



Big Data in Government: An Introduction

Matthew N. O. Sadiku¹, Uwakwe C. Chukwu², Abayomi Ajayi-Majebi³, Sarhan M. Musa¹

¹Roy G. Perry College of Engineering, Prairie View A&M University, Prairie View, TX, USA

²Department of Engineering Technology, South Carolina State University, Orangeburg, SC, USA

³Department of Manufacturing Engineering, Central State University, P.O. Box 1004, Wilberforce, OH, USA

Email: sadiku@ieee.org; uchukwu@scsu.edu; ajayi-majebi@centralstate.edu; smmusa@pvamu.edu

Abstract Data may be regarded as the new form of currency. It is now the lifeblood of twenty-first century government. Data is a critical asset for government agencies. It can have an enormous impact on government at local, state, national, and global levels. It allows governments to make faster decisions, monitor those decisions and make changes if necessary. Big data is a term that is widely used to describe the exponential growth of data. Big data technology is vitally important for governments all over the world. It can transform government and society. This paper is an overview on the application of big data in governments. Although the material covered in this paper is limited to US government, the information is applicable to other nations.

Keywords big data, data analytics, government, big data governance, big data management

Introduction

Data is the new gold in the digital era. It has become a valuable commodity just as important to production as labor, capital, and land. Data volumes have skyrocketed in our society and all over the world. Billions of people interact with their devices via the Internet and generate a lot of data every day. Some customers are ingesting up to 5 petabytes of data daily. Data continues to be generated and archived at growing rates across all sectors, including the government. As illustrated in Figure 1, big data is everywhere [1]. The amount of data generated worldwide is exploding. These large volumes of structured and unstructured data are known as big data (DB).



Figure 1: Big data is everywhere [1]



Big data refers to gathering, analyzing, and storing of complex data sets that are too large for traditional data processing and data management [2]. One of the major strengths of big data is its flexibility and universal applicability to so many industries. Industries that are improving their services through big data include healthcare, manufacturing, agriculture, telecommunications, retail, fitness, travel, finance, e-commerce, entertainment, banking, oil and gas, sports, social media, and government. Today, big data is synonymously known as business intelligence, business analytics, and data mining. It has shifted business intelligence from reporting and decision support to prediction and smart decision making [3]. As illustrated in Figure 2, four big data use activities have been identified as collection, combination, analytics, and use [4].

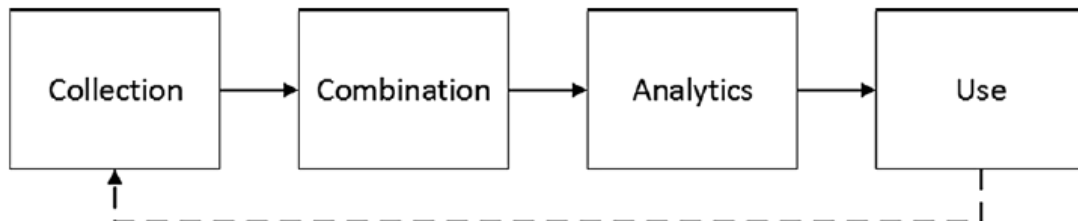


Figure 2: Big data use process [4]

The emergence of big data has led to the use of digital technologies in the delivery of public services and for realizing digital governance. Digital transformation has recently become a major issue to governments worldwide. Governments generate and collect vast quantities of data through various activities such as salary payments, pension management, tax collection, recording traffic data, and issuing official documents. Government data generation is increasing due to the rapid growth in the use of laptops, tablets, smart phones, and embedded sensors in government. As government leaders realize that information is a strategic asset, they try to protect, leverage, and analyze both structured and unstructured data to better serve and meet mission requirements. Government organizations are now deploying big data technologies to analyze massive data sets in science and research. They expect big data to help their ability to address national issues such as the economy, healthcare, education, job creation, national security, natural disasters, and terrorism.

Review on Big Data

Big data (BD) refers to a collection of data that cannot be captured, managed, and processed by conventional software tools. It is a relatively new technology that can help many industries, including government. The three main sources of big data are machines, people, and companies. Big data can be described with 42 Vs [5]. The first five Vs are volume, velocity, variety, veracity, and value [6].

