Deep Science Publishing
https://doi.org/10.70593/978-93-49910-54-6_3

DeepScience
Open Access Books

Chapter 3: Utilizing big data to strengthen financial oversight and increase transparency in public revenue

3.1 Introduction

Effective management of public resources is crucial to ensuring the smooth functioning of society, as well as the provision of quality goods and services to citizens. Thus, governments must act efficiently and responsibly, properly allocating and utilizing the resources of each nation. That said, the reality is quite different everywhere across the world, with numerous government officials being accused or even prosecuted for embezzling public funds every year. This raises a legitimate question: how is it possible that civil servants in charge of overseeing the correct use of public resources can commit such acts with impunity? The answer lies in the difficulties of the auditing process, which is generally conducted in an off-line, retrospective fashion. Consequently, auditing is often conducted months or years after transactions and the official in charge of certifying that the money has been spent correctly is usually the same as the one who has decided how the funds will be spent, making it difficult to avoid conflicts of interest in the process (Kassen, 2014; Kim et al., 2014; Janssen et al., 2017).

Developments in information technology in recent decades have contributed to the emergence of big data concepts and techniques, and the use of big data analytics has mushroomed in the fields of marketing, health, management, and education, inter alia. However, the concept of using such data in auditing has only very recently been put forward and has not yet been practically implemented anywhere. This paper explores the potential contribution of big data and its analysis to making public financial oversight and auditing more efficient. Important sources of big data already exist, such as online service directories, payment systems, and e-commerce platforms, among many others. The web has evolved into a database containing a wealth of information on economic

activity, including email exchanges, publication of company information like their tax identification number, and identification of suppliers through brochure publication services, among others. Such information can be utilized to conduct public audits more dynamically, guiding the allocation of resources deeper into the monitoring process (You, 2012; Zuiderwijk et al., 2015).

3.1.1. Overview of the Study

Research and development activities in the public sector are traditionally burdened by high transactional and supervisory costs. The development of information technologies in general and the emergence of big data and machine learning algorithms, in particular, will allow us to make vast strides in the total or partial elimination of these costs, with the additional advantage of their application to the analysis of information at a microscale level, which is more complete and precise compared to currently available public data. The objective of this paper is to present an overview of the main themes, technologies, and levels of application of big data in the public sector, especially concerning its applications in public financial management, and public revenue management in particular. The paper is a work in progress, which eventually aims to be a checklist for future researchers, or even a roadmap for public managers who wish to capitalize on these new instruments. It is organized in the following way: First, we explore the extent of the application of big data in the public sector, and how the current state of technological development and the availability of resources have opened the possibility and desire for greater innovation by public institutions. We then discuss what big data is and what its main applications to public revenue management are, particularly in the areas of tax policy design, tax fraud detection and audit support, and the enhancement of citizen relations. Finally, we present a discussion of the advantages, problems, and limitations associated with its use.

The ability to collect, store, and analyze vast quantities of information about all aspects of social life opens extraordinary opportunities to strengthen the analysis and evaluations that support the decision-making process. In particular, policy design and ex-ante or ex-post evaluation of efficiency and sustainability will benefit from an analysis of effects and identification of causal relations, which are more robust than the results based on pretty or quasi-experimental, or classic econometric techniques. The availability of technology for the profound transformation of processes applied by some large private services companies since the onset of the digital revolution allows us to carry out a critical reengineering of internal control processes. The increasing access to new forms of data holds disruptive potential to stimulate a change in the way forensic services are provided. The formal incorporations of statisticians and data scientists to audits carried out by private firms, the development of programs to process unstructured information

as well and the development of specific tools to monitor economic and financial matters used by the specialized services of public service institutions, suggest that its implementation by public auditors is near.

3.2. Understanding Big Data

1. Definition and Characteristics

Big Data is characterized by its volume, velocity, and variety, among other limiting factors, which create a data overflow concerning the processing capacity of traditional mechanisms. There is no single, commonly accepted definition of Big Data. It refers to data sets that are so large or complex that traditional data processing applications are inadequate. The most popular definition describes it as high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision-making, insight discovery, and process optimization. Big Data is defined in terms of the volume and complexity of the data, the need for new technologies and skills to uncover its value, its potential for innovation, and its implications for public policy, which is related to the availability of data at a low cost, as well as its use in new techniques that create simulations, predictions, and discoveries. It generally refers to rapidly growing amounts of data that cannot be managed and processed by traditional means and that cannot be fully exploited by classical methods.

Fig 3 . 1 : Understanding Big Data

2. Sources of Big Data in Finance

The data sources that are capable of producing Big Data come from various existing databases, which originate from multiple infrastructures: mobile monitoring systems, e-commerce sites, social networks, and sensors, among others. There is ongoing social interaction. In public finance, data is produced and made available through the websites

of the Executive, Legislative, and Judicial branches, as well as autonomous public entities, both in terms of their outputs and inputs.

