



## Research Article

© 2018 Elezaj et al.  
This is an open access article licensed under the Creative Commons  
Attribution-NonCommercial-NoDerivs License  
(<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

# Big Data in e-Government Environments: Albania as a Case Study

**Ogerta Elezaj**

*Faculty of Economy,  
University of Tirana*

**Prof. Dr. Dhimiter Tole**

*Faculty of Economy,  
University of Tirana*

**Prof. Assoc. Dr. Nevila Baci**

*Faculty of Economy,  
University of Tirana*

**Doi: 10.2478/ajis-2018-0052**

### **Abstract**

*Nowadays, the new trend in e-government environment is Big Data which is considered to be one of the most critical issues due to the managing, quality and privacy challenges. Creating and storing the huge volume of structured and unstructured data generated by governments in different formats does not provide any benefit for decision-making. Governments have to benefit from this data deluge to use its different aspects such as fraud detection or to measure citizen's needs and desire for services in different area such as finance, healthcare, environment, economic and social statistics, etc. The objective of this paper is to bring Big Data in the context of e-government environment addressing main challenges that governments faced and describing opportunities in this field. This research summarizes the current situation of big data initiatives in e-government in different countries. This paper presents the current situation regarding digital transformation and e-government environment in Albania. The e-government development and main statistics in the e-services it offers to Albanian is described as it will provide vital input for the potential of using big data as no current national initiatives exist on this field, in the country level. In this paper, we identify challenges and opportunities to start using Big Data in public organizations in Albania and we analyze methodological, organizational and processing challenges when big data is used, instead of traditional sources.*

**Keywords:** *big data, interoperability, e-government, analytics*

## **1. Introduction**

Nowadays, digital technologies are transforming the social and economic aspects of public and private institutions. Digitalisation, social media and the increased use of electronic devices are considered as the main reason for the increase of the data created and generated by public and private institutions. Relationships with banks, suppliers, customers are increasingly continuous (non-stop) and automated, where the number of transactions that are registered online has increased considerably. The changes made by Information and Communications Technology (ICT) require the emergence of radical innovations of traditional systems for data storage and data processing. Thus, for example, through the use of ICT in the health field, there is a change in paradigm shifting, from a reactive system with the focus on disease cure into a proactive system to

prevent disease. In addition, the education system is undergoing radical change through the use of distance learning platforms (E-learning) that use the Internet. Increase use of equipment, electronic networks and the digitalization of data processing processes generates a huge amount of data from the economic and social activity that people perform.

Base on some data deluge forecast, the amount of data produced worldwide is doubled every two years; is expected to increase from 4.4 zettabytes (or 4.4 trillion gigabytes) in 2013 to 44 zettabytes in 2020 (Cyclone, 2014). These huge data come from different heterogeneous sources such as the web, commercial online transactions, e-government data, social media, cellular) data, mobile applications, and the internet of things.

In 2013, only 22% of the information in the digital universe would be valuable and useful to analyze, as metadata is missing. By 2020, the percentage of valuable information to be analyzed is estimated to reach more than 35%, mainly due to the increase in investment in technical and human capacity in the Big Data field. The world market in Big Data technology is expected to mark an annual growth of 23% in 2014-2019, while worldwide revenue for Big Data will grow by 50% from 122 billion \$ in 2015 to 187 billion \$ (WordWide Big Data, 2016). The sectors with the greatest weight in this data revolution are the manufacturing sector, the banking sector, telecommunications, health, transport and the public sector.

The potential of using Big Data in the private sector is widely explored and many researches exist (Chen et al., 2015). Big data solutions have attracted interest in the private sector, but on the other hand, there is not a clear interest in applying them in public institutions, leaving a gap and emergent need for research of putting it in the public sector (Desouza et al., 2014). Big Data offers a potential to address many challenges that all public organizations face today, such as efficiency, transparency, and productivity (Kim et al., 2014). The purpose of this paper is to contribute to the discussion of the application and challenges of Big Data in government environments and its potential in Albanian public institutions for improving customer services. The article is structured as follows: Section 2 gives an overview of big data initiatives in public sector by European Commission. Description of big data sources, classification, and challenges is presented in section 3. Section 4 presents the statistical implications when traditional data sources are substituted by big data. Section 5 presents the current Albanian e-government infrastructure and interoperability framework to assess the potential of starting a national big data initiative. Section 6 presents the conclusions and discusses the further work.

