Chapter 13
Big Data for Digital Transformation of Public Services

Deepak Saxena
https://orcid.org/0000-0002-9331-3799
Birla Institute of Technology and Science, Pilani, India

ABSTRACT

Big data is presently considered integral to the management and strategies for digital enterprise transformation. Beyond being 'a lot of data', big data can be characterized in terms of seven Vs: volume, velocity, variety, variability, veracity, visualization, and value. Already being applied in private businesses, big data has immense potential for the digital transformation of public services in advancing the e-governance agenda. This chapter explores the nature of big data in public service and discusses its application in areas such as tax administration, transportation, energy, public health, and disaster management. Challenges and concerns are noted in terms of data quality, infrastructure cost, availability of suitable human resources, privacy, and security. Possible solutions such as shared services, cloud computing, open source software, open data framework, and regulatory compliance are noted. The chapter ends by noting future research directions to realize the full potential of Big data application in digital transformation of public services.

INTRODUCTION

Digital enterprise transformation may be defined as end-to-end value creation for the enterprise via management and application of digital technology aligned with business strategy. Big data is presently considered integral to the management and strategies for digital enterprise transformation. However, it is not entirely a new term. In the context of data and storage management, the term Big data was already in use in computing circles (e.g. Bourgoin & Smith, 1995; Mashey, 1997) even before the dawn of the new millennium. However, the term became well-known at the start of the decade when popular business research outlets such as McKinsey Global Institute (Brown, Chui, & Manyika, 2011) and Har-

DOI: 10.4018/978-1-7998-8583-2.ch013

vard Business Review (McAfee & Brynjolfsson, 2012) started publishing about it. By 2015, Gartner dropped the term from the hype cycle, noting that Big data has become so prevalent that it is now part of multiple emerging technologies (Heudecker, 2015) such as business intelligence, data science, social media analytics, enterprise information management or machine learning. The central role of Big data in digital enterprise transformation reflects in wide availability of technological solutions such as Apache Hadoop, Spark, Microsoft HDInsight, NoSQL, or Hive. In fact, Apache Hadoop has emerged as a best in-class Big data ecosystem, supporting data processing capabilities and system connectivity across heterogeneous databases and systems. However, since the focus of this chapter is on management and strategies for the transformation of digital enterprises, specific technologies are not discussed in this chapter. The interested reader is directed towards some recent works (Luengo, García-Gil, Ramírez-Gallego, García, & Herrera, 2020; Rao, Mitra, Bhatt, & Goswami, 2019) for a discussion on Big data ecosystem, technology and tools.

The key role of Big data in digital enterprise transformation and value creation may be inferred from the fact that KPMG / Harvey Nash CIO survey (2019) identifies Big data analytics as the most important (and the scarcest) skillset required in organizations. Big Data capabilities are found to be an important predictor of value creation and firm performance (Wamba, Gunasekaran, Akter, Ren, Dubey, & Childe, 2017; Mikalef, Boura, Lekakos, & Krogstie, 2019), especially in the private sector. However, the scope of Big data is not just limited to the private sector firms. Big data is frequently noted (Joseph & Johnson, 2013; Gaardboe, Svarre, & Kanstrup, 2015; Patel, Roy, Bhattacharyya, & Kim, 2017; Shukla & Mathur, 2020) as a key element in the management and strategies of digital public service transformation. This chapter explores the potential of Big data in advancing the e-Governance agenda via digital transformation of public services. Hence, the research question for this study is: *How does Big data contribute towards the digital transformation of public services?*. The research approach is exploratory in nature and thus relies on secondary data and studies. The remainder of the chapter is as follows. The next section provides a general background to Big data and outlines seven Vs that underpin the Big data revolution. This is followed by a discussion on how seven Vs are applicable to Big data in public service. Key application areas from the public service that may be transformed by Big data applications are discussed in detail. Thereafter. challenges and concerns associated with the development of Big data capabilities in the government are also noted. This is followed by offering solutions and recommendations to the problems raised earlier. Future research directions are noted before concluding the chapter.

Background

McAfee and Brynjolfsson (2012) note that Big data is more than 'a lot of data' owing to three key factors – Volume, Velocity and Variety – a categorization originally introduced by Laney (2001) to characterize e-Commerce data management. *Volume* refers to the generation of large chunk of data. It is estimated that the volume of digital data would be around 50 Zettabytes (1 Zettabyte equals 10^{21} bytes) in 2020, reaching a staggering 175 Zettabytes by 2025. *Velocity* refers to the speed with which data is generated. Correct and rapid processing of data in the real time has become extremely crucial for companies. For instance, Chinese retailer Alibaba processed around 544,000 orders per second on Singles day 2019. Similarly, Amazon US processed 20.2 million transactions during its Prime Day, a three-day event, in 2019. *Variety* refers to various forms of data in disparate formats that are combined to generate a holistic picture. For instance, Google collects and combines data from various sources (Android operating system, Chrome browser, Gmail, Maps, Search history, Voice, and YouTube activity to name a few) to personalize its offerings to the users.

