak@efos.hr
Phone: +385958728021*

BIG DATA: CHANGES IN DATA MANAGEMENT

Abstract

We live in a digital environment where everything we do leaves a digital trace. Data are collected from various sources – social network posts, e-mails, sensors, image and video content, search engines, online sales, etc. The biggest reason for this growth of data could be found in technological advancement, since data can be easily and cheaply stored and shared today. This new trend in acquiring data demands a completely new approach to their processing and analysis. The aim of this paper, based on analysis of actual and relevant sources, is to present the situation and trends in the collection, processing, analysis and use of data that are complex, fast-growing, and diverse in type and content. In the introduction, the research problem has been defined. Afterwards, the term "Big Data" and its basic four dimensions have been explained. In the main part of the paper, the examples of Big Data analyses have been shown, as well as interesting results yielded by those analyses. Also, the special review about Big Data in management has been presented.

Keywords: Big Data, data, management

1. Introduction

The digital environment, in which we live, creates a large amount of data every moment. The biggest reason for this growth of data could be found in technological progress. Nowadays, data can be easily and cheaply stored and shared, which was not the case just a few years ago. Data are now collected by almost everything that surrounds us. Almost everything we do leaves a digital trace that can be used for various analyses later. Frequently used devices (like mobile phones) contain several sensors that generate a bunch of data. More and more devices are connected to the Internet and can store and share data at high speed.

According to IBM[1], people create 2.5 quintillion bytes of data every day — so much that 90% of the data in the world 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. According to Zikopoulos (2015), for a number of years, Facebook was adding a new user every three seconds; today these users collectively generate double-digit terabytes of data every day. In fact, in a typical day, Facebook experiences more than 3.5 billion posts and about 155 million "Likes".

Let's take our own example into account. Every day we receive and send e-mails, use search engines intensively, chat with friends and colleagues via chat and text messaging, buy online, pay bills via mobile phones or the Internet, share images and videos on Facebook, use cloud data stores, write letters and do some spreadsheet calculations in Google Docs, etc. All this information, along with much more, is stored (usually on the cloud) and later used for detailed analysis.

Figure 1 What happens in an Internet minute?

What happens in an **Internet Minute?**

6 new Wikipedia articles published

204 million emails sent

1300 new mobile users

47,000 app downloads

amazon **$83,000** in sales

100+ new LinkedIn accounts

20 million photo views
3,000 photo uploads

320+ new Twitter accounts
100,000 new tweets

6 million Facebook views

Google **2+ million** search queries

You Tube **1.3 million** video views
30 hours of video uploaded

Source: Authors, adapted from: http://www.sparkyhub.com/what-happens-on-the-internet-in-60-seconds-infographic/ (Accessed on: November 27, 2015)

According to the McKinsey report (2011), enterprises and consumers stored more than 13 exabytes of new data on disk drives in 2010, which is more than 50,000 times the information stored in the US Library of Congress. McKinsey also projected that demand for deep analytical positions in the big data world could exceed the supply being produced on current trends by 140,000 to 190,000 positions. In addition, they projected a need for 1.5 million additional managers and analysts in the US who will have to become "data literate". A huge amount of data could be very useful for enterprises if they have the tools and skills to extract a value from these data.

The aim of this paper, based on analysis of actual and relevant sources, is to present the situation and trends in the collection, processing, analysis and use of data that are complex, fast-growing, and diverse in type and content.

H_1: Gathering and analysis of Big Data increases the efficiency in organizations and allows management to make better business decisions.

2. Big Data

Lately, there has been a lot of conversation about the concept of Big Data. This seemingly simple concept encompasses much more than one might think at first sight. As mentioned in the introduction, data about almost everything are monitored and stored today and, therefore, we have access to a large number of data. This is why the "Big Data" term is commonly accepted. Marr (2015: 10) in his latest book claims that the real value is not in large volumes of data but in what can be done now with that data. He also points out that it is not the amount of data that is making the difference, but our ability to analyze vast and complex data sets beyond anything we could ever do before. Knapp (2013) defines Big Data as the tools, processes, and procedures that allow an organization to create, manipulate, and manage very large sets and storage facilities. SAS Institute Inc. states that Big Data is a term that describes the large volume of data –

