Ljiljana Ilić
Aleksandar Đorđević *
Milan Erić

BIG DATA AND OPEN DATA IN PUBLIC ENTERPRISES AS AN INITIAL STEP OF DIGITIZATION

Abstract: *The Big Data expansion is emerging alongside the development of information technology. Increasing the number of processes drastically increases the number of data. Data are in various forms from paper, audio, geological, digital and other forms. At the same time, computers are becoming increasingly accessible in all spheres of life. Computers take over the role of Big Data management, manipulating them, switching from one location to another and processing them. New technologies are being developed with the aim of managing data in order to improve business reporting, analysis and identifying opportunities to improve business. Therefore, digitization becomes an imperative in the data management strategy.*

Keywords: *Big data, digitization, public enterprises*

1. Introduction

Big Data (BD) and open data are new phenomena and are closely related, but not the same. Open data brings a perspective that can make large data more useful, more accessible and safer.

While BD is defined by size, open data is defined by its use. BD is a term used to describe very large, complex, and variable datasets. But these divisions are subjective and technology-dependent: today's big data may not look so great in a few years when computer technology improves (Brown et al., 2011, Yi et al., 2017).

Open data is available public data that people, companies and organizations can use to launch new ventures, analyze patterns and trends, make informed decisions, and solve complex problems. All definitions of open data include two basic characteristics: data must be publicly available to anyone who needs to use them, and must be licensed in a way that allows it to be reused (Zhang et al., 2015). Open data should be relatively easy to use, although there are gradations of "openness". There is agreement that open data should be available for free or at minimal cost.

Digitization is a process that transfers data from one form to digital form. Digitalization of business processes begins with a digital strategy that involves completely eliminating manual data entry and duplication of work. By setting the company's digital strategy, it is defined how the process of digitalisation of business will begin.

The number of data generated and shared by enterprises, public administrations, numerous industrial and non-profit sectors and scientific research has increased immensely (Agarwal & Dhar, 2014, Al Nuaimi et al. 2015). These data include textual content (ie structured, semi-structured and unstructured) on multimedia content (e.g., videos, images and audio) on many platforms (eg machine-machine communications, social networking sites,

* Corresponding author: Aleksandar Đorđević
Email: adjordjevic@kg.ac.rs

sensor networks, cyber-physical Systems and Internet Things [IoT]) (Chen et al., 2013). Dobre and Xhafa (2014) report that every day the world produces about 2.5 quintillion bytes of data (ie 1 exabyte is equal to 1 quintillion bytes or 1 exabyte equal to 1 billion gigabytes), with 90% of these data generated in the world that are not systematized. Gantz and Reinsel (2012) claim that by 2020, more than 40 Zettabits (or 40 trillion gigabytes) of data will be produced, imitated and consumed. With this huge amount of complex and heterogeneous data coming out of any place, at any time and any device, there is no doubt that this is the age of BD - a phenomenon also called Data Deluge. The potential of BD is obvious because it is included in Gartner's Top 10 Strategic Technological Trends for 2013 (Bahrami & Singhal, 2015) and Top 10 critical technological trends in the next five years (Barnaghi et al. 2013, Waller & Fawcett, 2013). This is as important as nanotechnology and quantum computing in the present era.

In everyday life there are more and more devices and services that help to realize all needs faster and more efficiently. There are: eGovernment, e-banking, eTrade, eHealth, eEducation ... These services are in different environments at different levels of development, but they are progressing everywhere and becoming more and more present in the everyday life of citizens.

Digitization will bring more and more dynamic changes into our lives. Many questions arise, such as whether artificial intelligence in the foreseeable future will overcome the natural and what will happen then, whether people will continue to compromise their personal data, whether privacy will disappear, and if the abuse of personal data can be effectively restrained, are robots self-reproducing and will one day estimate that a person is unnecessary, whether more and more artificial devices (cyborgs) will be installed in people, and whether they will continue to practice on the basis of the analysis of large data (now Facebook) affects the results of elections and making social decisions.

A large amount of data is generated in the public sector. Most are generated, distributed and shared with users of various agencies and public companies. These are mainly data that are not systematized and analyzed to be standardized. These data are very important for the government, since they represent data about citizens with social and other data that are important for analysis. Results of processing and data analysis can be very important because they can help in financial plans as well as the identification of certain groups of users, such as: social assistance, state subsidies and loans, and similar. On the other hand, such data are sensitive because of their importance, so it is necessary for them to be protected and processed, and to be accessible only by a target groups or to the government itself. It is precisely this part of the BD processing that must be done with software tools that are in service or that represent artificial intelligence (Bihani & Patil, 2014).