It was soon realized that three Vs do not adequately capture the nature and potential of Big data. Noting its increasing importance, some additional Vs are suggested (Van Rijmenam, 2013; Khan, Uddin & Gupta, 2014) in relation to Big data – Variability, Veracity, Visualization, and Value. *Variability* refers to the variability in the meaning when processing natural language data. For example, the term 'great service' would differ in meaning depending on being preceded by 'got the reply within two hours' or 'still waiting for reply since last 15 days'. *Veracity* refers to the authenticity and truthfulness of the data (Rubin & Lukoianova, 2013). This is more relevant in case of unstructured data, for instance making sure that service reviews are coming from authentic users and not from automated bots. *Visualization* primarily relates to the appropriation of data as opposed to its inherent nature. Since Big data is unstructured and comes from a variety of sources, visualization of the trends helps in making sense of the vast tapestry of data. Therefore, all Big data applications include a dashboard (Pauwels, 2014) to assist top management in the decision-making. These dashboards include features like trend analysis or scenario planning, thereby extending the strategy toolbox (Woerner & Wixom, 2015). Finally, *Value* refers to the outcomes, for instance efficiency or reputational gains, made possible by Big data analytics. Beyond operational gains, however, Big data capabilities may act as an enabler of digital enterprise transformation (Dremel, Wulf, Herterich, Waizmann, & Brenner, 2017).

However, it may be noted that not all Vs are of equal importance at all times. For instance, Davenport and Dyche (2013) report that, compared to the volume alone, managing a variety of data sources in an integrated manner and developing analytical capabilities is more valuable for the companies. In contrast, managing large volume of transactions at a high velocity becomes crucial during events like Singles' Day or Black Friday sale. In this regard, Wamba et al. (2017) propose a Big Data Analytics Capability model consisting of three core areas – infrastructure flexibility, management capabilities, and personnel expertise capabilities. Big data infrastructure flexibility refers to modularity, connectivity and compatibility of various systems. Big data management capability requires planning, investment, coordination, and control of Big data resources. Finally, personnel expertise capabilities in the Big data context include technology, business and relational knowledge, along with technology management capability.

While Big data capabilities matter for all types of organizations from the perspective of digital enterprise transformation, it may be argued that not all Vs are important for all types of organizations. The factors that are significant for private companies may not exactly be the same as those relevant for digital public services. Similarly, short-term profit motive of the private sector is usually less relevant for public services that focus on creating public value (Saxena & McDonagh, 2019) in the long run. Next section highlights these differences and discusses how Big data may propel digital transformation of public service across diverse domains.

BIG DATA IN PUBLIC SERVICE

This section begins by outlining the nature of big data in public service by highlighting how seven Vs noted earlier relate to the government and public service. This is followed by discussion on public service applications from diverse areas such as tax administration, transportation, energy, public health, and disaster management. Thereafter challenges and concerns associated with Big data applications in the public service are discussed.

Nature of Big Data in Public Service

Mergel, Rethemeyer, and Isett (2016) define Big data for the public affairs as *high volume data that frequently combines highly structured administrative data actively collected by public sector organizations with continuously and automatically collected structured and unstructured real-time data that are often passively created by public and private entities through their Internet interactions* (p. 931). This definition touches upon many Vs discussed earlier. With help of specific examples, this section notes how each V relates to public services.

Volume

Governments usually deal with entire population for many departments such as revenue or social security. Public service data tends to be exhaustive (Kitchin, 2014) as compared to their private sector counterparts who have the liberty of focusing only on their intended customer base. With the exception of global corporate giants such as Amazon, Google or Facebook, the data processed by populous countries such as China, India or US easily surpasses the volume of data processed by companies.

Velocity

Depending on the time of the year, public systems may need to deal with very high velocity of data. There are often reports of tax systems being overloaded closer to the deadline of tax assessment and filing. As another example, the Covid-19 pandemic provides multiple instances of such high velocity data – from testing results to contact tracing, and registration for unemployment payments – that need to be quickly processed by the public services.

Variety

This is perhaps the most crucial element for the Big data for public services. Government systems collect data from a wide variety of sources at various levels (city level to federal government level), across departments (from health to education to agriculture) and across diverse systems. Enormous amount of data is stored in silos (Kim, Trimi, & Chung, 2014; Rajagopalan & Vellaipandiyan, 2013) across numerous systems within the government with data integration being a main challenge. Goldstein (2019) suggests that, owing to the variety of data being stored and processed by the governments, data lakes are a useful strategy for Big data utilization in public services. As opposed to data warehousing that relies on structured data, data lakes allow for centralized storage of any type of structured and unstructured data in raw form. This is a crucial requirement for public services since different government departments often find it difficult to agree upon their data needs.

Variability

Traditionally, governments used to collect opinions and suggestions from the interest groups by meeting with their representatives or by inviting suggestions. Social media has given them a great tool for not only directly interacting with the public, but also acting as a barometer of public reaction (Mergel et al., 2016) to a policy announcement. While the governments are yet to formally adopt this approach, scholars have already conducted sentiment analysis on key government decisions (Singh, Sawhney, & Kahlon, 2018) and citizen involvement with local government (Zavattaro, French, & Mohanty, 2015). With the growing importance of public opinion on the social media, it is an imperative that governments include social media sentiment analysis in their consideration.

Veracity

Quite often, the fraud with the government services (tax evasion, for example) happen due to the siloed nature of government data. To curb this tendency, governments have moved towards joined-up governance (Klievink & Janssen, 2009; Carey & Crammond, 2015) to ensure the veracity and consistency of data across departments. This is often done by linking the data via national identity mechanisms such as Social Security Number is the US, National Insurance Number in the UK, and Aadhaar in India. Once again, this takes us back to the variety element of the Big data, making it the most important factor in utilizing Big data in government.