both structured and unstructured – that inundates a business on a day-to-day basis. This company also agrees that it is not the amount of data that is important, but what organizations do with the data. Big Data can be analyzed for insights that lead to better decisions and strategic business moves.[2] This term was even added in 2013 to the Oxford English Dictionary.[3] Forbes Magazine (Arthur, 2013) defines Big Data as a collection of data from traditional and digital sources inside and outside your company that represents a source for ongoing discovery and analysis. On its website IBM states that Big Data is being generated by everything around us at all times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it. Big Data is arriving from multiple sources at an alarming velocity, volume and variety. To extract meaningful value from Big Data, you need optimal processing power, analytics capabilities and skills.[4] O'Reilly explains that Big Data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures. To gain value from this data, you must choose an alternative way to process it (Dumbill, 2012). Jennifer Dutcher from Berkeley, School of Information, also tried to define Big Data. She asked more than 40 thought leaders in publishing, fashion, food, automobiles, medicine, marketing, and every industry in between how exactly they would define the phrase "Big Data".[5] They said the following: Jon Bruner (Editor-at-Large, O'Reilly Media): "Big Data is the result of collecting information at its most granular level"; Reid Bryant (Data Scientist, Brooks Bell): "Big Data will ultimately describe any dataset large enough to necessitate high-level programming skill and statistically defensible methodologies in order to transform the data asset into something of value"; Rohan Deuskar (CEO and Co-Founder, Stylitics): "Big Data refers to the approach to data of 'collect now, sort out later'...meaning you capture and store data on a very large volume of actions and transactions of different types, on a continuous basis, in order to make sense of it later"; AnnaLee Saxenian (Dean, UC Berkeley School of Information): "Big Data is data that can't be processed using standard databases because it is too big, too fast-moving, or too complex for traditional

data processing tools"; Anna Smith (Analytics Engineer, Rent the Runway): "Big Data is when data grows to the point that the technology supporting the data has to change"; Mark van Rijmenam (CEO/Founder, BigData-Startups): "Big Data is not all about volume, it is more about combining different data sets and to analyze it in real-time to get insights for your organization"; Timothy Weaver (CIO, Del Monte Foods): "A lot of different data coming fast and in different structures". Gartner, the leading global research company providing information technology related insights, also gave a definition of Big Data. They said that Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.[6]

A Google search for the phrase "Big Data", conducted on November 11, 2015, returned about 770 million results. On the first page of results there were mostly ads from SAS, IBM and Oracle. Power (2014) also conducted an identical Google search. In April 2013 the same phrase returned about 17.7 million results. It means that the number of results in 2.5 years increased more than 43 times! Power narrowed the search to the phrase "define Big Data" and got 2.68 million results, and for the phrase "What is Big Data" 24.1 million results. Nowadays, for the first phrase we are getting 71 million results (26 times more), while the second phrase is returning 805 million results (33 times more).

In 2014, Gartner released its annual report i.e., Hype Cycle for Emerging Technologies, which gives a view of how a technology will evolve over time. Each Hype Cycle drills down into the five key phases of a technology's life cycle:[7] Technology Trigger, Peak of Inflated Expectations, Trough of Disillusionment, Slope of Enlightenment, and Plateau of Productivity. According to this report (Figure 2), Big Data has officially passed the "peak of inflated expectations" and is now on a one-way trip to the "trough of disillusionment". Gartner says it has done so rather rapidly, because we already have consistency in the way we approach this technology, and because most new advances are additive rather than revolutionary.[8]

Figure 2 Gartner's Hype Cycle (2014)

Source: http://siliconangle.com/blog/2014/08/19/gartners-hype-cycle-big-datas-on-the-slippery-slope/ (Accessed on: November 17, 2015)

However, Big Data dropped from Gartner's 2015 Hype Cycle (Figure 3) and no longer appears. The reason for this lies in the fact that Big Data is already an integral part of other technologies such as autonomous vehicles, Internet of Things, Machine Learning, Data security, Citizen Data Scientist etc.[9]

Figure 3 Gartner's Hype Cycle (2014)

Source: http://www.gartner.com/newsroom/id/3114217 (Accessed on: November 17, 2015)

The richness and the power of Big Data lie in the enormous amount of data. Today, it is possible to collect and store that amount of data relatively cheaply. As a result, companies are able to carry out a variety of business analyses on ALL data, and not, as before, on a particular sample. That sample, as much as we try to make it representative, can never have the quality of the whole. This is where special software tools come in order to analyze the large amount of data, without sampling a smaller part. Most companies do not have access to a large amount of data, nor financial resources or data experts who know how to analyze this data, as is the case with global and large companies, like Google, Amazon, Facebook, eBay etc. However, on the small amount of different data types the unthinkable analysis can also be done. This is further proof that the term "Big Data" has been a bit clumsily created. According to Marr (2015), despite the exponential growth of data and information, less than 20 per cent of the data companies currently hold is used to inform decision-making. And these 20 per cent only took traditional structured KPI or financial data into account. Insight from the unstructured data represents a rich untapped vein of information gold that is currently largely ignored.