3.2.1. Definition and Characteristics

Big Data can be defined as a dataset with a predictive and/or descriptive power with such a volume, velocity, or variety that requires specialized technology or systematic processes to enable insights that are beyond the ability of commonly used analytical or decision-making processes. While technological and analytical capabilities are growing at pace, we are far away from being able to extract insights from any dataset. In other words, Big Data can be viewed as the "overflow" of the ability to use common analytical tools. Therefore, Big Data is not only about what we are used to think about when we think about "Big Data": texts, images, and social media interactions, but also any dataset that goes over what we usually analyze, such as complex financial derivatives, large transaction flows, or any database that requires unique processing to be functional and usable. However, wealth does not only mean size: the challenge is threefold. Huge datasets challenge our systems, meaning we have to use specific technology, computing power, and specialized analytic tools. These systems should be connected to the decision-making process to enable timely and reactive decisions based on insights generated through Big Data Analytics.

This connection is what defines Big Data as the overflow: it is what is challenging organizations around the world, to what extent they should consider alternative technology to infrastructure and stores decision-making systems to be able to process these datasets. Only when this advanced processing of the data transforms organizations' capacity to act, we can say we are engaging in Big Data. The remaining areas are simply more challenging data-driven operations. In Finance, like in other social science disciplines, we are in for a transformation.

3.2.2. Sources of Big Data in Finance

Financial institutions collect a wide range of data through their operations. Financial information is only one subset of existing big data as a preparation for implementing big data technology. A jumble of financial assets, the rapid growth of financial technology, and the proliferation of financial big data have created a paradigm shift in financial activities. Although it is crucial to assess stored big data to maximize its value, there is a limited supply of theoretical and empirical studies that lead to an appropriate classification for financial big data. A hierarchical taxonomy is developed based on the building blocks of financial big data to facilitate a discussion of businesses' problems,

challenges, and capabilities. After reviewing and discussing key issues in financial sources of big data, we propose possible future research directions.

Financial big data can be applied to a growing number of financial conglomerates and institutions. Such diverse works have combined with data from elsewhere to provide innovations in creating and delivering financial services. However, depending on the geographic area and scale of their development, there are differences in how financial big data is used. Today, innovation frontiers are cross-border implementation of big data technologies. Such technologies have shown effects in finance industry sectors, including online banking, e-commerce, automated trading, investment advisory, insurance brokerage, and risk management. Collecting and using financial big data from diverse sectors involves a diverse construct of the financial platform itself as well as a compilation of sources of financial big data.

3.3. The Role of Big Data in Financial Oversight

When dealing with large amounts of data, stakeholders involved in financial enforcement procedures have a vital role to play, intelligently managing the technology used and the procedures adopted to promote their strategic objectives. For financial oversight activities to pursue their purpose and be effective, they have to be characterized by technical and professional aptitude, co-responsibility as a data-oriented strategic choice of the organization, knowledge of risk criteria, and collaborative relations among different public authorities and with the private sector. In this regard, given their technical skills and professional standing, the audit offices have a proactive role in making the most of the enormous opportunities that big data can offer to the whole public sector. The increasing availability of information is expected to impact greatly on audit activities.

Due mainly to recent progress in data storage and processing technologies, big data favors the development of potentialities aimed at monitoring databases, at controls being carried out mainly during the financial year, and at reducing sampling and reference periods, with a new approach to the assurance of statements. Its use, therefore, enables higher quality audit planning procedures in identifying risks and direction of internal check systems, laying the basis for more efficient and effective procedures and relations with the other internal and external actors involved. Moreover, the transition to digital data incorporating tagging allows audits to be carried also through tools powered with Artificial Intelligence capable of operating autonomously considering predefined configurations and parameters. The pervasive use of computer systems for the accounting of financial statements produces, therefore, huge databases that agencies can exploit through big data-enhanced decision-making.

3.3.1. Enhancing Audit Processes

With its speed and breadth of power and capability, Big Data, through the use of Foundational Data Assets, has novel potential to radically transform audit processes, possibly completely automating aspects of the work currently engaged in by human auditors, analysts, and investigators. Big Data offers the opportunity to capture all of an auditor's traditional substantive procedures to test the entire population of transactions governing the revenue data as opposed to testing small samples, a qualitative game-changer for audits. In particular, advances in data architecture and technology allow audit processes to ingest, maintain, and process a volume of financial and non-financial data much larger than ever before. Such advances to many former limits on speed and volume offer the chance to employ analytical techniques that were not feasible only a decade ago. These advances allow the utilization of new taxonomies within the transactions to identify risk and outlier test areas. Non-financial transactional data that lie within the growing volume of data can improve either categorical probabilities or formal prediction models. These systems could operate fully or partially independently to identify outliers and risk areas. Such results would feed into some level of human-machine teams to analyze, evaluate, and prioritize cases for further validation or possible prosecution. The reduced costs from data that are either claimed or designed to be low-cost, efficient, and tax-facilitating—the increased probability of intrinsic low-risk transactions, and integrated data systems and analysis fed and supported by the use of data all mean that governments should receive a more substantial proportion of their potential revenues with higher levels of intrinsic compliance.