Visualization

Since the data in Big data is often in unstructured form, possibly the most profound application of Big data is visualization (Davenport & Dyche, 2013) to aid in decision-making. Leaders in the government are often short on time and need condensed information to make quick decisions. For instance, Geographical Information Systems (GIS) and mapping data may be used to visualize the distribution of resources or requirements across the country. Such data visualization would help the leaders in quickly understanding the situation and taking evidence-based policy decisions (Höchtl, Parycek, & Schöllhammer, 2016).

Value

Value is the desired outcome of the Big data application and processing (Van Rijmenam, 2013; Khan et al., 2014). While the notion of value is often captured in terms of greater efficiency and increased revenue for private businesses, the notion of public value is more complex. Economic logic, such as cost savings or increased economic activity, is still applicable (Munné, 2016) to public services. However, public agencies also need to cater for equity and effectiveness (Boyle, 2006) apart from economic efficiency, for instance in deciding the location of schools to ensure accessible education. In addition, public agencies may also need collated Big data from different sources to fulfill the auditing requirements.

Big Data Applications in Public Service

To illustrate to potential of Big data in public service, this section discusses various Big data applications with illustrative examples from around the globe. These applications include, but not limited to, areas such as tax administration, transportation, energy, public health, and disaster management.

Tax Administration

Taxes, especially indirect taxes, form the core of government revenue. However, there is always some gap between expected tax liabilities and actual taxes paid. Internal Revenue Service (2019) estimates that gross annual tax gap in US during 2011-13 was more than $440 billion, around 16.4 percent of the taxes due. In contrast, that the tax gap in UK (HMRC, 2018) for 2017-18 was around £35 billion or 5.6 percent of taxes due. A major reason for the relative efficiency of the UK tax authorities is the use of Big data. Big data helps them in two major ways. First, they collate data from data aggregators and merchant service providers, including those from overseas. This way they can triangulate the sales and revenue data furnished by the companies and infer any relationships between the entities. Second, UK authorities identify high risk individuals and companies based on tax data analytics. Similar exercise is being conducted by Australian taxation office (Walker-Munro, 2020) that is developing a Big data network analytics solution called ANGIE to identity multi-layer complex relationship between its clients, in order to assist its tax avoidance taskforce. Atanasijević, Jakovetić, Krejić, Krklec-Jerinkić and Marković (2019) also report a similar project in the Republic of Serbia that employs Big data analytics (based on the weighted norm distance between the average income distribution of the industry sectors and the income distribution of the business) to identify low-risk and high-risk entities. High risk entities are then investigated further to ensure that there is no tax evasion or fraud (Campbell, 2014).

Transportation

Big data can be used for the management of intelligent transport for efficient public services. This is applicable both to the freight transport and public transport. The area of freight transport is no longer applicable only to the big companies like Amazon or DHL, it has become equally important for the governments across the world, for instance in planning and tracking medical supplies amid the coronavirus pandemic. Reflecting the variety element of Big data, data from the GPS trackers, satellite, sensors, and mobile devices (Zhu, Yu, Wang, Ying, & Tang, 2018) can be used to plan and optimize the route as well as for real-time tracking of the shipments. In the area of public transport, public service utilities may apply big data for operational improvements as well as for strategic planning. For instance, *HTM Personenvervoer*, a tram operator from Hague in Netherlands, uses smartcard data to predict ridership behavior and to plan the services across the tram network (Van Oort & Cats, 2015). At the operational level, *Stockholmstag*, the train operator in Stockholm in Sweden, uses Big data to ensure the reliability of its service (Kepes, 2015). The system forecasts and visualizes the train operations for next two hours. If it detects any potential delays (e.g. a train is delayed between stations X and Y, thereby introducing delay at station Z), it alerts the operator. To ensure the reliability of the service, a new train is then introduced from station Z so that there are no delays for the passengers waiting at the station. Heaphy (2019) reports similar application in Irish capital Dublin, where Dublin Bus services use visualization tools based on a variety of data – bus timetables, GPS-based vehicle location, CCTV footage, smartcard scans – to drill down on specific issues and offer traffic-calming measures to reduce congestion. Apart from traditional transport means, Big data analytics may play a crucial role in providing electric mobility as a service (Jnr, Petersen, Ahlers, & Krogstie, 2020), by allowing for the sharing of high velocity data on electric vehicles and the utilization of charging infrastructure in smart cities. Apart from the improvement in transport service delivery, such initiatives would also contribute towards the sustainability of the planet. With the advent of self-driving cars in the future, the importance of Big data in transport would become more important than ever.

Energy

Big data could be an enabler of intelligent energy management. Zhou, Fu, and Yang (2016) discuss the application of Big data across energy lifecycle – power generation, power transmission, power distribution, and demand side management. Based on the data collected from power grid equipment, GIS data, weather data etc., power generation companies run simulations and forecast power requirements. At times, they need more complex simulation for load balancing if they generate power from diverse sources (e.g. coal, nuclear, hydro or wind energy). During power transmission and distribution, data collected from sensors may be used for fault detection, electric device health monitoring, predictive maintenance, and power quality monitoring (Zhang, Huang, & Bompard, 2018). In situations where granular data is not available or customer privacy is prioritized, big data may still be used to generate clusters (Ushakova & Mikhaylov, 2020) for efficient energy management. At the same time, Big data also enables effective demand-side management primarily via smart meters. Smart meters provide feedback to the end users on their energy consumption so that they can adjust their usage. Italy provides the best example of the smart metering system installed by the state-controlled energy provider *Enel* (now rebranded as *e-distribuzione*). These meters provide data on daily, weekly, and monthly usage and help the consumers in reducing cost and the power company in increasing efficiency. Italy is well-ahead of the curve and has commenced the installation of second-generation smart meters (Stagnaro & Leoni, 2019) that would provide more real-time data to the consumers and make the process more interactive. Among Scandinavian nations, Sweden and Finland are also gearing up for the rollout of second-generation smart meters in 2021.