Mayer-Schönberger and Cukler (2013) explain that data can tell us incredible things, discover unexpected correlations and solve, at first sight unsolvable, problems. When doing so, in data analysis, it is necessary to ask the question **WHAT**, and not **WHY**. This means that, in data analysis, there is no need to look for causes in obtained results. It is important what these results show, even if it might seem illogical. One of the examples is the famous American retail store Wallmart. A research of cus-

tomers' buying habits has revealed that a few days before an approaching storm, customers massively buy flashlights and a specific type of chips. Why they buy the flashlights is logical, but why the chips? Accordingly, before every storm, Wallmart has been putting a large amount of those specific chips at the store entrance. Their sales increased quite a lot.

2.1 Big Data dimensions

14 years ago, Doug Laney from the META Group (today in Gartner) wrote an article (Laney, 2001) in which he predicted the future of Big Data. He was the first one to mention three dimensions (3 V) in e-commerce:

- Volume
- Velocity
- Variety

Soubra (2012) claims that these three properties define the expansion of a data set along various fronts to where it merits to be called Big Data. This is an expansion that is accelerating to generate yet more data of various types.

Volume refers to the vast amount of data generated every second (Marr, 2015). The benefit gained from the ability to process large amounts of information is the main attraction of big data analytics. Having more data beats out having better models (Dumbill, 2012). On the e-commerce example, Laney (2001) explained that the lower cost of e-channels enables an enterprise to offer its goods or services to more individuals or trading partners, and up to 10x the quantity of data about an individual transaction may be collected – thereby increasing the overall volume of data to be managed.

Table 1 Data size example

1 byte	1 character, number, letter
2 kilobytes	1 typewritten page
1 megabyte	Short novel
10 megabytes	Digital chest X-ray
1 gigabyte	7 minutes of HD-TV Video
100 gigabytes	Library floor of academic journals
10 terabytes	Printed collection of the U.S. Library of Congress
1,5 petabytes	All 10 billion photos on Facebook
50 petabytes	Entire written works of mankind, from the beginning of recorded history, in all languages
5 exabytes	All the word ever spoken by mankind
1 zettabyte	250 billion DVDs
1 yottabyte	Size of the entire World Wide Web; it would take approximately 11 trillion years to download a Yottabyte file from the Internet using high-power broadband

Source: Adapted from: https://datascience.berkeley.edu/big-data-infographic/ (Accessed on: November 27, 2015)

Velocity refers to the speed at which new data is generated and moving around the world. For example, for detection of credit card fraud millions of transactions are monitored in almost real time in order to find unusual patterns (Marr, 2015). The importance of data's velocity — the increasing rate at which data flows into an organization — has followed a similar pattern to that of volume. The Internet and mobile era enables online retailers to compile large histories of customers' every click and interaction: not just the final sales. Those who are able to quickly utilize that information, by recommending additional purchases, for instance, gain competitive advantage (Dumbill, 2012). Laney (2001) explains that e-commerce has also increased point-of-interaction (POI) speed and, consequently, the pace data used to support interactions and generated by interactions. Soubra (2012) states that initially, companies analyzed data using a batch process. One takes a chunk of data, submits a job to the server and waits for delivery of the result. That scheme works when the incoming data rate is slower than the batch processing rate and when the result is useful despite the delay. With the new sources of data, such as social and mobile applications, the batch process breaks down. The data is now streaming into the server in real time, in a continuous fashion and the result is only useful if the delay is very short.

Variety refers to the different types of data being generated – from financial data to social networks data; from images to sensor data; from video to voice recording (Marr, 2015). Rarely does data present itself in a form perfectly ordered and ready for processing. A common theme in big data systems is that the source data is diverse, and does not fall into neat relational structures. It could be text from social networks, image data, a raw feed directly from a sensor source. None of these things come ready for integration into an application (Dumbill, 2012).

Figure 3 3V concept
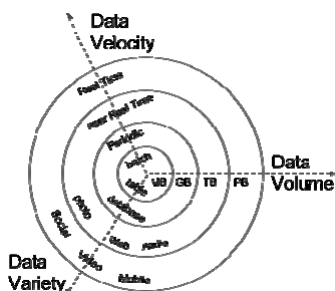
Lately, with the growth of interest and popularity of Big Data, the additional V letters are showing up. The 4th (added by IBM) is **Veracity**. It refers to generated data mess (Marr, 2015). Veracity deals with uncertain and imprecise data (Syed, 2013). Authenticity of the data increases with automation of data capture. With multiple sources of data, it would be possible to triangulate the results for authenticity (Moorthy et al., 2015). Borne (2014) mentioned additional 6Vs: Validity, Value, Variability, Venue, Vocabulary and Vagueness.