- **Volume:** This refers to the size of the data being generated both inside and outside organizations and is increasing annually. Some regard big data as data over one petabyte in volume.
- **Velocity:** This depicts the unprecedented speed at which data are generated by Internet users, mobile users, social media, etc. Data are generated and processed in a fast way to extract useful, relevant information. Big data could be analyzed in real time, and it has movement and velocity.
- **Variety:** This refers to the data types since big data may originate from heterogeneous sources and is in different formats (e.g., videos, images, audio, text, logs). BD comprises of structured, semi-structured or unstructured data.
- **Veracity:** By this, we mean the truthfulness of data, i.e. whether the data comes from a reputable, trustworthy, authentic, and accountable source. It suggests the inconsistency in the quality of different sources of big data. The data may not be 100% correct.
- **Value:** This is the most important aspect of the big data. It is the desired outcome of big data processing. It refers to the process of discovering hidden values from large datasets. It denotes the value derived from the analysis of the existing data. If one cannot extract some business value from the data, there is no use managing and storing it.

On this basis, small data can be regarded as having low volume, low velocity, low variety, low veracity, and low value. Additional five Vs has been added [7]:

- **Validity:** This refers to the accuracy and correctness of data. It also indicates how up to date it is.



- *Viability*: This identifies the relevancy of data for each use case. Relevancy of data is required to maintain the desired and accurate outcome through analytical and predictive measures.
- *Volatility*: Since data are generated and change at a rapid rate, volatility determines how quickly data change.
- *Vulnerability*: The vulnerability of data is essential because privacy and security are of utmost importance for personal data.
- *Visualization*: Data needs to be presented unambiguously and attractively to the user. Proper visualization of large and complex clinical reports helps in finding valuable insights.

Figure 3 shows the 10V's of big data. In addition, the 10V's above, some suggest the following 5V's: Venue, Variability, Vocabulary, Vagueness, and Validity) [8]. The future of big data will bring more Vs.

To thrive in today's complex business environment, businesses must adopt a data-driven culture and leverage analytics platforms to make key decisions that improve productivity. Industries that benefit from big data include the healthcare, financial, airline, travel, restaurants, automobile, sports, agriculture, manufacturing, hospitality industries, and government.