Public Health

Big data in healthcare (Pastorino et al., 2019) can include data from wearable devices, electronic health records, patient summaries, pharmaceutical data, telemedicine, health apps, social media, and socio-economic indicators, among others. Big data in healthcare can be used at multiple levels. At the clinical level, patient summary and electronic health records can provide an aid to the doctors for predicting the onset of disease and in designing the most appropriate treatment for a patient. Moreover, wearable devices contribute towards the patient-centered health monitoring to facilitate self-management of the chronic health condition (Chiauzzi, Rodarte & DasMahapatra, 2015), also supporting social distancing during the coronavirus pandemic. Big data also helps in patient-centered design in healthcare. For instance, MARIO project from Ireland has developed caring service robots for older people with dementia. With customized applications and user interface, the service robot helps the patients in fighting loneliness and isolation. At the institutional level, Big data can help is effective planning and cost optimization by using predictive analysis on disease occurrence and patient treatment. The eHealth project from Estonia has digitalized the patient information and prescriptions, thereby developing high-quality health services based on data and process analytics. However, the biggest application of Big data is apparent in the monitoring and surveillance of infectious disease (Hay, George, Moyes & Brownstein, 2013) such as the coronavirus pandemic that started in late 2019. Disease data can be clubbed with other environmental variables (e.g. population density, people movement, average age) to prepare a real-time map of the spread of the disease and to prepare containment strategies. Although having some privacy concerns, China (Health Code app) and Ireland (Covid Tracker app) have successfully used location-infection data and have applied Big data analytics to alert its population and public health officials on the spread of the disease. Moreover, Big data may also help in preventing further outbreaks. For instance, Chen et al (2020) report a case in Taiwan where a Covid-19 patient was identified based on their travel history, even before they could enter the general hospital.

Disaster Management

Apart from ongoing coronavirus crisis, there is always a possibility of natural and man-made disasters such as earthquake, flood or fire. A key element of effective disaster management is the availability of accurate and timely information (Akter & Wamba, 2019). Once a disaster occurs, emergency teams are often saddled with a barrage of information from multiple sources such as – telephone calls, news footage, social media posts, and satellite images. In this situation, Big data analytics may facilitate timely and targeted response. In the past, satellite images and GIS mapping have been used post-facto to delineate affected areas. However, with the ubiquitous availability of geolocation data and social media posts, Big data analytics can help in providing a real-time picture of the ongoing disaster, for instance in locating areas that require immediate assistance and identifying roads that are still approachable. Moreover, open source platforms like Ushahidi or OpenStreetMap allow volunteers across the world to pool data and capabilities to map the disaster in real time. Such form of crowdsourcing has been used, for instance, during earthquakes in Haiti (Norheim-Hagtun & Meier, 2010) and Nepal (Poiani, Rocha, Degrossi & Albuquerque, 2016), and flood relief in India (Anbalagan & Valliyammai, 2016). Once the disaster is over, machine learning can be applied on the Big data to accurately predict the next hotspots of fire or flood. For instance, analyzing data from three different sources, Dutta, Das and Aryal (2016) provide highly accurate (around 91 percent accuracy) estimate on bush-fire incidence in Australia. Such estimates can be used in planning early evacuation and mitigation. Thus, Big data analysis can help both in real-time response as well as in predictive planning.

Challenges and Concerns

While there is immense potential of Big data for the for digital transformation of public service, its management and strategic use is not without challenges. Across the domains outlined above, four key challenges could be identified for Big data application in the public service context.

Data Quality

A lot of historical data in public agencies is either in non-digital format (e.g. printouts, hardbound reports) or in cumbersome digital formats (e.g. scans). Even when digital data is present in information systems, it often exists in silos (Kim et al., 2014; Rajagopalan & Vellaipandiyan, 2013) across different units/departments. Consequently, there is little consistency and structure across disparate systems in the government (Joseph & Johnson, 2013). Data ownership may also become an issue (Pradhan & Shakya, 2018) if a system links dataset from different departments. Moreover, there is often resistance from public services in changing their legacy systems and data formats, further complicating Big data integration.

Infrastructure Costs

Due to technological limitations, quite often the most efficient way for the public service is to build state of the art infrastructure from scratch to support Big data applications. Hence, there is often a huge upfront infrastructure cost (Joseph & Johnson, 2013; Pradhan & Shakya, 2018) in terms of systems, servers, network, and data warehousing. Due to political urgency of other functions, government may not be interested in investing much in information technology. For instance, coronavirus crisis has resulted in budget deficits, with an increasing focus on investment in health and social services. However, as noted earlier, in the long term, investment in Big data infrastructure would be useful for public health in terms of disease tracing and epidemiological management.

Human Resource

Big data is not just about technology – one also needs skilled data scientists to make full use of the data. Due to their limited exposure of public service systems, outside consultants are often not able to appreciate the complexity and subtlety of government data. At the same time, there is often a lack of technological expertise in public service agencies. Due to inflexible nature of the hiring and salary scale, public agencies often find it difficult to attract suitable human resources (Pradhan & Shakya, 2018; Sarker, Hossin, Frimpong, & Xiaohua, 2018) with relevant information technology skills. The problem is exacerbated for Big data capabilities since finding human resources with relevant skillset is reported (KPMG, 2019) as a key issue even for the private sector. Taken together, limited domain knowledge among consultants and a lack of technological capabilities in government departments, create a barrier in the exploitation of Big data in public service.