2.2 Data types

Because of one "V", data types have become an important issue in Big Data analysis. It is about variety. Goes (2014) finds variety the most interesting dimension of Big Data from an IS (information system) perspective. Putting together data from sensors and the "Internet of things", the vast repository that we call the Web, user-generated content, social media, data generated and consumed on mobile platforms, and data from enterprise systems, allows researchers to ask and answer questions that explain and predict individual behavior and detect population trends. According to Syed et al. (2013), Big Data is made of structured and unstructured information (about 90% is unstructured). The most common form of **structured** data is a database where specific information is stored based on methodology of columns and rows. This type of information is searchable, efficiently organized for human readers and secure. However, less importance is given to this type of data because these data are already organized, it is easy to analyze them, and most importantly – they include only 10 per cent of all data. These are, for example, Point-of-Sale transactions, financial data, data about customers or suppliers, invoices etc.

Analysis of **unstructured** data is a big challenge for all companies. They cannot be aligned into columns and rows. Unstructured data can be images, videos, web pages, text files, e-mails, social networks posts, PDF files, PowerPoint presentations etc. According to Marr (2015), data could be analyzed from our activities, talks, pictures and videos, sensors, internet-of-things. All of that has been searched and is being stored. If we listen to music using our smart phone or digital music player, data is also collected on what we are listening to, how long we are listening and what tracks we are skipping past. And artists like

Lady Gaga are using this data to create playlists for live gigs and influence future song creation. Billions of emails are sent and stored every week. One million accounts are added to Twitter every day, every sixty seconds 293,000 status updates are posted on Facebook, each day 350 million photos are uploaded to Facebook. The average smartphone has at least 6 sensors (GPS sensor, accelerometer, gyroscope…) and each of them continuously captures new data.

Figure 4 Analysis of structured and unstructured data
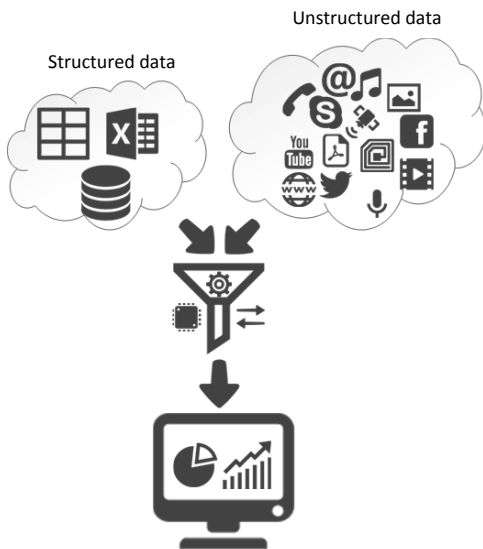
Source: Authors

All of the examples above show how much unstructured data is being captured today. Special software tools are needed for the analysis of that kind of data, but they can reveal unbelievable results.

3. Big Data in Management

As mentioned in the 2nd Oracle Academy day[10], Ljiljana Perica, Oracle Business Solution Leader, points out that over 80% of the data in the organization are unstructured: from phone calls, e-mails, social networks and sensors, to a variety of video, audio and image records. Financial transactions and official and unofficial organizational documents are only partly structured, so in a small percentage it can be truly spoken of structured data. The versatility and velocity make it difficult to manage the data, which of course means that the more data are collected, the more difficult it is to manage them. Technology has advanced so much that the sentences and phrases of telephone calls (for example in call centres) can be translated into mathematical records that will give information about customer preferences and their attitudes about products (Katz, 2015). The mobile environment plays an important role in data management providing incredible velocity and huge amounts of different data (Park et al., 2014). All the necessary data are already in the organization, but they need to be found and transformed into valid information. Managers usually have loads of data, but often do not know what to do with them (Finley et al., 2015). Different people may interpret a single data differently, which means that a variety of information can be drawn from one data (De et al., 2016). Satisfaction as a dependent variable is one of the most important items of each organization both internally and externally. Internal satisfaction refers to all employees of an organization, and external satisfaction is that expected by consumers. Big Data can contribute to the increase of both groups of satisfaction. For example, organizations that offer online services, such as Amazon, can increase the satisfaction of its customers in a simple way, following the "clicks", the exact time spent on their services, by observing the precise term that increases the attendance of their pages and the like. One of the ideal examples are organizations that offer online movie watching. Thanks to the Big Data platform, they can figure out the profiles of their customers, and thus foreshadow their future conduct. The user data occur within the types of movies they watch most often, a specific genre viewed at a given time, their possible ratings and reviews, a list of far more watched movies, list of movies that have not been viewed to the end, et cetera. In fact, all activities performed by the consumer, or through an organization, are recorded and stored. By virtue of monitoring of the aforementioned particles, it is possible to recommend to a consumer a movie that has not been viewed so far, given his previous habits and preferences that are "read out" from these data. This can be understood as a kind of market research, but this is a much simpler and cheaper variant which can quickly get to the desired information. Storage costs are low, and data storage is not created by people, but rather by devices. Managing the satisfaction may not be only at the level of questioning the existing needs and interests, but also in imposing new, even pilot studies. Penetrating the consumer profile is increasing by