3.3.2. Risk Assessment and Management

In its broadest sense, risk management is a discipline aimed at helping identify, assess, and prioritize risks in a way that avoids negative and enhances positive outcomes. This may involve limiting exposure to particular assets, developing business continuity strategies aimed at dealing with critical hazards or otherwise engaging in systematic efforts to keep risk levels within planned or acceptable limits. Within government, risk is managed on multiple levels; agencies face risks associated with the way they conduct their mandates, while the government also faces risks associated with the impact of agencies' work on constituents. Internal controls are generally capable of managing the first type of risk, while oversight is concerned with the second. Traditionally, both internal control and oversight have been more or less independent of each other. Agency management planned and implemented specific internal control frameworks, while controllers and auditors recommended enhancements based on views formed from assessing exposed vulnerabilities. With new IT developments, such as Big Data and Data Analytics, risk management evolutions are currently shaping up different types of

responses. Agency management systems can now be supplemented with oversight conducted in real time. Accountable executives can use interactive dashboards, with summary data analyzed using a risk-based approach, to constantly assess whether known risks are under control and whether they need to adjust their risk management plans.

3.4. Increasing Transparency in Public Revenue

The greatest need for budgeting, in terms of both oversight and data availability, continues to be present in public revenue. This is partly due to a historical imbalance: in the past, revenue was thought to be a technical function performed by specialized public officials, little interested in or needing any sort of supervision, while expenditures, considered more discretionary and prone to political pressures, required a more politicized control structure. But it is also the case that, despite the proliferation of instruments to improve budget formulation and monitoring on the spending side, the processes of revenue collection and disbursement remain opaque in many countries. Many stakeholders need to understand where funds are coming from, how they are being used, and why. This section provides some guidance on transparency in revenue processes.

Public Access to Financial Data

All financial data should be available to the public, at a level of detail that will meet the needs of both laypersons and specialized analysts. Such data can be made available in a coherent system designed for that purpose, or in any one of the excellent systems provided by the private sector, especially specialized Foundations. Macroeconomic data may be necessary to present contextual information and to compare the performance of different jurisdictions. The allocation of expenditures is generally of more interest to the public than the sources of revenue. All data needed to estimate balances need explanation but must also be available: revenue collected to date, expenditure disbursed to date, and outstanding obligations.

3.4.1. Public Access to Financial Data

The findings on information asymmetries affecting the relations between political actors and the electorate, and on the importance of reliable information for the functioning of democratic systems, suggest that government financial data be made accessible to the public. In principal-agent models of the delegation of tasks, the agent is responsible for the execution of the task (e.g. the collection of mandatory revenues). In addition, earmarking makes agents accountable for the allocation of the funds raised, in this case, government revenues. In these models, the sharing of information (by the principal

towards the agent, and by the agent towards the principal) reduces moral hazard, that is, the risk that the agent will take advantage of the asymmetry of information and deviate from activities beneficial for the principal. Access to fiscal data could be intensified as the supervision of funds flows creates a second line of checks and balances – supporting the work of specialized independent institutions – on the work of agencies.

A more formal empirical examination of the institutional structures privileging the role of the electorate is also suggested by the empirical work on these questions. Publicly accessible financial data on revenues, earmarked expenditures, and outputs of tax collection activities, as well as on the revenue allocation to branches of government and bodies responsible for the collection and allocation of the funds, facilitate the work of informed analysts within the electorate. They help compare the performance of government bodies dealing with these activities with that of similar branches and bodies in other countries or with similar populations in terms of their size and demand for public goods. Comparing performance and identifying underperformers is essential for the establishment of the informational feedback necessary for the efficient and accountable functioning of a democratic government.