BD represents a large set of data which on its own do not offer standard values from which conclusions can be drawn. If all the potential of these data is used, with satisfactory methods and processes, values that can be used are obtained. In this context, the analysis refers to the methods used to analyze and process BD and can be used under processes that provide or draw conclusions from the BD (Hammouda & Karray, 2000, Sandhu & Sood, 2015). This creates the possibility for the BD data set to be systematized and unified as needed, thus switching to the Open Data Data Form.

2. BD application in Serbina enterprisses

Since 2010, an eGovernment portal has been set up in Serbia, creating the first serious step in digitizing state administration data.

After that, digitization began in all public companies of importance.

Open Data Infrastructure consists of data sets, organizations that use and maintain them, and provide guidelines on how to use them properly. In order to be relevant, these infrastructures must be sustainable and users have to give an insight how to use the data in a way that is most beneficial to the society as a whole. Infrastructures must include technology, processes and organization.

Thanks to the expansion of open data, many communities working in this area have changed their priorities - from gaining access to data, building new open data infrastructure. Like physical tangible infrastructure, such as roads, railways, power lines and telecommunications networks, which has been a driving force for the development over the past 100 years, data sets have become a vital part of our modern infrastructure

The general objective of Open Data is to encourage greater coherence and strengthen co-operation leading to the adoption and implementation of common principles, standards and good practice of open data from various sectors across the world.

A number of institutions, such as the Ministry of State Administration and Local Self-Government, the Ministry of Finance, the Statistical Office of the Republic of Serbia, the Secretariat for Public Policies, the Business Registers Agency and the Central Registry of Compulsory Social Insurance (CROSO), expressed great interest in opening their data. This opens up significant opportunities for large-scale projects, which in turn will provide experience and motivation for other potential participants in this project.

Predictive analysis allows connecting data from various sources such as the web, text data, and Call Center data. Sophisticated business techniques allow easy linking of data and reliable conclusions, taking into account the current situation and the ability to predict future situations. It is very useful to businesses because they can rightly conclude with the right conclusions to the current situation and also plan properly.

For this reason, solutions that would have such a good BD processing model are very important in the operations of each profit company. The correct predictive analysis influences how to make the right decisions and planning, and to reduce costs and raise company competitiveness.

In business systems where there are a large number of processes and documentation moving through the system, it is often necessary to consider all the elements of the business system in order to improve its functioning.

In the private sector there is a greater desire to regulate the system because processes differently move and the flows of money are extremely simple, because everything is focused on the ultimate goal and that is profit.

3. An example of digitization in Public Enterprises

In the public sector there are state and local administrations and business systems or companies that operate within the administration.

Digitization of the process within the enterprise or administration, regardless of whether it is a state or local administration, is the basis of the process and the flow of documentation overview. By examining the process and the flow of documentation, all the advantages and disadvantages of the public enterprise or administration are noted.

The first step in digitizing public enterprises must be the introduction of data management on all processes.

Digitalization of a public company should include the following:

- ePisarnicu
- archive
- procurement
- eProcesor that processes geo data
- eDocumentation that implies the

handling of documents, contracts and business processes, such as addressing clients' requests.

In day-to-day work in public companies, a large number of processes are carried out, aimed at providing services to citizens, other companies and administrations.

In Kragujevac, the first step in digitization is the formation of a unique base of subscribers of utility services and the establishment of the Unified Payment System. The advantage of a unique database is consistency and centralized updating of user data (Jiang et al., 2015, ur Rehman et al., 2016).

The digitization of state public administrations and the opening of their data, primarily the Ministry of Interior Affairs, the cadastre, the geographic information system, enables better up-to-date and harmonization with the data not only of the users, but also allows the updating of the space data by accessing the cadastre or GIS.

The effects of consolidating the collection of all services of utility companies can be shown in the diagrams.

A comparative analysis of the utility billing before and after the introduction of the Unified Billing System (UBS) shows the effects of this type of business (Figure 1 and 2).