connections with social networks creating a more thorough insight into past experience, environmental, consumer, and purchasing habits, emotions and all forms of interaction. This enables gaining insight not only in demographic and transactional data, but also contextual.

Big Data is not applied only in online organizations. Retail organizations can maintain contact with consumers across multiple channels, the interactions and transactions of consumers stored in various databases are making it easier to interpret the consumer/customer's choice. This means that retail can accurately connect a potential takeover of customers with marketing campaigns, and can do the analysis of the basket. It serves for recognition of their buying habits, as well as predicting them. All these data, from structured such as customer's reviews to unstructured, are stored in the so-called silos, and Big Data enhances their segmentation. Thanks to sensors in stores, QR codes and their deep analytics, sale is increasing; it reduces storage costs, and most importantly - increases customer satisfaction. From the customer's perspective, it is possible for organizations that have successfully managed data, through on-line application and even by reading the codes on mobile applications, to find out whether there is a particular size of shoes in another shop in the world, and carry out an automatic on-line purchase of the required pair.

Except consumer satisfaction, employee satisfaction can be managed as well. Employees in today's organizations are not even aware of how they are "followed". It is not only about classic monitoring, but managing productivity. Namely, the sensors in the rooms allow monitoring of employee's movements. It gives an insight into absenteeism, but also presentism, or at the time that employees physically spend within the organization. If it is a fixed working place, internal movements in the organization are also recorded. This practice should not be used for unethical monitoring, but actually for identifying trends that should be reduced. The monetary value of time is very important because time loss consumes resources. Sensors can see the values such as excessive movement to the copier, which can point to the need for dislocating a machine or purchasing additional ones. The human eye cannot easily measure the time spent on such movements, but devices can accurately calculate the time losses in seconds. Big Data can also be a tool for the analysis of individual and team behaviours of employees. Modern

organizations have sensors and badges which are recording individual and team work and time spent on general and specific work tasks (Gerard et al., 2014). Big Data is in that context a control system that helps connect behaviour with organizational results, and measures performance that represents employee productivity (Warren et al., 2015). The aforementioned telephone calls as data source can be used not only for control of consumers and customers, but also employee control. Banks and other organizations that offer their services and advice through phone calls can monitor employees checking whether they provide consumers with the necessary and valid information. Mathematical records of conversations will direct superiors to possible deviations and errors, and show that employees need to be alert, additionally trained, or removed from the work place. The dynamics of the employee therefore has never been easier to follow, so the current diary of work is considered to be a very outdated tool. Measuring the impact of formal and informal communication channels can also determine the organizational behaviour and predict the future. However, many HRM experts consider the Big Data platform is less applicable in human resources management than in other economic and non-economic sectors (Chynoweth, 2015). They point out that the quality of data varies because people are sensitive particles that cannot be easily objectified and generalized as in the case of material resources. Human behaviour is complex and predicting the future behaviour of employees is not easy relying only on objective data. In fact, HR managers report that the learning data management in their area is not feasible because the decisions about people are often better made based on empirical experience and intuition. Nevertheless, this criticism must not deny the benefits of Big Data in management because Big Data helps more than it hinders.

Big Data is not only helpful in the managing of absenteeism, productivity and satisfaction, but also in the process of selection of candidates for certain jobs. At the same time, out of structured and unstructured data, it is possible to find out the age, sex, education, previous work experience, interests, hobbies and recreation, residence, social events that someone visits, membership and belonging to different communities, purchasing habits, family relationships, social relationships, achievements, political leanings and so on. If additional testing of candidates is performed with IT technology

and in a virtual way, it is possible to actually eliminate a huge number of candidates, thus narrowing the choice to possibly several desirable ones. This avoids the cost of time, money and energy which is especially important in the modern labour market where there are many candidates with a variety of personality profiles, education levels, competencies and skills. It is understood that computers and robots will never be able to replace some segments of human contact so "face to face" interaction will never lose its importance, but it is also true that due to the proper management of data in organizations HR managers do not have to interview hundreds of candidates, but only the few who meet all the requirements and criteria.