3.4.2. Real-Time Monitoring of Revenues

While timely data may be useful for analysis and evaluating accountability, it alone is not enough to strengthen financial oversight. Public finance is, however, often not static, expenditures can change every day, and we need up-to-date information about available

Fig 3 . 2 : Real-Time Monitoring of Revenues

funds at any point in time. It is, therefore, not only desirable but also possible to receive real-time information that could facilitate timely oversight and monitoring of actual

financial developments and their comparison with budgetary plans. Real-time monitoring may be facilitated through big-data initiatives. Sensor data-generating technologies that monitor behavior and activities in the real world, offer a low-cost alternative to periodic surveys. Automatic language processing and data-mining techniques have lowered the costs of text-based data collection and analysis, such as sentiment analysis. Optical character recognition allows computers to read text in images of natural language and has lowered the costs of converting books, newspapers, and academic journals into machine-readable data. Cost and quality constraints have also led to the development of large global proprietary databases that charge fees for access. In addition, the growing preponderance of publicly available open-source data offers opportunities for no-cost or low-cost data collection by the public sector or by non-profit non-governmental organizations. The Government Financial Officers Association has for years been working with its members to develop best practices for the use of technology to make government financial information more accessible and comprehensible to the general public. It has been working to identify how best to take advantage of the new research programs being developed by increasingly powerful data-processing technology and widely available access to the general public keystones of increasing government transparency.

3.5. Technological Frameworks for Big Data Analytics

Big Data technology tools represent the foundation that enables the storage, recording, processing, and analysis of data in ways that were not previously possible. The increasing volume, velocity, variety, and complexity of data, and the low cost of data storage coupled with innovations in machine learning and artificial intelligence open up new possibilities for effectively utilizing data to help with decision making, as well as with the discovery of insights, knowledge, and intelligent action, while increasing the reliability of the analysis of Big Data. The primary goal of the present work is to propose innovative ways to effectively utilize Big Data to help with solving important Government problems in the area of Financial Oversight and Transparency, which is discussed in detail in the subsequent sections. Our goal is to provide tools at the forefront of innovation to assist Governments in their efforts to protect public funds. In this effort, we will consider critical issues such as the size and diversity of datasets, and the need for near real-time data processing and automated decision-making for Actionable Intelligence and discovery of knowledge. The objective is to provide an Automated Intelligence Assisting Tool that takes as input varied public domain data and provides as output, inquiries, answers, and knowledge that serve as input for human decision-making. The results of this effort, as a practical contribution, would take the form of efficient, modular, flexible, and extensible software models and tools that can efficiently perform the tasks of Data Accumulation, Transformation and Standardization,

Processing, and Analysis of these diverse datasets from various domains including Banking Transactions, Suspicious Transactions, Abridged Financial Statements, Corporate Ownership Structure and Relationships, and Social Network Ties, among others. These tools need to work with different scale sizes and types, depending mostly on the government population size, and the nature, complexity, and volume of the Government problems addressed.

3.5.1. Data Collection Techniques

The more traditional view of data only considered structured data. However, the fast growth of other types of Big Data, such as unstructured and semistructured types, as well as the amounting volume of data open to the world today has enticed a large mass of computer scientists and business administrators to look for new methods to collect, dissect, and make use of it. This has resulted in the development of new analytical techniques for dealing with many forms of Big Data coming from different sources. This data has been collected primarily due to technological improvements associated with interactive applications that create transactional data and digital sources of content and opinion that are available publicly. These forms of Big Data have blossomed in prominence and importance.

Discovering and extracting knowledge from nontraditional forms and sources of data from around the world are now the goals of many Big Data analytics initiatives. The available and established methods to discover and extract knowledge from Big Data aim to support efficient data management and analysis, allowing managers to make informed decisions based on insights, solutions, and recommendations. The basic technologies of Big Data that help organizations collect these kinds of information are Web scrapers and APIs. Altogether, they facilitate the access of large amounts of data contained within the World Wide Web and other open sources. In general, these types of data collection techniques function by performing a collection from predetermined and programmed data sources. When new information becomes available, it is automatically fetched and stored in the organization’s databases.

3.5.2. Analytical Tools and Software

Data collected from various sources rarely speak for themselves. They need to be processed, cleaned, organized, refined, and potentially transformed, stored, and curated before they can be analyzed for generating knowledge that can spark the building of a data culture. Hence, the timeless data lifecycle is essential to consider and make decisions about at each step. The data will be processed with analytical processes that

ensure both the principles of data ethics and the practices of data quality, security, and permanency are considered and worked to achieve.

Many tools abound to support various technical and analytic capabilities; the availability of digital pipelines has grown enormously. Therefore, the investigation and judicious selection of such tools are crucial to the usefulness and cost-effectiveness of any data engagement effort. We review below some of the more prominent companies offering some of the more popular and highest-rated solutions.