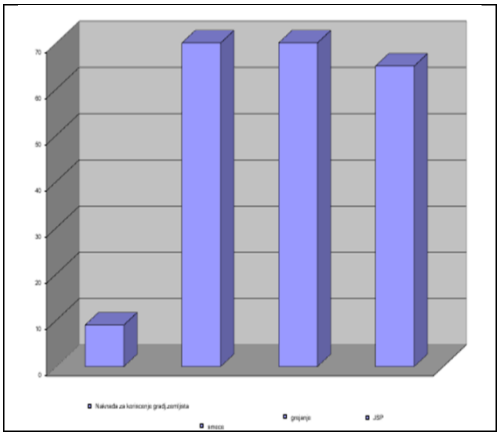

Figure 1. Payment for the utility services before the introduction of UBS

The strategy of utilities is to maintain regular communication with citizens. A global information revolution is in progress, aimed at improving cooperation on key social issues, providing more effective public oversight of the work of utilities, and encouraging innovation, sustainable economic development, efficient public policies in the field of ecology, and programs for improving the lives of individuals and society as a whole.

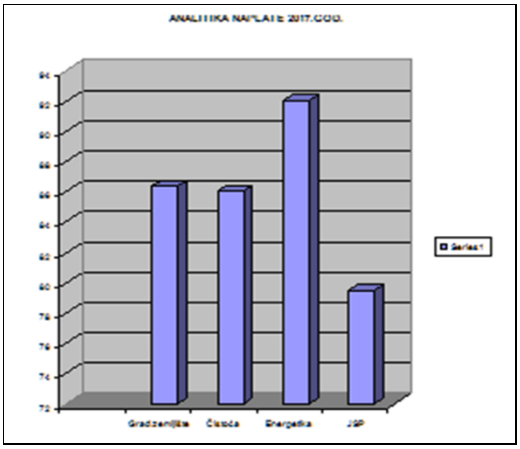

Figure 2. Payment for the utility services after the introduction of UBS

The effects of this digitization are multiple. There is a reduction and rationalization of costs, optimization of business processes, more efficient work of services and interconnection.

After centralizing the database of utility services, the next step would be the introduction of data management in the field of manipulation of paper documents. Regular ePisarnica would be centralized at the level of all Public Enterprises in local self-government.

A centralized office would allow data management in utility companies, as well as their better connections with citizens.

Introduction of CRM (Customer Relationship Management) would also facilitate the introduction of a central ePisarnica, as the flow of information would be transparent. At any time, it would be possible to send documents to public companies and document status (Gurin, 2014).

This way of business would open the possibility of using BDA (Business Data Analysis) tools for analysis and business improvement.

The results of the processing of the obtained data have the following effects:

- Analysis of UBS and perception of the integration of utility services
- Forecast of economic and technological development of the system
- Identifying trends in CRM business
- Support management
- Strategic planning
- Technological improvement of UBS.

All of these elements contribute to the creation of a management model for utility services.

Using the BDA tool set, a management model is created.

Figure 3 shows the scheme of centralized ePisarnica as a proposal for possible improvement of work and the next step in the process of digitization of public utility companies in Kragujevac.

The benefit of the electronic office system is multiple in the digitalization era because the results of the work can be measured and the real needs of the business system are observed. This business prevents corruption and raises the efficiency of the business system in general.

The main advantage is the faster functioning of the office and the improvement of its basic function of classifying documents at the entrance to the business system. Each document is electronically monitored to resolve the case or archiving. Thus, for each document in the system, the "Life cycle" is recorded.

The notion of digitization and systematization of data contributes to the efficient operation of the business system.

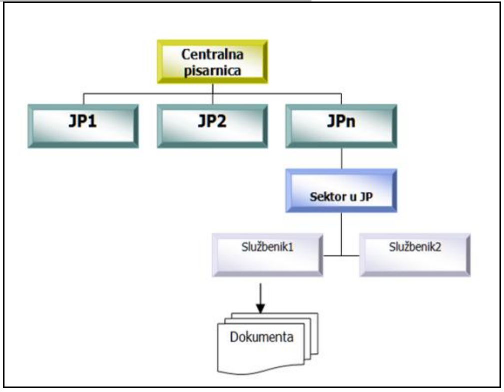

Figure 3. ePisarnica structure scheme

4. Conclusion

This systematic processing and storage of data allows primarily easy and fast locating of documents in the business system, which also gives the possibility to manage data processing.