## 2. Big Data Initiatives

Government agencies use technology innovations to increase openness, efficiency, citizen satisfaction. Their goal is to be service-oriented and they are aware of the potential of data collection, processing and analyzing as follow:

- Open government and data sharing: Providing accurate and fast information from public organizations to citizens is a requirement for greater government transparency, promoting greater trust between citizens and government, in line with open data initiatives.
- Citizen sentiment analysis: It is considered as government new eye where information from both traditional and new social media is opening new opportunities to achieve transparency and citizen's engagement.
- Economic analysis: Correlation, pattern recognition and rich analytics of data coming from multiple sources help policymakers in making decisions based on knowledge.
- Tax agencies: Integration and matching of structured and unstructured data, data coming from different sources, has a big impact on potential fraud detection.
- Smart city and Internet of things (IoT) applications: Based on technology innovation, public institutions all over the world are using sensors to measure physical phenomena such as traffic volume, the location of vehicles, etc. Analyzing this huge amount of data which are characterized by velocity, gives to the public institutions the potential to improve urban management with the main aim of improving quality of life for their citizens.
- Cybersecurity: Collecting and analyzing data coming from government network computers

and their logs increase the ability of institutions to detect and prevent malicious attacks.

Historically, the public sector often develops many disconnected systems trying to serve a single citizen population. Unlocking and connecting these systems can enable greater coordination and integration across agencies providing more efficient, effective, and seamless services to citizens. Many initiatives are taken by governments worldwide to transform services offered to citizens and businesses using a whole-of-government approach.

Interoperability is considered as the most important key element for transforming e-government. Interoperability is defined as the connection of diverse data, systems, people and processes, which gives to the governments the possibility to combine services for new systems. Also, by interoperability, governments can benefit from connecting existing infrastructure with the new technological environment. But when we speak about interoperability, it should not be treated on as a technical aspect, because it is more a political issue. Leaders of different governments are discussing the need for better interoperability to promote data sharing among different systems across the country.

In public sector the concept of Big Data is new, and it is considered as a paradigm to switch to smart government. It can reduce costs, improve decision-making and decrease the time needed to process the data (Yusifov et al., 2016). Large volume of data generated by e-government systems has to deal with problems mainly related to the data semantics. Information shared among different systems, which are considered as Big Data sources, should be done in a meaningful and effective way to achieve semantic interoperability. Different countries such as United States, Brazil, Germany, and Estonia have developed their Government Interoperability Framework (GIF) (Arch-int et al., 2013). Integration of data coming from multiple systems is the key to unlocking big data potential to enhance decision-making.

European Commission considers Big Data as a key economic asset, which provides a competitive advantage. There are around 20 companies dominating the market in Big Data which are making big changes offering a new paradigm for data storing and processing. Only two of these companies are European companies, and this happens because Europe has to invest more in the data value chain. A big project, launched in 2005, which costs around €500million, has started which is a Public Private Partnership agreement, among European Commission, academia, researcher, and industries. The goal of this project is to initiate actions and contribute to the development of IT infrastructure and applications that support Big Data, to improve decision-making. The benefits of these projects are :

- 30% of global data market will be available;
- 100.000 new jobs in Europe by 2020;
- 10 % lower energy consumption and better health-care

### 3. Big Data Sources, Classification, and Challenges

Big Data sources are heterogenous and are collected from various applications such as: sequential or time series data (historical records); continuously growing data streams (video data, data received by different sensor); geospatial data; multimedia files (text, audio, and video); social networks data; web data, etc.

In 2013 UNECE has classified Big Data sources in three main categories: 1. Social networks: unstructured data created by people which are digitized and ubiquitous; traditional business systems: data coming from processes where business events are recorded and monitored, such as registering a customer, producing a product, receiving an order, etc. These data are structured data and are composed of transactions, reference tables, and relationships, as well as metadata that determine their context (Devlin et al., 2012). This category of data includes data that a company manages and processes, both in operational systems and in business intelligence systems and their main characteristics is being structured and mainly stored in relational database systems. These include administrative data, which are defined as data collected by or on behalf of central and local institutions, for administrative purposes, in accordance with the legal and sub-legal acts on which their activity is based; 3. The Internet of Things (IoT) - the new technology development philosophy, based on smart objects that help to interact and improve people's lives. Every subject

of everyday life from the car to the gas supplier may have, through networks, a digital identity, so it becomes more functional, more economical and efficient, where "intelligent objects" reach to share information about themselves and the environment that surrounds them. "Smart" sensors increase the number of things that can be connected, but naturally also provide control over the information. Opportunities and challenges of using Big Data in government institutions are matter of an open debate because of the complexity of data management and the nature of the products services government offers to their citizens and businesses. Below an overview is given of the main challenges.