Analytics and the associated field of data journalism matured alongside the advances in data pipelines for transformation and storage, along with web services and tools that can be layered on top of and take advantage of big data technologies. The relatively recent formalization of the role of the data scientist in organizations and teams as a mixologist – someone who mixes expertise in computing, statistics, and business knowledge – is a clear indication of the increasing importance of good data skills in organizations. This increases the likelihood of successful insights-to-action cascading in organizations across the globe. We emphasize the use of organizations and corporations not simply to indicate the non-academic application domains that this technical report can cover, but to address the inherent goal of many data efforts, of generating non-academic, disruptive, and economic impact products and services.

3.6. Challenges in Implementation

Several challenges in using big data-analyze government spending and revenue may limit the success of this strategy. This section discusses these challenges, including data privacy concerns and integration of systems.

1. Data Privacy Concerns

Governments collect extensive amounts of data about individuals. In many cases, this data is collected in the course of regulating or providing services to citizens, so the citizens did not necessarily provide the data voluntarily. In addition, government agencies use advanced software to analyze these datasets, including predictive analytics on citizen behavior. Increasingly, there are concerns about government data use. Citizens expect privacy and are concerned that the government may misuse its data. Thus, many citizens want to impose limits on what the government can do with their data after it is collected. Their concerns with government data use stem from the inevitability that data mining tools could be utilized to discriminate against vulnerable populations, creating false accusations or stigma.

2. Integration with Existing Systems

Governments already generate considerable amounts of data regarding budget preparatory actions and execution activities using various information systems. In the majority of cases, these systems are implemented by different departments and agencies, are of different types and levels of sophistication, and are designed to meet specific local needs only. This widest of combinations creates severe barriers to the effort for reusing existing government data into big data analysis. This issue is thought to be one of the important constraints in making a useful real-time monitoring of government actions since the existing government systems are not interlinked to provide real-time information and are not designed to share data needed for effective auditing support and forensics. Ensuring that government department and agency systems are integrated systems that support a common data exchange standard and promote continuous monitoring of existing data in government systems is paramount.

3.6.1. Data Privacy Concerns

Data is the backbone of big data. In big data applications, data is abundant "and its manipulation is, in some ways, voyeuristic." Given the nature of the data, it is probable that many individuals would have privacy concerns, and it's currently possible to color data differently for various individuals. The general trend in society is to expect more privacy and IPR over personal information than is available under the current regime. One of the legal concerns with the use of personal data in data mining, pattern recognition, and machine learning is the potential to infer some information, and possibly private information, about an individual even when that individual's record has been omitted from the data set used to induce patterns, or even when the individual's data are included and anonymized in some way.

Another common concern regarding the use of data in data mining and machine learning is the risk of exposure of latent data – biometric attributes, political affiliations, etc. Declaring a variable as "sensitive" does not immunize it from a successful data mining attack. Some of these latent variables could be of serious concern – such as religious affiliation, political views, and sexual orientation. Big data can expose such latent variables; analysts have required only indirect mechanical signatures to infer these attributes.

3.6.2. Integration with Existing Systems

Evaluating the challenges in the implementation of Big Data tools could also help identify the conditions under which systems can be effectively developed and integrated into the day-to-day work of public finance managers. Integration with existing systems is a potentially complicated issue for information systems intended to be used by

government agencies with established procedures and processes. The legacy systems could either be fully proprietary or have custom-made components that are designed and integrated to work for a specific agency. This creates two fronts of concern in managing the transition to a more modern system. The first issue is to reconcile the creative facility built into Big Data systems with their open-source architecture and the standards such architectures impose in the interfaces, structures, and protocols in the specific agencies. The second concern is the sheer scale of operations and volumes deployed by legacy systems. Since their last upgrade, these systems may have absorbed and validated millions of static templates in addition to copious transaction record files.

The policy recommendation is a genuine integration approach since a near real-time flow of critical data on expenditures supplemented by periodic data uploads of observations is intended to support the legacy system used for pre-audit functions. The core function of any public financial management system is to ensure that no expenditure should be actually paid out except for verified and validated appropriations, and that no payments can be made except in satisfaction of validated debt obligations, and this can only be supported by a custom-designed high-security system. Data uploads should ideally be carried out by an intelligent data preparation system that optimally re-organizes large document-based files for indirect uploads into the data warehouses of Big Data systems, and not rely on manual data entry which necessitates a long run delay, considerable manpower allocation, and quality control resources both at the uploading and at the preprocessing end.

3.7. Policy Recommendations

The study recognized the lack of regulatory frameworks surrounding Big Data use in government and the need for security mechanisms so that Big Data does not have negative consequences. However, the aim was to explain ways in which Big Data can be both safely and powerfully used. To this end, this chapter seeks to discuss some policy recommendations that relate to the experiences of several countries and the result of the last chapter.