An especially important benefit of Big Data is reflected in the examples presented at the Oracle Academy lecture that emphasize the great use of Big Data in telecommunications companies, banks, insurance companies and the entertainment industry. In the public sector, this platform is certainly applicable in health care, education and transport. The Chicago Police Department for criminal investigation, has sped up the detection of criminal acts through the data of prior criminal incidents, their notes, data on suspects, transcripts of 311 calls and the like (Oracle Academy lecture). The benefit of the program is reflected in the rapid finding of suspects and potential suspects.

Big Data brings to the management completely new, so far unexplored possibilities that may contribute to the creation of added value. The issue of competitiveness on the market is one of the key problems of modern organizations so there is often a case of two or more competing organizations on the market with complementary products. In fact, these organizations operate within the same market, and even if there is no other link, they share the same market. In practice, there are more links: from common suppliers, customers, geographic proximity, the same technology and manufacturing processes. Detailed information about such relationships may not be formally available, but they can be found in the unstructured data of different shapes. Competition should be monitored in all possible ways, from tracking articles and interviews of competitive organizations in the media, to monitoring their websites, blogs and social networks. The intensity of the monitoring of competitive organizations should not be significantly lower than the monitoring of own organization. Online reviews, "clicks", "likes" and "tweets" are a source of information about own products and services, but also about competing ones. Any observed deviation of own products from the desired state is an alert for fast action on consumer reactions. Such situations occur also with competitors, so active monitoring of their consumers' reactions can open up space for "attack" on the weak points of competitors. The most important results of the implementation of Big Data are reflected in the possibility for rapid decision making, resolving problems quickly and predicting future events, greater productivity of employees and the entire organization, and greater competitiveness in the market. In the area of management, Big Data is becoming an increasingly dominant concept in Supply Chain Management (Waller, Fawcett, 2013). Predictive analytics in logistics due to current phenomena and behaviours is trying to predict the future, storage inventories and their costs. Structured data in the logistics are not new, but the collection of unstructured data could greatly change the current perception of SCM. The sensors can detect the stocks in the warehouse, the locations where the shelves are running out of certain products, overcrowding in the distribution centre, misplaced inventory and so on.

There is no education system that produces specifically "data managers" because data manipulation requires a multidisciplinary approach. It is hard for engineers to independently solve medical issues for example. Therefore, a prerequisite for the successful implementation of Big Data in the supply chain requires primarily excellent knowledge and expertise, that is, theory and practice of SCM, with upgraded management and Big Data knowledge. Therefore, sometimes it is not enough to have a person who has only experience in analysing the data in the information and technical sense, but at the same time an employee must understand the primary work that the organization does. For this reason, perhaps it is not too helpful to outsource this kind of work, because only internal employees, with long experience in the business that is the subject of analysis, and extensive experience in the organization, can adequately use Big Data. Gob (2014) points out that the successful use of Big Data is preceded by a high level of statistical knowledge and skills. Analytical skills should never exclude managerial skills. Gerard et al. (2014) point out that successful "data" experts should be able to find the real business problem, but from the data point of view. A person

who would be a decision maker with successful use of Big Data is expected to possess managerial skills, technical and IT skills, and above all, knowledge of the organization itself, so it can be assumed that this job in the organization could be performed by more than one person. Big Data in the supply chain requires knowledge of statistics, forecasting methods and probability, optimization, analytical mathematical modelling, finance, marketing and accounting (Waller, Fawcett, 2013). This leads to the conclusion that it is necessary to have at the same time the possibility of quantitative and qualitative considerations. It is difficult to find in one person all the necessary requirements because it is not enough just to find and identify, but also to understand some phenomena and predict the future. The combination of youth and experience is often a winning combination. Older and more experienced employees have great knowledge of their profession, but they are not as flexible, fast and prone to technological changes as the younger members of staff who, on the other hand, lack professional and organizational experience.

Big Data greatly improves the accounting tasks and audit (Griffin, Wright, 2015). Thanks to the Big Data, external and internal auditors have access to the history of the entire organizational culture. Quality control becomes more realistic as availability and interpretation of the data minimizes the risk of subjective assessment which greatly facilitates the work for auditors, as well as their conscience. Sharing data must be protected and privacy ensured. Everyone cannot have access to data, but those who do, must operate ethically. The social aspect of Big Data is a very important component of data management as unethical behaviour is the main criticism directed at the digital economy. The fear of data theft is becoming greater and greater with use of Big Data as more paths are opened for thieves. Hackers are able to take over all financial information about an organization, but also about their customers and consumers in a matter of hours (Katz, 2015) and thus undermine consumer loyalty to an organization. At the level of the organizational system, the targets are mostly official and unofficial e-mails, through which it is possible to get the "meta-data" (such as a person's address, phone number or account number). That's why Big Data is also expected to serve as protection against criminal activities. For users of banking services it is especially important and many banks are warning users who change the usual ATM, country or way of access to accounts by controlling user authenticity. Even accounts and profiles on networks and on-line services warn users when they try to log on from a new IP address, thus trying to preserve the privacy of its users.