Privacy and Security

Key benefits of Big data emerge from the standardization and integration of a variety of data sources. For public service agencies, this often means collecting and collating data on its citizens. This results in some privacy concerns in terms of potential state surveillance and right to privacy. For instance, Walker-Munro (2020) discusses concerns associated with privacy, inaccuracy, opacity and bias associated with Big data driven tax administration in Australia. Similar concerns are often raised on the Aadhaar project from India in terms of its failure to protect the privacy of its citizens (Dixon, 2017), especially when analyzed from the point of view of privacy frameworks in the west. Security of Big data is also an issue (Kitchin, 2016) since there is always a risk of hacking of government data. In fact, a rationale often advanced by government departments for not integrating their datasets with other departments is that it may undermine the security of their own database.

SOLUTIONS AND RECOMMENDATIONS

For any system, there would always be a tradeoff between the gains and the risks of using the system. Despite the challenges and concerns noted earlier, the benefits of Big data utilization in public service are manifold and potentially outweigh the risks. This section outlines some solutions and recommendations that may be adopted to ease the concerns and to mitigate the challenges of Big data management in public service transformation.

Shared Services

Considering the infrastructure cost, public agencies may invest in developing shared services (Richter & Brühl, 2020) that provide common service (e.g. human resources, payroll, citizen interface) across public agencies. For instance, Estonian public authorities are using two shared services – one for county councils and the other for central government – that have resulted in overall cost savings (Tammel, 2017) for the government. The shared services model may be applied to Big data management in government. This also aligns with the vision of Big data as an integrated data application. This may take the form of an autonomous public data agency that would benefit from resource pooling and would be able to attract suitable human resources with required competencies.

Cloud Computing

Recently, software firms have moved from providing from software-as-a-product to software-as-a-service. Moreover, the infrastructure is also being provided as a service over the internet. Taken together, this trend is marketed as cloud computing model for providing information systems services, including Big data (Lnenicka & Komarkova, 2019; Yang, Huang, Li, Liu & Hu, 2017). Adoption of cloud computing has the benefits of higher efficiency (since the processing is optimized at the level of the vendor) and lower cost (since many clients share the services provided by the cloud vendor). Moreover, cloud vendors heavily invest on the security of the system so that they do not loose their market due to possible security breaches. However, since the services are shared across clients, there is very little room for the customization. Hence, public agencies may use this architecture mainly for the services that are standard across departments, such as payroll or human resource management.

Open Source/Data

If a public agency wishes to have the in-house implementation requiring a high degree of customization, it may use open source Big data frameworks (Davenport & Dyche, 2013) such as Apache Hadoop, Cassandra or MongoDB. Moreover, using open source platforms allows for crowdsourcing (Anbalagan & Valliyammai, 2016; Norheim-Hagtun & Meier, 2010; Poiani et al., 2016) in areas such as disaster management or in public health management. In fact, now there is an increasing movement towards open data in government (Bertot, Gorham, Jaeger, Sarin & Choi, 2014) where public agencies share standardized datasets not only with other public agencies but also with the general population. Provision of open data helps towards fulfilling the obligations under the Freedom of Information acts enacted in various countries. This may also completement the cloud computing initiatives (Lnenicka & Komarkova, 2019) discussed earlier.

Regulatory Compliance

Irrespective of the solutions used, a key requirement for public agencies is regulatory compliance (Walker-Munro, 2020) in terms of government audit and towards ensuring the privacy of its citizens. Frameworks like *General Data Protection Regulation* in the European Union or *Australian Privacy Principles* support public agencies in ensuring regulatory compliance. In fact, such regulations may indirectly support the cause of Big data as data formats need to be standardized to fulfil regulatory requirements. This would, for instance, involve structuring data according to its sensitivity and implementing privacy preserving techniques (Gruschka, Mavroeidis, Vishi & Jensen, 2018) as per regulations. Such privacy preserving techniques would not only ensure Big data integration in a legally accepted way, it would also help in accurate data mining (Chamikara, Bertók, Liu, Camtepe & Khalil, 2020).

FUTURE RESEARCH DIRECTIONS

Based on the recommendations, there are two distinct research directions that apply to the management and application of Big data in public service transformation. The first direction relates to research on management strategies and architectural approaches (for instance, use of shared services and cloud computing) to exploit the potential of Big data in government. While a lot of evidence is available from the private sector on providing shared services, research in public sector is required in view of their specific institutional context. Second research direction relates to the variety element of Big data. There is a need for the development of open data standards and sharing protocols for linking the variety of data sources (e.g. formal report, quantitative data, textual data, social media data, sensors, maps, speech, video, crowdsourced data etc.) applicable to public services in a privacy preserving manner. This would support the cause of Big data integration without compromising the privacy of the citizens.