Not many governments have developed robust regulatory frameworks for using Big Data that apply to all government ministries. Recently, however, there have been some attempts by various governments to provide some regulatory clarity. While these frameworks focus on issues such as privacy and personal data, it is hoped that governments can expand their frameworks to cover potential negative outcomes from any Bias, Data Quality, Performance Instead of Results, Security, and Misinterpretation Risks that the use of Big Data can lead to. Several countries have already put some regulatory frameworks in place that seek to formulate regulatory practice guidelines. These efforts can provide some useful insights. So far, the European Union has led the

way in the development of guidelines on Big Data use. Several of these frameworks have focused on protecting privacy and obtaining users' consent. Other countries, however, are expanding these efforts to address risk factors that are specific to the use of Big Data in different government sectors.

Apart from regulatory frameworks, what would be more helpful for government ministries is detailed direct instructions from regulatory agencies on the exact measures and processes involved in managing Big Data collection and use for their government sectors. Data management activities could lead to numerous data outcomes, so government ministries in each sector could benefit from sector-specific guidelines on best practices. Some countries are already in the process of launching initiatives to provide detailed instructions on the best practices for data management specific to different sectors.

3.7.1. Regulatory Frameworks

To deepen the use of Big Data in the public sector and increase transparency in revenue matters, regulatory frameworks are proposed that promote, on one hand, the government's intent to share data and, on the other, explicit guidelines on some policy issues of considerable relevance for the good use of data. These frameworks are aimed at being minimal and flexible, so that local variations can be incorporated, according to the degree of development of the country. The frameworks are inspired by many developed countries' experiences, which already have a more advanced stage in terms of the implementation of open government data policies, in addition to having a tax administration with greater capacity to influence and promote the use of data by public entities.

Currently, in Latin America, some countries have developed institutional frameworks that provide support for initiatives that aim to increase the use of Big Data in decision-making. This is the case of countries such as Chile, Costa Rica, Colombia, Mexico, Peru, and Uruguay, which have adopted policies that promote the use of Open Government Data or the implementation of Digital Government strategies. The problem lies in the existence of cross-sectional actions aimed at exploiting the potentials offered by new information technologies. In this sense, there is little recognition of economic, social, and political aspects that limit the volume of executives and decision-making. As a general common denominator in the region, we can find high levels of availability and low-density use of open data. That is, the use of data portals is very variable in the Public Sector.

3.7.2. Best Practices for Data Management

Big data technology is not an absolute solution. It needs to be carefully adapted and adopted, usually together with a revised budget architecture, more investment in the training of the employed public servants, and the availability of adequate enabling technology. Therefore, the results achieved will depend critically on the choices made by those responsible for the institutional arrangements, and regulations, as well as the budget architecture affecting the financial oversight and external auditing of public expenditure. In this context, for big data technology to strengthen external auditing of public expenditure and more widely utilized big data to strengthen financial oversight and increase transparency in public revenue, the data-management solutions followed by public institutions and agencies, as well as the existing regulations and budget architecture that influence their choices assume exceptional importance.

The experience of the main external auditing agencies and bodies from all over the world indicates that they must follow a series of best practices for data management to guarantee the success of the technology investments made, adapting the existing frameworks to the new digital finance. First, because those agencies and bodies must manage, and then analyze, the increasing amount of data that are intrinsically generated by the satisfactory and efficient implementation of the frameworks, progressively eliminating the possibility of doing this task based on only or largely sampling the transactions analyzed, given the big number of digital footprints of public revenue and expenditure. Secondly, there is a growing demand for enhanced transparency from citizens and other stakeholders about the use of the resources allocated by external auditors. The indication of good practices for data management is based on the analysis of the requirement and functions of external auditors in carrying out the external audit of reported financial statements and transactions on public revenue and expenditure by government entities, bodies, and agencies, as well as the criteria to fully satisfy their accountability.

3.8. Future Trends in Big Data and Public Finance

The base of future public finance regulation and control is being laid today. In the foreseeable future, the role of new information technologies such as artificial intelligence, machine learning, and distributed ledger technologies will increase in both public finance management and the oversight of its transparency. They will not only enhance the state's capability to collect and analyze big data, including from nontraditional sources, but also improve the character of such analyses, as the quality of forecasting and expert assessments can significantly substantiate decision-making and limit the influence of human feelings and moods. These changes will allow us to move from merely monitoring available data to predictive analytics, alerting about potential risks, and diagnosing emerging problems in the automatic mode, with minimal human

involvement. These changes can lead to the appearance shortly of the following services and products in government activity. Governments will have a powerful tool for forecasting various negative phenomena. Such forecasting will be based on data from social networks, online resources, and demographic, economic, and sociological observations. Recommendation system for public actions. The analysis of publicly available big data, including data about previous negative phenomena, will make it possible to create a recommendation system that will tell what exactly the government should do to minimize adverse effects and secure public welfare. Automated, resident feedback systems. With the implementation of continuous two-way communication channels between the government and society, it will be possible to automate feedback on satisfaction with the quality of public services. Digital identity, ratings, and regulation. Shortly, individuals and companies will have their ratings. Digital space will reinforce their functions, and big data analysis will allow the private sector to meet greater transparency requirements.