Big Data is not a fashion in management, but a necessity. It helps to answer the question of how to do something better, faster and cheaper. It is the question of time when organizations will be divided into those who use Big Data and those who do not (Finley et al., 2015). Organizations that raise the possibility of recording, processing, analysing and distribution of data for the purpose of decision-making are likely to outperform their competitors and respond faster to the needs of their customers.

4. Conclusion

Big Data is certainly changing the way companies and their managers make decisions. Increasing attention is given to the unstructured data (image and video content, sensor data, social network data, logs…) that, until recently, have been neglected due to the small storage capacity for these data and the impossibility of their in-depth analysis. With the technology development and facilitating of data manipulation, unstructured data have become the main interest of business analysis. The companies that have the ability to gather and store a large number of data acquire a huge advantage. Such kind of data might show the incredible (and seemingly illogical) correlation between two or more variables. It is important to emphasize that most of these results were the result of the analysis of all data (N = all), rather than analysis of a representative sample. As it can be seen in the example of Wallmart, the obtained correlation does not need to have a logical explanation. It is important that, on the basis of such analysis, business decisions could be made.

As is evident form this literature review, Big Data analysis is given great importance in organizations and it helps management in decision-making. Managers are now beginning to be aware that all the necessary data are already in the organization, and, thanks to new technologies, they could be stored and transformed into valid information.

References

1. Arthur, L. (2013), "What is Big Data", Forbes Magazine, Available at: http://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/ (Accessed on: November 16, 2015)

2. Borne, K. (2014), "Top 10 Big Data challenges – A serious look at 10 Big Data V's.", Available at: https://www.mapr.com/blog/top-10-big-data-challenges-%E2%80%93-serious-look-10-big-data-v%E2%80%99s#.VLk8Iy6mRYo (Accessed on: November 17, 2015)

3. Chynoweth, C. (2015), "Stop doing dumb things with data", People Management, December 2015, pp. 42-44.

4. De, K., Klimentov, A., Maeno, T., Mashinistov, R., Nilsson, P., Oleynik, D., Panitkin, S., Ryabinkin, E., Wenaus, T. (2016), "Accelerating Science Impact through Big Data Workflow Management and Super-computing", EPJ Web of Conferences.

5. Dumbill, E. (2012), "What is big data", O'Reilly Media, Available at: https://www.oreilly.com/ideas/what-is-big-data (Accessed on: November 17, 2015)

6. Dutcher, J. (2013). Data Size Matters [Infographic]. Available at: https://datascience.berkeley.edu/big-data-infographic/ (Accessed on: November 17, 2015)

7. Finley, F., Blaeser, J., Djavairian, A. (2014), "Building and Maintaining Capabilities that Deliver Results", Supply Chain Management Review, Vol. 19, No. 5, pp. 24-31.

8. George, G., Haas, M. R., Pentland, Alex (2014), "Big Data and Management", Academy of Management Journal, Vol. 57, No. 2, pp. 321-326.

9. Göb, R. (2014), "Discussion of Reliability Meets Big Data: Opportunities and Challenges", Quality Engineering, Vol. 26, No. 1, pp. 121-126.

10. Goes, P. B. (2014), "Big Data and IS Research", MIS Quarterly, Vol. 38, No. 3, pp. 3-8.

11. Griffin, P. A., Wright, A. M. (2015), "Commentaries on Big Data's Importance for Accounting and Auditing", Accounting Horizons, Vol. 29, No. 2, pp. 377-379.

12. Katz, D. M. (2015), "Big Data, Smaller Risk", CFO, Vol. 31, No. 8, pp. 38-41.

13. Knapp, M. (2013), "Big Data", Journal of Electronic Resources in Medical Libraries, Vol. 10, No. 4, pp. 215-222.

14. Laney, D. (2001), "3D Data Management: Controlling Data Volumen, Velocity, and Variety", META Group, Available at: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (Accessed on: November 17, 2015)

15. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, Ch., Byers Hung, A. (2011), "Big Data: The next frontier for innovation, competition, and productivity", McKinsey Global Institute. Available at: http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_exec_summary.ashx (Accessed on: November 18, 2015)

16. Marr, B. (2015). Big Data: Using smart Big Data, analytics and metrics to make better decisions and improve performance. Hoboken, NJ: John Wiley & Sons Ltd.