CONCLUSION

Big data is being widely utilized in the private sector to support the management and strategies for digital enterprise transformation. This chapter notes the potential of Big data for digital transformation of public services. Seven Vs of Big data are discussed in the context of public service, noting that volume and variety are perhaps more crucial for public service than the velocity alone. Big data applications from the area of tax administration, transportation, energy, public health, and disaster management, are discussed with relevant examples from across the globe. However, there are also challenges associated with data quality, infrastructure cost, availability of suitable human resources, privacy and security. These challenges can be mitigated with initiatives such as shared services, cloud computing, open source software, open data framework, and regulatory compliance. For these initiatives to be successful, future research should focus upon Big data management strategies and architectural approaches, along with the development of open data standards and sharing protocols in privacy preserving manner. This would allow the governments to realize the full potential of Big data application in digital transformation of public services.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Akter, S., & Wamba, S. F. (2019). Big data and disaster management: A systematic review and agenda for future research. *Annals of Operations Research*, *283*(1-2), 939–959. doi:10.100710479-017-2584-2

Anbalagan, B., & Valliyammai, C. (2016, December). #ChennaiFloods: Leveraging Human and Machine Learning for Crisis Mapping during Disasters Using Social Media. In *2016 IEEE 23rd International Conference on High Performance Computing Workshops (HiPCW)* (pp. 50-59). IEEE.

Atanasijević, J., Jakovetić, D., Krejić, N., Krklec-Jerinkić, N., & Marković, D. (2019). Using big data analytics to improve efficiency of tax collection in the tax Administration of the Republic of Serbia. *Ekonomika Preduzeća*, *67*(1-2), 115–130. doi:10.5937/EkoPre1808115A

Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information Polity, 19*(1, 2), 5-16.

Bourgoin, M. O., & Smith, S. J. (1995). *Big data—Better Returns, Leveraging Your Hidden Data Assets to Improve ROI. In Artificial Intelligence in the Capital Markets*. Probus Publishing Company.

Boyle, R. (2006). *Measuring public sector productivity: lessons from international experience*. CPMR Discussion Paper 35. Institute of Public Administration.

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

Campbell, N. (2014). Tax Policy and Administration in an Era of Big data. *Tax Planning International*, *12*(12), 2–6.

Carey, G., & Crammond, B. (2015). What works in joined-up government? An evidence synthesis. *International Journal of Public Administration*, *38*(13-14), 1020–1029. doi:10.1080/01900692.2014.982292

Chamikara, M. A. P., Bertók, P., Liu, D., Camtepe, S., & Khalil, I. (2020). Efficient privacy preservation of big data for accurate data mining. *Information Sciences*, *527*, 420–443. doi:10.1016/j.ins.2019.05.053

Chen, F. M., Feng, M. C., Chen, T. C., Hsieh, M. H., Kuo, S. H., Chang, H. L., Yang, C. J., & Chen, Y. H. (2020). Big data integration and analytics to prevent a potential hospital outbreak of COVID-19 in Taiwan. *Journal of Microbiology, Immunology, and Infection*. Advance online publication. doi:10.1016/j.jmii.2020.04.010 PMID:32327328

Chiauzzi, E., Rodarte, C., & DasMahapatra, P. (2015). Patient-centered activity monitoring in the self-management of chronic health conditions. *BMC Medicine*, *13*(1), 77. doi:10.118612916-015-0319-2 PMID:25889598

Davenport, T. H., & Dyché, J. (2013). *Big data in big companies*. International Institute for Analytics.

Dixon, P. (2017). A Failure to "Do No Harm"—India's Aadhaar biometric ID program and its inability to protect privacy in relation to measures in Europe and the US. *Health and Technology*, *7*(4), 539–567. doi:10.100712553-017-0202-6 PMID:29308348

Dremel, C., Wulf, J., Herterich, M. M., Waizmann, J. C., & Brenner, W. (2017). How AUDI AG Established Big data Analytics in Its Digital Transformation. *MIS Quarterly Executive*, *16*(2), 81–100.

Dutta, R., Das, A., & Aryal, J. (2016). Big data integration shows Australian bush-fire frequency is increasing significantly. *Royal Society Open Science*, *3*(2), 150241. doi:10.1098/rsos.150241 PMID:26998312

Gaardboe, R., Svarre, T., & Kanstrup, A. M. (2015). Characteristics of business intelligence and big data in e-government: Preliminary findings. *Innovation and the Public Sector*, 109.

Goldstein, P. (2019). *Data Lakes: What They Are and How They Can Benefit Feds*. Retrieved from https://fedtechmagazine.com/article/2019/01/data-lakes-what-they-are-and-how-they-can-benefit-feds-perfcon

Gruschka, N., Mavroeidis, V., Vishi, K., & Jensen, M. (2018, December). Privacy issues and data protection in big data: a case study analysis under GDPR. In *2018 IEEE International Conference on Big data* (pp. 5027-5033). IEEE. 10.1109/BigData.2018.8622621

Hay, S. I., George, D. B., Moyes, C. L., & Brownstein, J. S. (2013). Big data opportunities for global infectious disease surveillance. *PLoS Medicine*, *10*(4), e1001413. doi:10.1371/journal.pmed.1001413 PMID:23565065

Heaphy, L. (2019). Data ratcheting and data-driven organisational change in transport. *Big Data & Society*, *6*(2), 1–12. doi:10.1177/2053951719867359

Heudecker, N. (2015). *Big data Isn't Obsolete. It's Normal*. Retrieved from https://blogs.gartner.com/nick-heudecker/big-data-is-now-normal/

HMRC. (2019). *Tax gap remains low*. Retrieved from https://www.gov.uk/government/news/tax-gap-remains-low

Höchtl, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, *26*(1-2), 147–169. doi:10.1080/10919392.2015.1125187

Internal Revenue Service. (2019). *The Tax Gap*. Retrieved from https://www.irs.gov/newsroom/the-tax-gap