Through the strategic plans of public companies, it is necessary to anticipate changes in the way of doing business in the economic sense, as well as in terms of upgrading the IT system and the development of Hardwer. All of these are harmonized and implemented in Public Enterprises in relation to the core business.

After the digital transformation of the data and the business systems itself, the introduction of intelligent business is carried out with the help of predictive analytics and data management.

The business system that performs the entire digitization becomes part of a smart city (Gharaibeh et al., 2017).

Using the latest technologies after digitization and intelligent business development, the future of this area is the development of robotics and artificial intelligence that would be used to perform some activities.

There are jobs in the communal area where people can be replaced by robots. In addition to technical and technological development, the costs would be reduced, the number of work-related injuries would be reduced to a

minimum, the number of workers in the field and other.
In such a duty, the application of predictive analysis, the development of Data Mining and Business Intelligence would represent a set of tools that would improve the quality of life of both society and the individual.

Recognizing the progressiveness and speed of digitization in all segments of life, starting with technological production, it is imposing, through the digitization of data in business information systems, the conclusion that smart cities represent the future.

References:

Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research* 25, 3, 443-448.

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal Of Internet Services And Applications*, 6(1). doi: 10.1186/s13174-015-0041-5.

Anon., (2018). [online] Available at: https://www.euprava.gov.rs/ [Accessed 5 Dec. 2018].

Bahrami, M., & Singhal, M. (2015). The role of cloud computing architecture in big data. In Information granularity, big data, and computational intelligence (pp. 275-295). Springer, Cham.

Barnaghi, P., Sheth, A., & Henson, C. (2013). From data to actionable knowledge: big data challenges in the web of things. IEEE Intelligent Systems, (6), 6-11.

Bihani, P., & Patil, S. T. (2014). A comparative study of data analysis techniques. International journal of emerging trends & technology in computer science, 3(2), 95-101.

Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'. McKinsey Quarterly, 4(1), 24-35.

Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S., & Zhou, X. (2013). Big data challenge: a data management perspective. Frontiers of Computer Science, 7(2), 157-164.

Dobre, C., & Xhafa, F. (2014). Intelligent services for big data science. *Future Generation Computer Systems*, 37, 267-281.

Gharaibeh, A., Salahuddin, M. A., Hussini, S. J., Khreishah, A., Khalil, I., Guizani, M., & Al-Fuqaha, A. (2017). Smart cities: A survey on data management, security, and enabling technologies. IEEE Communications Surveys & Tutorials, 19(4), 2456-2501.

Gurin, J. (2014). Open data now: the secret to hot startups, smart investing, savvy marketing, and fast innovation. McGraw Hill Professional.

Hammouda, K., & Karray, F. (2000). A comparative study of data clustering techniques. University of Waterloo, Ontario, Canada, 1.

Jiang, H., Chen, Y., Qiao, Z., Weng, T. H., & Li, K. C. (2015). Scaling up MapReduce-based big data processing on multi-GPU systems. Cluster Computing, 18(1), 369-383.

John Lu, Z. Q. (2010). The elements of statistical learning: data mining, inference, and prediction. Journal of the Royal Statistical Society: Series A (Statistics in Society), 173(3), 693-694.

Sandhu, R., & Sood, S. K. (2015). Scheduling of big data applications on distributed cloud based on QoS parameters. Cluster Computing, 18(2), 817-828.

ur Rehman, M. H., Chang, V., Batool, A., & Wah, T. Y. (2016). Big data reduction framework for value creation in sustainable enterprises. International Journal of Information Management, 36(6), 917-928.

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, 34(2), 77-84.
Yi, X., Liu, F., Liu, J., & Jin, H. (2014). Building a network highway for big data: architecture and challenges. IEEE Network, 28(4), 5-13.
Zhang, F., Liu, M., Gui, F., Shen, W., Shami, A., & Ma, Y. (2015). A distributed frequent itemset mining algorithm using Spark for Big Data analytics. Cluster Computing, 18(4), 1493-1501.

Ljiljana Ilić
Javno stambeno preduzeće Kragujevac,
Kragujevac,
Serbia
lj_ilic@yahoo.com

Aleksandar Đorđević
Faculty of Engineering
University of Kragujevac,
Kragujevac,
Serbia
adjordjevic@kg.ac.rs

Milan Erić
Faculty of Engineering
University of Kragujevac,
Kragujevac,
Serbia
meric@kg.ac.rs