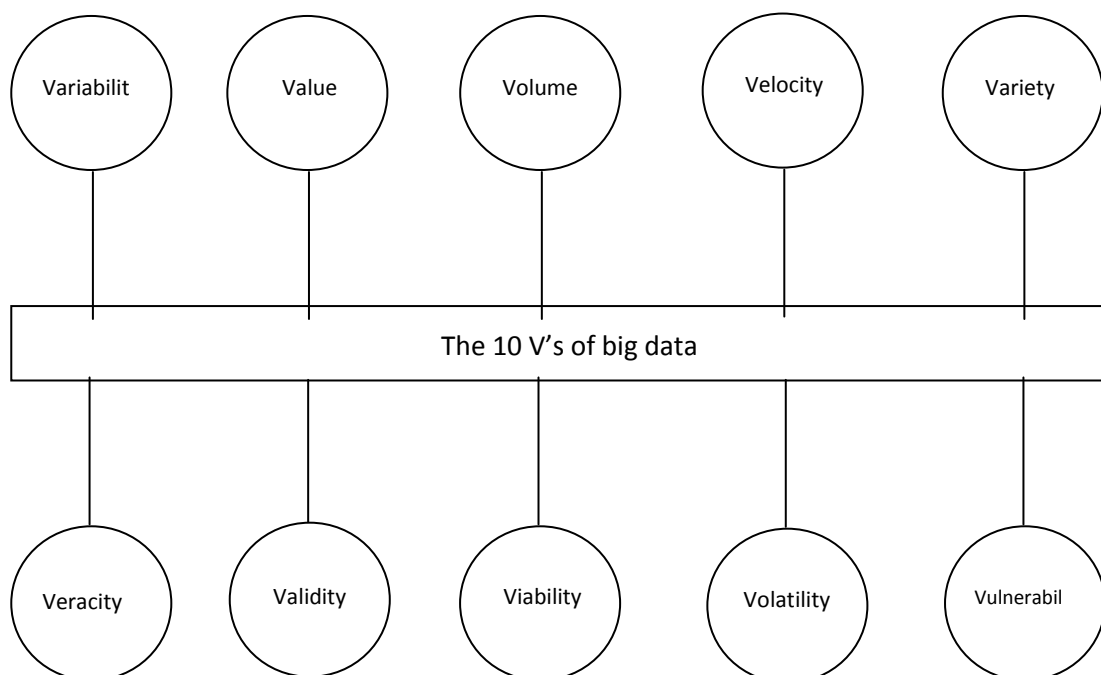


Figure 3: The 10V's of big data

Big Data Analytics

Big data sets can be staggering in size so that its analysis is daunting. Every day, data is growing bigger and bigger, and big data analysis (BDA) has become a requirement for gaining invaluable insights into data such that companies could gain significant profits in the global market. Big data analytics can leverage the gap within structured and unstructured data sources. Once the big data is ready for analysis, we use advanced software programs such as Hadoop, MapReduce, MongoDB, Spark, Cassandra, Apache Storm, and NoSQL databases [9]. Big data analytics refers to how we can extract, validate, translate, and utilize big data as a new currency of information transactions. It is an emerging field that is aimed at creating empirical predictions. Data-driven organizations use analytics to guide decisions at all levels [10].

Big data analytics is capable of processing massive amounts of dirty data and extract the gold information from it. It basically learns from data to predict the way individuals will behave in the future. Governments have adopted evidence-driven policy-making through innovative and efficient use of big data analytics. Big data analytics has the potential to transform the way in which government operates, delivers public services, and



engages with the citizens. It can help the government achieve greater efficiency, combat fraud, and boost productivity and growth. Governments can also use big data analytics to unlock key information and improve transparency. The use of data analytics delivers significant gains in the following ways [11]:

- *Cost Reduction:* Using big data technologies like Hadoop or cloud-based analytics allows organizations to store large amounts of data in a cost-effective, efficient manner.
- *Better Decision-Making:* Big data's primary value comes from its ability to facilitate smarter decision-making.
- *Improved Products & Services:* Big data analytics applications also allow companies to come up with new and improved solutions and products.

Big Data Governance

Big data governance refers to the management of the mind-boggling volume of an organization's data. Big data in government comes from disparate sources such as traffic, cameras, sensors, satellites, emails, mobile phones, RFID, and social media. A data leader serves the public as a data evangelist or advocate for data-driven government. State and local governments want to make their data visible, accessible, and useful to its constituents. The following government responsibilities can benefit from timely analysis of big data [12]:

- Providing national defense and security
- Ensuring public safety with police, fire, disaster and emergency services
- Educating citizens
- Delivering healthcare, public health and social programs
- Building and maintaining roads, bridges and other infrastructure
- Providing transportation services
- Protecting the environment
- Managing taxation

The government market is beginning to explore the value of big data and is budgeting accordingly. More and more government agencies regard big data as important and view it as a strategic asset that can be used to make better, faster decisions. They realize that data needs to be of high quality, accurate, complete, timely, protected, and trustworthy. Many have established data strategy goals that typically include these principles [13]:

- Make data visible so data consumers can easily locate it.
- Make data accessible so data consumers can easily retrieve information when they need it.
- Make data understandable so consumers can recognize what the reports mean. This is critical to enable enhanced, accurate, and timely decision-making.
- Maintain trustworthiness of data to deliver needed value to all stakeholders for making timely decisions.
- Make data interoperable by properly exchanging data between systems.
- Maintain security so consumers know that data is protected from unauthorized use and manipulation.
- Prepared data in a timely manner.

The US government has recognized the importance of big data by promoting and supporting research on advancing big data science and engineering. In March 2012, the US federal government launched the Big Data Research and Development Initiative with \$200 million in new spending to improve the tools/techniques needed to track, access, organize, store, model, and analyze data [14]. The US Government Accountability Office (GAO) provides reference tools to the public to help perform audits in the methodical and thorough fashion.