3.8.1. Emerging Technologies

While experts disagree about the long-term viability of blockchain and AI technologies, we argue that the bulk of techniques commonly associated with Big Data are in their infancy and have many avenues for development. New technologies grow on one another, and as more tools evolve, the power and scope of what we can do with data expands. Cloud computing, edge processing, powerful computational architecture, and rapid feedback loops of action and reaction have made online services deeper and broader. Completely new areas of data are emerging. Currently, we are only scratching the surface of human interaction data; celebration and mourning; love and lust; altruism and violence; shared moments of space and time; interactions and intent; avarice and generosity. Soon we will be able to translate these into meaningful measures, find deeper interrelationships within and across subpopulations, and then accompany people throughout their life journeys. We can visualize how people change, as they shift their social and political interactions, modify social connections, migrate along with family and friends, adapt to historical events, and alter their economic consumption. Improved predictive capabilities, combined with these moves toward dollarized data, will yield even more commercial efficiencies; as a result, businesses will continue to allocate an increasing share of their capital budgets toward predictive technology.

In conjunction with these exciting new developments, there are five macro technology trends especially relevant to the potential use of Big Data in predicting human activities, decision-making, and infrastructure replacement. The first is the convergence of a wide spectrum of advanced technologies, such as artificial intelligence, life sciences, nanotechnology, cognitive science, and robotics. The combination of these technologies

will allow firms to harness group behavior at unprecedented levels of analysis. One of the more powerful positive scenarios for these technologies is the aggregation of small, focused organizations worldwide leveraging complex technology to do what was until now only possible for large businesses, at a lower cost. Using just-in-time delivery methods, these companies would take months rather than the years traditional prime contractors might spend building a facility.

3.8.2. Potential Impacts on Governance

Public finance is one of the most important functions of states. Consequently, oversight over public revenue should be crucial to all governments. It is through the collection of taxes and other revenue streams that states finance public goods needed for society to function. Furthermore, the regulation of the economy through macroeconomic policy as well as economic development through public investment should ideally work for the benefit of all. Citizens should have faith that their country is being governed correctly and that public service is being provided in the best interests of all, using resources available wisely. Unfortunately, the reality in many countries is far from this ideal. Financial crises, corruption scandals, and poor governance have plagued countless nations around the world. Citizens are increasingly distrustful of their governments, damaging the relationship of confidence that should exist in a democracy. Effectively