- **Big data management**  
These challenges include data processing challenges, managing data across organizations and an absence of the technologies and expertise required to manage large amounts of data. Traditional statistical analyses are based on target populations, structured data and the methods for data processing are based on sampling theory. Since Big Data are mostly unstructured and in different formats such as video, image, textual data, techniques of information retrieval such as data mining, data reduction and artificial intelligence should be considered to be used by official statisticians to cope with these challenges (Darren et al.,2014). Retrieving, storing, processing and transferring the huge volume of data sets is a challenge. Technological innovations like high performance computing, storage facilities and high bandwidth data channels may partially solve these issues. As data confidentiality and security is a top priority in the organizations, using cheap cloud-computing solutions is not the best option to be considered by them. People working groups in big data projects should be composed by members with multidisciplinary skills from different professional backgrounds. Statistical agencies, banks and public agencies should train their staff.
- **Access, ethical and privacy concerns**  
In majority of the cases, the owners of Big Data are outside of the control of national or international institutions. The data are held by private sector and accessing and using them requires memorandum of understanding to be agreed with data owners where confidentiality and data privacy should be well specified. AS Big Data contain personal data, institutions should invest IT infrastructure adapting security techniques such as cryptography, data anonymisation and statistical disclosure control (Moreno et al., 2016). Furthermore, there is a conflict between demands for privacy and those for more openness and transparency (Henninger, 2013).
- **Data quality**  
The quality of indicators produced using Big Data should be assessed to ensure that the new indicators produced meet the minimum data quality standards for real fiscal, monetary, financial or other statistics. Most of the data coming from Big Data is unstructured and proper data manipulation techniques should be applied to transform them into time series data. The processing means cleaning, outlier detection and treatment, imputation of missing variables which should be done in line with tradition statistical methods to ensure data quality. Still, there is a lot to be done regarding methodology of Big Data processing to adhere data quality standard, but could still uncover insights by alerting that something is happened. For example government based on social network data, using sentiment analyses can draw conclusions regarding different social-economic trends or opinions.

#### 4. Statistical Implications

Big Data plays an important contribution to the production of official statistics in producing social-economic indicators. Traditionally, providers of official statistics use administrative data, censuses, and surveys data as main data sources. Recently, many international statistical institutions worldwide have taken initiatives to accommodate big data sources as a potential source to achieve their main goal in producing relevant and timeless statistics. While these data sources will continue to be important, efforts should be made to integrate them with unstructured data coming from big

data sources, to provide more relevant and insightful statistics that are vital to be used in decisions. Big data provide the following opportunities for policymaking: i) producing new indicators, ii) innovative data sources in production of statistics, iii) forecasting trends of existing indicators.

Big Data looks appealing to the national institute of statistics (NIS) because of the possibility to improve data quality, reduce response burden and give a new possibility for producing new indicators. Till now, NSIs use a top-down traditional paradigm where data are planned and collected based on the needs to produce specific indicators. This paradigm is dealing with traditional inferences approaches such as sampling survey methodology and model-based inference. Using Big Data as a data source requires an emergent need to move to a next paradigm, bottom-up, where data exist. The methods to be used are explanatory analyses and knowledge discovery approach. These methodological issues are considered as the main challenge faced by data scientist.

**Table 1:** Comparing small, administrative and big data sources

	Primary Sources (Censuses, Surveys)	Administrative sources	Big Data sources
Data designed for statistical purposes	yes	no	no
Concepts, definitions and classification well-specified	yes	often	no
Population is known and defined	yes	often	no
Metadata available	yes	often	no
Data structured	yes	often	rarely
Data refer to units of the population of interest	yes	yes	no
Advance Pre-processing methods need to be used	no	no	yes
Interest variables are directly available	yes	yes	no
Data cover target (sub)- population	yes	often	not yet
Data are representative	yes	often	no
Data values are clean	no	sometimes	rarely

Data processing methodology is a huge challenge for NSIs in processing Big Data sources because official statistics are based on sampling theory. The data processing follows a well-defined workflow, a target population is identified, a sample is selected, data are collected and are processed. The sampling units are selected using survey sampling methods to ensure representatives. On the other hand, Big Data sources are a continuous flow of unstructured data and extracting useful information from them is difficult and requires machine learning algorithms and data explanatory techniques to be used. Sub-population of Big data is not target populations. For example, the data collected from online social networks do not represent the whole structure of the society. To cope with this discrepancy, the target population characteristics should be compared with the covered population. In practice, sometimes this is not easy or even possible. For example, data coming from online social networks often cannot identify persons because they are registered with usernames (Daas et al., 2015).