Applications

We consider some examples of program areas in government agencies that are actively applying big data strategies [15-18]. Figure 4 shows some applications of big data in government [17].





Figure 4: Big data in government [17]

- **Smart Cities:** Cities often launch initiatives to focus on a specific aspect of life. The smart city incorporates data into city planning and hopes to achieve spans across various departments including water, energy, communication, housing, and mobility, all with the intention of improving people's quality of life. There are five main pillars in a close city: intelligent mobility, intelligent economy, intelligent governance, intelligent environment, and intelligent people. Typical examples of smart cities exist in the United States, Korea, China, Japan, India, Canada, Singapore, the United Kingdom, and the European Union. For example, the city government of San Francisco has employed a form of the big data approach to reduce frequent traffic collisions in the city. As cities collect data, they can gather more impactful insights into local problems facing the citizens.
- **Policy-making:** Big data holds great potential for improving decision-making processes and policy-making. Big data is becoming a key part of transactions and decision-making processes. Big data serves as a cue to improve the implementation and effectiveness of government policies. It can provide profound insights into human mobility and economic behavior, to better inform policy makers. Governments have extensively built their capacities to make choices based on big data inputs. They have also turned to big data to inform policy decisions. The global policymakers are showing a keen interest in governance through big data technology. They employ satellite imagery, cell phones, vehicle sensors, and social media for policy-making and urban planning. There has been a growing awareness that data can inform a better policy making process.
- **Public Safety:** Police wear body cameras and video that records continuously. Meanwhile, police departments receive more 9-1-1 calls than ever as cities grow denser. Law enforcement and justice departments now can acquire warrants for information from giants like Google, Apple, or Facebook including location data from GPS, Wi-Fi connections, and cell towers.
- **Fraud Detection:** Fraud is a significant problem for government. The idea that government should focus more on preventing problems instead of just reacting to them is not new. Experts warn that fraud against government may get worse in coming years. These forces will drive more fraudsters and organized crime syndicates to turn to government targets. Predictive analytics tools are being used to detect tax fraud and emergency services. They allow more governments to enhance transparency, increase citizen engagement in public affairs, prevent fraud and crime, improve national security, and support the well-being of people through better education and healthcare. They also allow the government to get ahead of problems before they waste money and harm lives.
- **Trust in Government:** In recent years, public trust in government has reached historic lows. Public trust in government can be improved greatly with greater transparency of operations using big data and



analytic techniques. Research show that as the public learns more about government operations and has more visibility, resident trust and satisfaction levels increase. Other data-driven or data-enabled methods to improve public trust in government include keeping constituents safe and secure, increasing transparency via open data, breaking down the silos of government, providing more convenient ways to engage the public, and include public voice in their government's decision making.

- *Poverty Eradication:* There is so much poverty all over the world. Poverty excludes many people from being part of the new world of data and information. Many governments have tried for years to eradicate poverty. Public and private leaders will need to focus on these poor people if they are to achieve zero extreme poverty. It is through the use of big data that governments are equipped to make better decisions for decreasing poverty all over the world.
- *Open Government:* Data is open if people can use and distribute it freely. Open data requires government and other sources to make some data freely available for users to use as they deem fit without restrictions from copyright or other control mechanisms. It refers to the idea that data should be available freely to everyone to use without restrictions. In return, governments obtain benefits such as long-term transparency, citizen participation, and innovation. Open data should lead to an open government and play an important role in good e-government [19]. Figure 5 illustrates how open data relates to other types of data [20]. The concept of open government is a new phenomenon which has attracted a lot of attention in recent years. It includes free and open access to government information. Making government information available publicly has been termed a civil right by the United Nations. The Open Government movement started in January 21, 2009 when President Barack Obama issued a Memorandum on Transparency and Open Government. On his first day in office, Obama committed his administration to an unprecedented level of openness in government. Open government has increased in importance and become an essential global issue. Open government allows open data initiatives to increase trust, transparency, and accountability between citizens and government [21].

Other applications of big data in government include public health, transportation, healthcare, education, agriculture, taxation, fraud detection, fighting crime, environmental protection, energy exploration, national security, cybersecurity, criminal justice, weather forecasting, and open data initiatives.