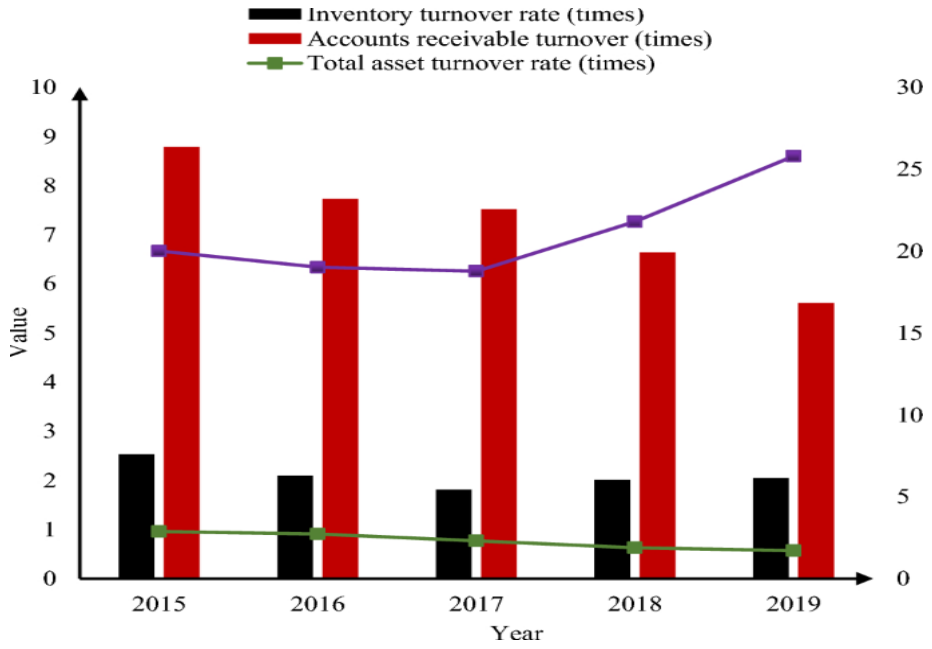

Fig 3 . 3 : Financial big data management and intelligence

governing additional problems requires data. State regulation involves supplementing or overriding individual or group behavior with rules and special charges. Addressing market failures, maintaining order, providing public goods, and redistributing income requires data and very often the cooperation of third parties, involving either monetary or non-monetary transactions. Democratically accountable institutions must coordinate these interactions for effective delivery. The advancement of information and communication technologies has made these interactions easier to track, share, and process by creating a system of fast communication between engaged parties. Data can now be transmitted and processed in real time by many institutions. These emerging technologies can strengthen the result-based governance of public finance institutions by holding them accountable and providing them with information to manage and execute budgets efficiently. Citizens need to be encouraged to capture these technologies, to understand how to better monitor the state, and to gather evidence to back up their claims. This requires partners from civil society, international organizations, the economy, and the public sector.

3.9. Conclusion

The increasing availability of real-time data in many areas of economic activity has offered great potential for the implementation of tools for the monitoring of economic phenomena. This capability finds strong application in the public sector where strong information asymmetries related to the management of citizen resources exist, and real-time data analysis could strengthen control actions and increase the transparency of other information related to public revenue. This paper has presented a comprehensive study of the possibilities offered by big data technology to create a new approach, complementary to traditional tools, to strengthen financial oversight and increase the transparency of public revenue management processes.

Given the numerous concerns expressed regarding the potential negative consequences of big data on issues such as the individual right to privacy and its possible misuse in creating dangerous social discrimination, increasing attention has also been paid to possible ethical spreading models that can be adopted. The present research suggests setting the creation of big data used for public benefit in the context of a public-private partnership approach. This model is deemed to be able to guarantee a more ethical approach than possible alternatives, thanks to the control activities that can be implemented by public actors, but also respecting the interests of the private actors involved, avoiding the emergence of mistrust potentially damaging to the social contract. Finally, in light of the latest events relating to the pandemic crisis and the war situation, there seems to be an urgent need for the development of tools capable of enhancing data and information sharing between different public functions and different countries, to

address the changes in the economic situation or welfare policies on the actions of public administrations.

3.9.1. Final Reflections and Implications for the Future

Big Data, Machine Learning, NLP, and DNNs, presumably combined into vast complex ecosystems, have made macroanalysis and microanalysis of massive datasets possible. Moving masses of data to the cloud or to anywhere is no longer a problem. The era of small data models is gone and we have entered the global space of providing all levels and domains of Digital Public Financial Information and budget tax revenue analysis and simulation hosts, in collaboration with digital platforms that do predictive simulations with extensive use of Big Data, Artificial Intelligence, Machine Learning models and their clouds, which interact with Loan Agreement Data and a growing number of countries implementing their Public Financial Management systems in the web 3.0 space. Digital countries, in some cases, also dive into these digital environments and allow an ecosystem of budgetary, tax revenue, and Public Financial Management macroeconomic information and credit simulations for predictive purposes. Countries embark on these digital ecosystems and even make available, as in the case of augmented cryptography, tax payments that can anticipate payments with predictive models of tax revenues for the budget year.

The automation of public revenue planning and management, the increase in the detail of the effects of tax policy, and the alliance of technological companies with universities, in the sharing with the tax agencies of the databases of the countries that have little face-to-face economic activity, allow us to foresee a future in which the taxpayer is almost an object of study to whom it is enough to analyze to fulfill the obligation since the possibility of error, avoidance, or surveillance by the administration has little effect. The transparency of the information in the Public Financial Management systems will reveal little by little what is left behind in the present.

References

Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors Influencing Big Data Decision-Making Quality. Journal of Business Research, 70, 338–345. https://doi.org/10.1016/j.jbusres.2016.08.007

Kassen, M. (2014). Big Data and Open Government: Perspectives on the Development of Open Public Data in Russia. Government Information Quarterly, 31(1), 1–11. https://doi.org/10.1016/j.giq.2013.07.007

Zuiderwijk, A., Janssen, M., & Dwivedi, Y. K. (2015). Acceptance and Use Predictors of Open Data Technologies: Drawing Upon the Unified Theory of Acceptance and

Use of Technology. Government Information Quarterly, 32(4), 429–440. https://doi.org/10.1016/j.giq.2015.09.005

You, C. (2012). The Big Data Opportunity: Making Government Faster, Smarter and More Personal. Policy Exchange. https://doi.org/10.2139/ssrn.2635825

Kim, S., Trimi, S., & Chung, J. H. (2014). Big-Data Applications in the Government Sector. Communications of the ACM, 57(3), 78–85. https://doi.org/10.1145/2500873