## 5. Albania E-Government and Potential for Big Data Use

With the latest information technology development, governments are increasingly relying on computerized information systems for monitoring and decision-making. Information and Communication Technologies (ICT) play an important role in achieving a strategic goal for a more efficient government. E-government refers to the use of modern technology resources such as the internet, mobile, etc. to improve e-services provided to the citizens and businesses. The main benefit of technological systems of e-government is seen in building an open information society by delivering public services on the Internet.

E-government is an instrument of an information society in the form of governing principles, strategies, systems and tools that enable the use of ICT in mutual interaction between key elements of society - the government, citizens, and businesses –to strengthen democracy and support development.

The government portal e-albania.al is developed and is administered by the National Agency

for Information Society (NAIS) as a multifunctional portal, and it is considered as one stop shop where citizens and businesses access public electronic services. It provides services 24 hours, 7 days a week. The portal started as a European Union investment in 2009. At the early stage, it was very simple including 6 electronic services and 4 systems linked to the government interoperability platform. The portal is linked to the Governmental Interoperability Platform, which is the basic architecture that combines hardware, software, and services enabling the interaction between all the connected systems of the government institutions.

In the E-Albania portal, there are already listed more than 571 services that government institutions currently offer to the citizens and businesses. Some of these services are transactions and provided through different online systems.

In April there are around 500,000 users already registered on the portal E-Albania. This is not just a number, but a clear indicator that reflects the usefulness of a multitude of electronic services, which Albanian institutions offer over the internet for their citizens and businesses.

In total, only during the first three months of 2018, 114,000 new users have been registered, showing a positive trend. E-Albania is offering hundreds of electronic services for citizens and businesses too. Many of these services such as certificates and the health card have facilitated their lives by saving them at all costs and time. Many public institutions which are responsible for offering these online services interact and exchange real-time data between them using the interoperability framework. During 2016-2017, over 50 million transactions have been recorded among interconnected systems. The existence of this platform is why today in Albania 30 different documents can be downloaded directly from the portal equipped with a digital seal. The impact of this development has been tangible and significant in reducing paper documents.

In May 2018, in e-Albania portal, there were 25403 newly registered users, and more than half of them were using mobile applications. The total number of e-services provided to them is more than 300.000, illustrated in table 2.

**Table 2-** e-Albania statistics, May 2018

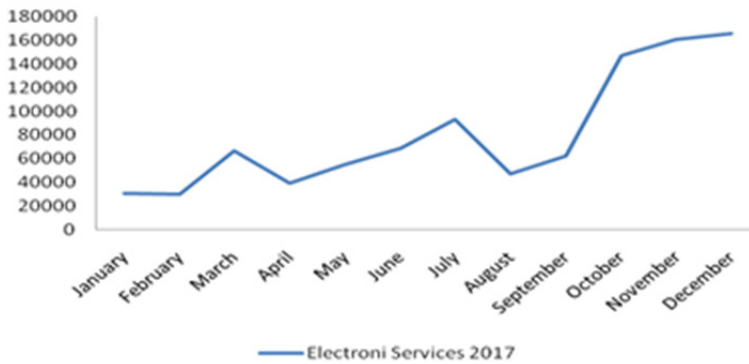
Visited pages 2,151,718	New registered users 25,403	Mobile app users 14,880	Total number of eservices used 346,502	Services 1,341	Application for Construction Permit 1,251
----------------------------	--------------------------------	----------------------------	---	-------------------	--

- The most used services on the portal are presented in table 2. The users that are registered in the portal receive an extract in the form of a certificate, which can be family or personal certificate. The data are stored in the national database, civil register, which is owned by the General Directorate of Civil Status, Ministry of Interior.

**Table 3:** 10 most used services on the portal-source

Nr	Services	Uses
1	Family Certificate	100,584
2	Generating the Health Card	31,298
3	Proof of payment of contributions to the individual	26,425
4	Personal Certificate	23,792
5	Request for an unregistered individual attestation	10,853
6	Application for Construction Permit (e-Permit)	10,692
7	Confirmation of the vehicle's active condition	9,428
8	Attestation of status (active, passive, including date)	8,221
9	Declaration and payment of contributions by employers	7,707
10	Application for a passport and identity card	6,926

This information constitutes a national asset and should be managed and analyzed in a way that can be used not only for ensuring government transparency but also in increasing national economic benefits.



The access to the e-Albania platform generate large volumes of data and collecting different data such as length time spend on the page, frequency of visits, number of transactions completed per visit, user location and so on. Analyzing this information by doing web analytics, the policymakers can develop metrics for monitoring the effectiveness of the e-Albania platform.