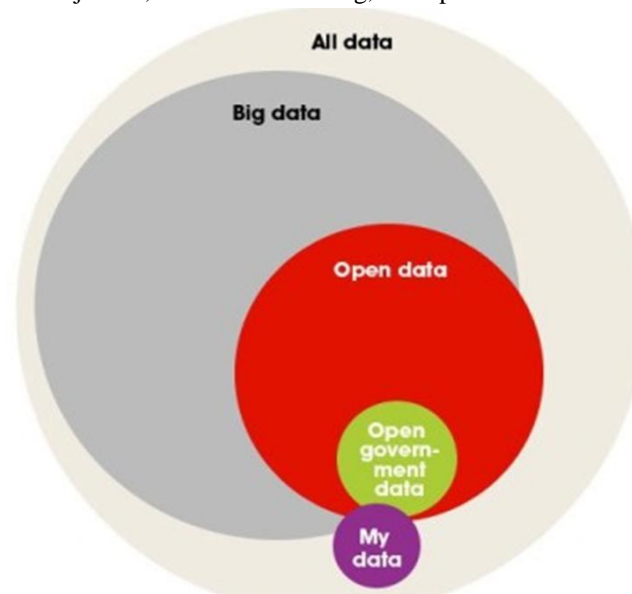


Figure 5: How open data relates to other types of data [20]

Benefits

Big data from citizens can help go break down the silos of government leaders, society, and policy-making to improve community's lives. They have revolutionized the way governments engage with the citizens. The reasons government embraces big data include the following [22]:



- Replacing or supporting human decision-making with automated algorithms
- Reducing inefficiencies within an agency
- Creating transparency
- Improving performance by enabling experimentation to discover needs and expose variability
- Improving ROI for IT investments
- Improved decision-making and operational intelligence
- Providing predictive capabilities to improve mission outcomes
- Reducing security threats and crime
- Eliminating waste, fraud, and abuse
- Innovating new business models and stakeholder services

Big data empowers governments to serve citizens more effectively. Government-wide digital transformation enables open, transparent government and to facilitate government service delivery. Big data technologies are at the heart of the intelligent economy. Government organizations are mining data to prevent bad actors from committing acts of terror and prevent waste, fraud, and abuse. Government leaders want access to big data to improve program accuracy, better serve constituents, protect the warfighter, prevent cyber attacks, protect critical information, and overcome challenges in identity management [23]. Big data holds a great deal of promise for driving efficiencies that could improve the lives of people around the world. It allows governments to better oversee transportation to ensure safer roadways.

Many governments have come to the conclusion that the potential benefit outweighs the potential negative impact. Politicians around the world tend to play down people's fears by pointing to increased transparency and accountability [22]. Once the data is open and available, opportunities to apply them in different areas are endless. Big data plays a major role in mass citizen surveillance. Figure 6 shows some of the benefits of big data for governments [17].