Jnr, B. A., Petersen, S. A., Ahlers, D., & Krogstie, J. (2020). Big data driven multi-tier architecture for electric mobility as a service in smart cities. *International Journal of Energy Sector Management*. . doi:10.1108/IJESM-08-2019-0001

Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. *IT Professional*, *15*(6), 43–48. doi:10.1109/MITP.2013.61

Kepes, B. (2015). *How big data predicts and helps prevent train delays in Sweden*. Retrieved from https://www.networkworld.com/article/2982447/how-big-data-predicts-and-helps-prevent-train-delays-in-sweden.html

Khan, M. A., Uddin, M. F., & Gupta, N. (2014, April). Seven V's of Big data understanding Big data to extract value. In *Proceedings of the 2014 zone 1 conference of the American Society for Engineering Education* (pp. 1-5). IEEE. 10.1109/ASEEZone1.2014.6820689

Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. *Communications of the ACM*, *57*(3), 78–85. doi:10.1145/2500873

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

Kitchin, R. (2016). Getting smarter about smart cities: Improving data privacy and data security. Data Protection Unit, Department of the Taoiseach, Dublin, Ireland.

Klievink, B., & Janssen, M. (2009). Realizing joined-up government—Dynamic capabilities and stage models for transformation. *Government Information Quarterly*, *26*(2), 275–284. doi:10.1016/j.giq.2008.12.007

KPMG. (2019). *Harvey Nash/KPMG CIO Survey – A Changing Perspective*. Retrieved from https://home.kpmg/xx/en/home/insights/2019/06/harvey-nash-kpmg-cio-survey-2019.html

Laney, D. (2001). *Application delivery strategies*. META Group.

Lnenicka, M., & Komarkova, J. (2019). Developing a government enterprise architecture framework to support the requirements of big and open linked data with the use of cloud computing. *International Journal of Information Management*, *46*, 124–141. doi:10.1016/j.ijinfomgt.2018.12.003

Luengo, J., García-Gil, D., Ramírez-Gallego, S., García, S., & Herrera, F. (2020). *Big Data Preprocessing: Enabling Smart Data*. Springer Nature. doi:10.1007/978-3-030-39105-8

Mashey, J. R. (1997, October). *Big data... And the next wave of infrastress*. Paper presented in Computer Science Division Seminar, University of California, Berkeley, CA.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: The management revolution. *Harvard Business Review*, *90*(10), 60–68. PMID:23074865

Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big data in public affairs. *Public Administration Review*, *76*(6), 928–937. doi:10.1111/puar.12625

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, *98*, 261–276. doi:10.1016/j.jbusres.2019.01.044

Munné, R. (2016). Big data in the public sector. In New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe (pp. 195-208). London: Springer Open. doi:10.1007/978-3-319-21569-3_11

Norheim-Hagtun, I., & Meier, P. (2010). Crowdsourcing for crisis mapping in Haiti. *Innovations: Technology, Governance, Globalization*, *5*(4), 81–89. doi:10.1162/INOV_a_00046

Pastorino, R., De Vito, C., Migliara, G., Glocker, K., Binenbaum, I., Ricciardi, W., & Boccia, S. (2019). Benefits and challenges of Big data in healthcare: An overview of the European initiatives. *European Journal of Public Health*, *29*(Supplement_3), 23–27. doi:10.1093/eurpub/ckz168 PMID:31738444

Patel, B., Roy, S., Bhattacharyya, D., & Kim, T. H. (2017). Necessity of big data and analytics for good e-governance. *International Journal of Grid and Distributed Computing*, *10*(8), 11–20. doi:10.14257/ijgdc.2017.10.8.02

Pauwels, K. (2014). *It's Not the Size of the Data—it's how You Use it: Smarter Marketing with Analytics and Dashboards*. Amacom.

Poiani, T. H., dos Santos Rocha, R., Degrossi, L. C., & de Albuquerque, J. P. (2016, January). Potential of collaborative mapping for disaster relief: A case study of OpenStreetMap in the Nepal earthquake 2015. In *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 188-197). IEEE 10.1109/HICSS.2016.31

Pradhan, P., & Shakya, S. (2018). Big data Challenges for e-Government Services in Nepal. *Journal of the Institute of Engineering*, *14*(1), 216–222. doi:10.3126/jie.v14i1.20087

Rajagopalan, M. R., & Vellaipandiyan, S. (2013, November). Big data framework for national e-governance plan. In *2013 Eleventh International Conference on ICT and Knowledge Engineering* (pp. 1-5). IEEE. 10.1109/ICTKE.2013.6756283

Rao, T. R., Mitra, P., Bhatt, R., & Goswami, A. (2019). The big data system, components, tools, and technologies: A survey. *Knowledge and Information Systems*, *60*(3), 1165–1245. doi:10.100710115-018-1248-0

Richter, P. C., & Brühl, R. (2020). Ahead of the game: Antecedents for the success of shared service centers. *European Management Journal*, *38*(3), 477–488. doi:10.1016/j.emj.2019.10.006

Rubin, V., & Lukoianova, T. (2013). Veracity roadmap: Is big data objective, truthful and credible? *Advances in Classification Research Online*, *24*(1), 4.

Sarker, M. N. I., Hossin, M. A., Frimpong, A. N. K., & Xiaohua, Y. (2018, August). Promoting information resource management for e-government through big data approach. In *Proceedings of the 2018 International Conference on Information Management & Management Science* (pp. 99-104). 10.1145/3277139.3277155

Saxena, D., & McDonagh, J. (2019). Evaluating ERP implementations: The case for a lifecycle-based interpretive approach. *Electronic Journal of Information Systems Evaluation*, *22*(1), 29–37.