Law No. 10 325, of 23 September 2010 On State Databases, has created the initial legal basis for information-interchange between different Government agencies. However, the realization of the information-interchange in practice necessitates the regulation of information services provided with the help of state databases. In total there are 74 state databases from which 45 are registered as state databases, and the others are in process. Despite the fact that many of the government institutions in our country digitized their systems, in many cases, these systems have been developed without a communication strategy among them, with the only purpose to meet the internal information needs. The current technological transformations and the needs for quick, accurate, transparent and up-to-date information for citizens and businesses, require that these systems interact with each other sharing large volumes of data. All the data part of in e-government can't be processed using traditional analytic methods because of their characteristics, big volume, processing speed, and diversity.

## 6. Conclusions

The next trend of e-government institutions will be data analytics and governments worldwide have to address challenges and potentials of using Big Data as it offers substantial value for them to become more efficient, transparent and democratic (Joseph, 2013). By using Big Data, governments can drastically improve decision-making, based on more information which is available and analytic techniques used to get faster insight from processing it in real time. Based on the analysis of technological developments of the Albanian government, we conclude that there are no initiatives for the use of the Big Data and no technology is being developed to respond to these huge data capacities. NAIS's efforts to reduce the margin of error in the Big Data analysis by guiding users to the right path are inadequate. NAIS does not use Big Data to help decision-makers make the right decisions based on data quality assurance and analysis. Today, we still continue to lose data that are not collected or making unusable huge data sets by not processing per day, week, month, precisely because of the missing adequate Big Data infrastructure which results in a weaker administrative capacity of the country, fewer opportunities to have extensive knowledge of society. The government should make a data inventory, strategies for linking different data coming from different systems and prepare strategies and plans before taking initiatives to Big Data. Policymakers in Albania should consider Big Data as a step to move to smart government development and not just a data initiative.



## References

- Chen, Jidong & Tao, Ye & Wang, Haoran & Chen, Tao (2015) Big data based fraud risk management at Alibaba, *The Journal of Finance and Data Science*, 1 (1): 1-10.
- Daas, Piet JH, Marco J. Puts, Bart Bulenes, and Paul AM van den Hurk. "Big Data as a Source for Official Statistics." *Journal of Official Statistics* 32, no. 2 (2015): 249-262.
- Darren & Choo, Kim-Kwang Raymond (2014) Impacts of increasing volume of digital forensic data: A survey and future research challenges, *Digital Investigation*, 11 (4): 273-294.
- Desouza, Kevin C. & Jacob, Benoy (2014) Big Data in the Public Sector: Lessons for Practitioners and Scholars, *Administration & Society*: 1-22.
- Devlin, B., 2012. The Big Data Zoo-Taming the Beasts: The Need for an Integrated Platform for Enterprise Information. 9sight Consulting.
- Henninger, Maureen (2013) The Value and Challenges of Public Sector Information, *Cosmopolitan Civil Societies: An Interdisciplinary Journal*, 5 (3): 75-95.
- Cyclone Interactive Multimedia Group, Inc., & Cyclone Interactive Multimedia Group, Inc. (n.d.). The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. Retrieved from <https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>
- Worldwide Big Data and Business Analytics Revenues Forecast to Reach \$187 Billion in 2019, According to IDC. (2016, May 23). Retrieved from <https://www.businesswire.com/news/home/20160523005153/en/Worldwide-Big-Data-Business-Analytics-Revenues-Forecast>
- Joseph, R. C., & Johnson, N. A. (2013, November). Big Data and Transformational Government. *Computing Now*, pp. 43-48.
- Kim, Gang-Hoon & Trimi, Silvana & Chung, Ji-Hyong (2014) Big-data applications in the government sector, *Communications of the ACM*, 57 (3): 78-85.
- Moreno, J., M. A., Serrano, and E. Fernández-Medina. (2016). "Main Issue in Big Data Security." *Futur. Internet*, vol. 8, no. 3, p. 44, 2016
- N. Arch-int and S. Arch-int, "Semantic Ontology Mapping for interoperability of Learning Resource Systems using a rule-based reasoning approach," *Expert Syst. Appl.*, vol. 40, no. 18, pp. 7428-7443, Dec. 2013.
- Yusifov, Farhad. "Big Data in e-Government: Issues, Opportunities and Prospects." In *ECEG2016 Proceedings of 16th European Conference on e-Government ECEG 2016*, p. 352. Academic Conferences and publishing limited, 2016.