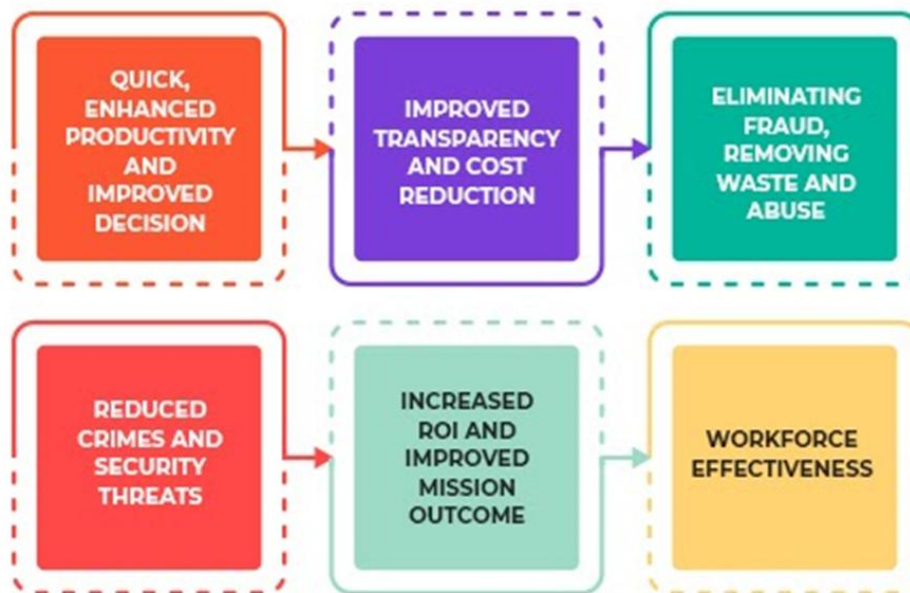


Figure 6: Benefits of big data for governments [17]

Challenges

Big data presents both benefits and challenges for government. The collection, processing, and analysis of big data is a built-in culture challenge for many years to come. Managing big data in government is a major challenge as public servants are held back by budgetary issues, security issues, or a lack of resources. Government faces many challenges with implementing big data strategies. Managing big data as government's traditional silo approach discourages information sharing and hinder movement from process driven to data



driven. Budgetary constraints are compelling government agencies to right size their big data information storage. Technical challenges include shortage of relevant talent, lack of relevant software tools, integration of multiple data sources, and data storage and access.

Privacy issues are often the biggest challenge presented by the rise of big data.

Fundamental elements of human rights (privacy, ethics, access, and respect for data sovereignty) must be safeguarded. Proper data protection measures must be put in place to prevent data misuse or mishandling. Major gaps are already opening up between the data haves and have-nots. Many people are excluded from the new era of big data by language, poverty, lack of education, lack of technology infrastructure, and inequality, especially in developing countries. Women and girls, for example, suffer social constraints and are marginalized in both private and public decision making. Many governments do not have access to adequate data on their entire populations, particularly for the poorest and most marginalized [2].

Conclusion

Governments at all levels collect an increasing amount of data daily. Big data is one of the emerging, critical issues facing governments in the digital era. It is transforming governments. Governments must be data-driven to manage these new, immense inputs, and to satisfy constituents who evolved along with this global data revolution. In view of the massive data influx from millions of devices, agency systems, cameras, and more, the public sector has become data-driven. Governments must use the power of big data to be more citizen-focused, design cost-effective services, and prepare for the challenges of a digitally transformed society. More information about big data in industry can be found in the books in [24,25] and the following related journals:

- *Journal of Big Data*
- *Big Data Research*
- *Big Data & Society*
- *Government Information Quarterly*

References

- [1]. L. Andrews, "What are examples of big data in manufacturing?" October 2019, <https://www.sensrtrix.com/what-are-examples-of-big-data-in-manufacturing/>
- [2]. [2] J. W. Lee, "Big data strategies for government, society and policy-making," *The Journal of Asian Finance, Economics and Business*, vol. 7, no. 7, 2020, pp. 474-487.
- [3]. G. H. Kim, S. Trimi, and J. H. Chung, "Big-Data Applications in the government sector," *Communications of the ACM*, vol. 57, no. 3, March 2014, pp. 78-85.
- [4]. B. Klievink et al., "Big data in the public sector: Uncertainties and readiness," *Information Systems Frontiers*, vol. 19, 2017, pp. 267-283.
- [5]. "The 42 V's of big data and data science," <https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html>
- [6]. M. N. O. Sadiku, M. Tembely, and S. M. Musa, "Big data: An introduction for engineers," *Journal of Scientific and Engineering Research*, vol. 3, no. 2, 2016, pp. 106-108.
- [7]. P. K. D. Pramanik, S. Pal, and M. Mukhopadhyay, "Healthcare big data: A comprehensive overview," in N. Bouchemal (ed.), *Intelligent Systems for Healthcare Management and Delivery*. IGI Global, chapter 4, 2019, pp. 72-100.
- [8]. J. Moorthy et al., "Big data: Prospects and challenges," *The Journal for Decision Makers*, vol. 40, no. 1, 2015, pp. 74-96. <https://www.grandviewresearch.com/industry-analysis/industrial-wireless-sensor-networks-iwsn-market>
- [9]. M. N. O. Sadiku, J. Foreman, and S. M. Musa, "Big data analytics: A primer," *International Journal of Technologies and Management Research*, vol. 5, no. 9, September 2018, pp. 44-49.
- [10]. C. M. M. Kotteti, M. N. O. Sadiku, and S. M. Musa, "Big data analytics," *Invention Journal of Research Technology in Engineering & Management*, vol. 2, no. 10, Oct. 2018, pp. 2455-3689.
- [11]. "How big data is transforming industries in big ways," June 2020,