17. Mayer-Schönberger, V., Cukier, K. (2013). Big Data. A revolution that will transform how we live, work and think. Boston, MA: Houghton Mifflin Harcourt.

18. Moorthy, J., Lahiri, R., Biswas, N., Sanyal, D., Ranjan, J., Nanath, K., Ghosh, P. (2015), "Big Data: Prospects and Challenges", The Journal for Decision Makers, Vol. 40, No. 1, pp. 74-96.

19. Park, J., Kim, H., Jeong, Y., Lee, E. (2014), "Two-phase grouping-based resource management for big data processing in mobile cloud computing", International Journal of Communication Systems, Vol. 27, No. 6, pp. 839-851.

20. Power, D. J. (2014), "Using 'Big Data' for analytics and decision support", Journal of Decision Systems, Vol. 23, No. 2, pp. 222-228.

21. Soubra, D. (2012), "The 3Vs that define Big Data", Available at: http://www.datasciencecentral.com/forum/topics/the-3vs-that-define-big-data (pristup: 17.11.2015.)

22. Syed, A. R., Gillela, K., Venugopal, C. (2013), "The future revolution on big data", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, No. 6, pp. 2446-2451.

23. Waller, M. A., Fawcett, S. E. (2013), "Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management", Journal of Business Logistics, Vol. 34, No. 2, pp. 77-84.

24. Warren, D., Moffitt, K., Byrnes, P. (2015), "How accounting records will change with Big Data", Accounting Horizons, Vol. 29, No. 2.

25. Zikopoulos, P., deRoos, D., Bienko, Ch., Buglio, R., Andrews, M. (2015). Big Data Beyond the Hype. New York: McGraw-Hill Education.

(Endnotes)

1 Available at: https://www.01.ibm.com/software/data/bigdata/what-is-big-data.html (Accessed on: November 13, 2015)

2 Available at: http://www.sas.com/en_us/insights/big-data/what-is-big-data.html (Accessed on: November 13, 2015)

3 Available at: http://mashable.com/2013/06/13/dictionary-new-words-2013/#IqSgIfPZsPq9 (Accessed on: November 16, 2015)

4 Available at: http://www.ibm.com/big-data/us/en/ (Accessed on: November 16, 2015)

5 Available at: https://datascience.berkeley.edu/what-is-big-data/#JohnAkred (Accessed on: November 16, 2015)

6 Available at: http://www.gartner.com/it-glossary/big-data/ (Accessed on: November 16, 2015)

7 Available at: http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp (Accessed on: November 16, 2015)

8 Available at: http://siliconangle.com/blog/2014/08/19/gartners-hype-cycle-big-datas-on-the-slippery-slope/ (Accessed on: November 16, 2015)

9 Available at: http://www.contexti.com/blog/big-data-dropped-from-gartners-2015-hype-cycle-for-emerging-technologies/ (Accessed on: November 17, 2015)

10 2nd Oracle Academy Day was held in Zagreb, in February 2016. The theme was "Big Data".

Dario Šebalj
Ana Živković
Kristina Hodak

BIG DATA: PROMJENE U UPRAVLJANJU PODATCIMA

Sažetak

Živimo u digitalnome okruženju u kojemu gotovo sve što radimo ostavlja digitalni trag. Podatci se prikupljaju iz raznih izvora - objava na društvenim mrežama, e-pošte, različitih senzora, slikovnih i video sadržaja, pretraga na tražilicama, online kupovina i mnogo drugih. Najveći razlog ovakvoga rasta podataka može se naći u tehnološkom napretku budući da se danas podatci mogu lako i jeftino pohranjivati i dijeliti. Ovaj novi trend u generiranju i prikupljanju podataka zahtijeva potpuno novi pristup njihovoj obradi i analizi. Cilj je ovoga rada da se, na temelju analize aktualnih i relevantnih izvora, prikažu stanja i trendovi u prikupljanju, obradama, analizi i korištenju podataka koji su kompleksni, brzorastući, različiti po tipu i sadržajima. U uvodu je najprije definiran problem istraživanja, zatim je istražen i objašnjen pojam „*Big data*", a potom su objašnjene i njegove osnovne četiri dimenzije, odnosno gledišta. U središnjem dijelu rada prikazani su primjeri analize *Big data* i zanimljivih rezultata koje su te analize polučile, a poseban osvrt dan je na *Big Data* u menadžmentu.

Ključne riječi: *Big data*, menadžment, podatci