Shukla, P. S., & Mathur, M. (2020). Conceptualizing the Role of Data Analytics and Technology in E-Governance: An Insight. *International Journal of Business Analytics*, *7*(2), 1–12. doi:10.4018/IJBAN.2020040101

Singh, P., Sawhney, R. S., & Kahlon, K. S. (2018). Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government. *ICT Express*, *4*(3), 124–129. doi:10.1016/j.icte.2017.03.001

Stagnaro, C., & Leoni, I. B. (2019). Second-generation smart meter roll-out in Italy: A cost-benefit analysis. Paper presented in Eurelectric Power Summit 2019, Florence, Italy.

Tammel, K. (2017). Shared services and cost reduction motive in the public sector. *International Journal of Public Administration*, *40*(9), 792–804. doi:10.1080/01900692.2016.1204617

Ushakova, A., & Mikhaylov, S. J. (2020). Big data to the rescue? Challenges in analysing granular household electricity consumption in the United Kingdom. *Energy Research & Social Science*, *64*, 101428. doi:10.1016/j.erss.2020.101428

Van Oort, N., & Cats, O. (2015, September). Improving public transport decision making, planning and operations by using big data: Cases from Sweden and the Netherlands. In *2015 IEEE 18th international conference on intelligent transportation systems* (pp. 19-24). IEEE. 10.1109/ITSC.2015.11

Van Rijmenam, M. (2013). *Why the 3v's are not sufficient to describe big data*. Retrieved from https://datafloq.com/read/3vs-sufficient-describe-big-data/166

Walker-Munro, B. (2020). Use of Big Data Analytics by Tax Authorities. In M. Jackson & M. Shelly (Eds.), *Legal Regulations, Implications, and Issues Surrounding Digital Data* (pp. 86–110). IGI Global. doi:10.4018/978-1-7998-3130-3.ch005

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356–365. doi:10.1016/j.jbusres.2016.08.009

Woerner, S. L., & Wixom, B. H. (2015). Big data: Extending the business strategy toolbox. *Journal of Information Technology*, *30*(1), 60–62. doi:10.1057/jit.2014.31

Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big data and cloud computing: Innovation opportunities and challenges. *International Journal of Digital Earth*, *10*(1), 13–53. doi:10.1080/17538947.2016.1239771

Zavattaro, S. M., French, P. E., & Mohanty, S. D. (2015). A sentiment analysis of US local government tweets: The connection between tone and citizen involvement. *Government Information Quarterly*, *32*(3), 333–341. doi:10.1016/j.giq.2015.03.003

Zhang, Y., Huang, T., & Bompard, E. F. (2018). Big data analytics in smart grids: A review. *Energy Informatics*, *1*(1), 8. doi:10.118642162-018-0007-5

Zhou, K., Fu, C., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable & Sustainable Energy Reviews*, *56*, 215–225. doi:10.1016/j.rser.2015.11.050

Zhu, L., Yu, F. R., Wang, Y., Ning, B., & Tang, T. (2018). Big data analytics in intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, *20*(1), 383–398. doi:10.1109/TITS.2018.2815678

ADDITIONAL READING

Baig, A., Cadena, A., Chui, M., & Dua, A. (2014). *Innovation in local government: Open data and information technology*. McKinsey & Company.

Cavanillas, J. M., Curry, E., & Wahlster, W. (Eds.). (2016). *New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe*. Springer. doi:10.1007/978-3-319-21569-3

Dey, N., & Tamane, S. (2019). *Big data Analytics for Smart and Connected Cities*. IGI Global. doi:10.4018/978-1-5225-6207-8

Information Resources Management Association. (2016). *Big data: concepts, methodologies, tools, and applications*. IGI Global.

Radnor, Z. J., Bateman, N., Esain, A., Kumar, M., Williams, S. J., & Upton, D. M. (Eds.). (2015). *Public service operations management: A research handbook*. Routledge. doi:10.4324/9781315747972

Sandhu, K. (Ed.). (2019). *Leadership, Management, and Adoption Techniques for Digital Service Innovation*. IGI Global.

Sun, Z. (Ed.). (2019). *Managerial Perspectives on Intelligent Big data Analytics*. IGI Global. doi:10.4018/978-1-5225-7277-0

KEY TERMS AND DEFINITIONS

Big Data: Datasets characterized by Volume, Velocity, Variety, Variability, Veracity, Visualization, and Value.

Digital Enterprise Transformation: End-to-end value creation for the enterprise via the management and application of digital technology aligned with business strategy.

E-Governance: Application of digital technology for the purpose of efficient, effective, and responsive public governance.

Enterprise: An organizational entity with a definite objectives, scopes, and operational rules. An enterprise focus may be conducting a business or providing public services.

Management: The set of principles and actions geared toward implementing the strategy to attain the objectives of the enterprise.

Public Service: Essential service (e.g., health, electricity, law and order), that are required by the citizens for proper functioning of the society.

Strategy: A long term direction outlining the value proposition and growth of the enterprise to attain its objective.

Value: Intended outcomes, for instance efficiency or reputational gains, made possible by Big data analytics.

Variability: Variation in the meaning when processing natural language data.

Variety: Various forms of data in disparate formats.

Velocity: The speed with which data is generated.

Visualization: Visual interpretation of data and analysis to assist in decision-making.

Volume: Amount of data generated.