- <https://www.tiempodev.com/blog/how-big-data-is-transforming-industries-in-big-ways/>
- [12]. IBM, "Addressing government challenges with big data analytics," <https://docplayer.net/1130110-Addressing-government-challenges-with-big-data-analytics.html>
- [13]. C. Gregory, "A modern big data strategy for vaulting federal agencies toward digital modernization," April 2021, <https://looker.com/blog/big-data-strategy-government-sector>
- [14]. "The impact of big data on government," <https://www.ironmountain.com/resources/whitepapers/t/the-impact-of-big-data-on-government>
- [15]. B. Marr, "6 incredible ways big data is used by the US government," <https://www.smartdatacollective.com/incredible-ways-big-data-used-us-government/>
- [16]. J. Wiseman, "The case for government investment in analytics: Why every government executive should care about data," September 2019, <https://datasmart.ash.harvard.edu/news/article/case-government-investment-analytics>
- [17]. M. Sahu, "Big data in the public sector: Applications and benefits," August 2021, <https://analyticssteps.com/blogs/big-data-public-sector-applications-and-benefits>
- [18]. R. Joseph, "Big data analytics in government: How the public sector leverages data insights," June 2019, <https://www.intellectyx.com/blog/big-data-analytics-in-government-how-the-public-sector-leverages-data-insights/>
- [19]. M. N. O. Sadiku, M. Tembely, and S.M. Musa, "Open data: Opportunities and challenges," *Journal of Multidisciplinary Engineering Science and Technology*, vol. 3, no. 11, Nov. 2016, pp. 6006-6008.
- [20]. "Open data: Regulation and cooperation result in business efficiency," <https://oliviazhuyan.wordpress.com/2015/03/30/open-data-regulation-and-cooperation-result-in-business-efficiency/>
- [21]. M. N. O. Sadiku, and S. M. Musa and O. D. Momoh, "Open government," *International Journal of Engineering Research and Advanced Technology*, vol. 2, no.12, Dec. 2016, pp. 17-20.
- [22]. "Demystifying big data," https://bigdatawg.nist.gov/_uploadfiles/M0068_v1_3903747095.pdf
- [23]. "The US government: Using big data to run a country," in B. Marr (ed.), *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*. John Wiley & Sons, chapter 36, 2016.
- [24]. M. Lowman (ed.), *A Practical Guide to Analytics for Governments: Using Big Data for Good*. John Wiley & Sons, 2017.
- [25]. J. M. Cavanillas, E. Curry, and W. Wahlster (eds.), *New Horizons for a Data-driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe*. Springer, 2016.

About Authors

Matthew N. O. Sadiku is a professor emeritus in the Department of Electrical and Computer Engineering at Prairie View A&M University, Prairie View, Texas. He is the author of several books and papers. His areas of research interest include computational electromagnetics and computer networks. He is a fellow of IEEE.

Uwakwe C. Chukwu is an Associate professor in the Department of Industrial & Electrical Engineering Technology of South Carolina State University. He has published several books and papers. His research interests are power systems, smart grid, V2G, energy scavenging, renewable energies, and microgrids.

Abayomi Ajayi-Majebi is a professor in the Department of Manufacturing Engineering at Central State University in Wilberforce, Ohio. In 2015 he was honored by the White House as a Champion of Change for his significant contributions to the engineering education of minority students. He is a senior member of both the Society of Manufacturing Engineers and the American Society for Quality.

Sarhan M. Musa is a professor in the Department Electrical and Computer Engineering at Prairie View A&M University, Texas. He has been the director of Prairie View Networking Academy, Texas, since 2004. He is an LTD Sprint and Boeing Welliver Fellow. His areas of research interest include computational electromagnetics and computer networks.

