

Review

Not peer-reviewed version

A Review on Data Capture, Storage, Processing, Interoperability, and Visualization for the Smart Supervision of Public Expenditure

Jaime Alonso Restrepo-Carmona , Juan C. Zuluaga , [Manuela Velásquez](#) , [Carolina Zuluaga](#) , Rosse M. Villamil , Olguer Morales , Ángela M Hurtado , Carlos A. Escobar , [Julián Sierra-Pérez](#) , [Rafael E. Vásquez](#) *

Posted Date: 1 July 2024

doi: 10.20944/preprints202407.0044.v1

Keywords: Digital transformation; Smart government; Public expenditure; Fiscal control; Governmental accounting; Industry 5.0



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

A Review on Data Capture, Storage, Processing, Interoperability, and Visualization for the Smart Supervision of Public Expenditure

Jaime A. Restrepo-Carmona ¹, Juan C. Zuluaga ², Manuela Velásquez ³, Carolina Zuluaga ³, Rosse M. Villamil ², Olguer Morales ², Ángela M. Hurtado ², Carlos A. Escobar ¹, Julián Sierra-Pérez ^{1,3}, Rafael E. Vásquez ^{1,3,*}

¹ Corporación Rotorr, Universidad Nacional de Colombia, Cr. 45 26-85, Bogotá, 111311, Colombia; jarestrepoca@unal.edu.co (J.A.R.C.); ealejandro101@gmail.com (C.A.E.)

² Dirección de Información, Análisis y Reacción Inmediata, Contraloría General de la República, Cr. 69 No 44-35, Bogotá, 111071, Colombia; juan.zuluaga@contraloria.gov.co (J.C.Z.); rosse.villamil@contraloria.gov.co (R.M.V.); olguer.morales@contraloria.gov.co (O.M.); angela.hurtado@contraloria.gov.co (A.M.H.);

³ School of Engineering, Universidad Pontificia Bolivariana, Medellín 050031, Colombia; manuela.velasquezj@upb.edu.co (M.V.); carolina.zuluagaa@upb.edu.co (C.Z.); julian.sierra@upb.edu.co (J.S.P.); rafael.vasquez@upb.edu.co (R.E.V.)

* Correspondence: rafael.vasquez@upb.edu.co; Tel.: +57-4-448-8388

Abstract: Fiscal control and monitoring of public management are issues of great importance and relevance, as they focus on preventing and mitigating corruption, thus benefiting governments in their management and performance, and thus citizens. Given the large amount of data involved in the management and functioning of a government, it is pertinent to apply data analysis methods to contribute to fiscal control. This paper aims to investigate data management for this purpose, starting with a bibliometric study to demonstrate the relevance and pertinence of the current study. It examines existing data collection techniques that may facilitate such control, data storage considering its nature and the possibilities that contribute to this end, and data processing through techniques that allow obtaining adequate information for further analysis and decision making. The necessary architecture and interoperability between systems is also discussed, as the existence of large amounts and types of data in governments makes it crucial to promote these practices for accurate and truthful analysis. Finally, data visualization is addressed since this topic is highly relevant as it allows for the identification of anomalies, where proper visualization promotes appropriate analysis and thus informed decision making. This work is useful for modern audit and fiscal control entities, since data capture, storage, processing, interoperability, and visualization are crucial elements that collectively enhance the ability to supervise public expenditure effectively, ensuring that public funds are managed with the highest standards of transparency, accountability, and efficiency.

Keywords: digital transformation; smart government; public expenditure; fiscal control; governmental accounting; industry 5.0

1. Introduction

Today, the use and management of data in various fields has increased significantly, facilitating business and industry analysis for decision making. The amount of data generated is growing exponentially [1]. It is estimated that around 2.5 quintillion bytes of data are created every day [2]. In other words, data is fully available to those willing to mine it. Data is a valuable asset for businesses in the 21st century [3]. In 2006, Clive Humby, a British mathematician, coined the phrase 'data is the new oil'. Like oil, data is not valuable in its raw state; its value comes when it is collected quickly, completely and accurately, and linked to other relevant data [4].

The public sector and tax analytics are also seizing the opportunity to use data to combat tax evasion [5]. Public expenditure management is a critical government function that helps ensure

fiscal responsibility and accountability. The public budget, usually set for one year, is a fundamental plan of action that reflects the government's priorities and objectives through the amounts allocated for revenue and expenditure [6]. Public expenditure management ensures that resources are used efficiently and effectively and supports the delivery of essential public services. It is therefore a key element of good governance.

In addition, public expenditure management is essential for achieving macroeconomic stability and can help control inflation and reduce deficits [7]. It also promotes transparency and accountability of public finances [8,9]. By providing a clear and understandable picture of spending, citizens can better assess economic factors and their impact on society. This allows them to support or challenge spending decisions in a more informed way, while governments can demonstrate progress against their commitments [10].

Today, problems such as corruption, lack of transparency and inefficiency are among the main difficulties facing the public sector [11]. These problems impede the proper functioning and management of government, which has a negative impact on the development of a nation and the well-being of its citizens. The institutional design of local governments has a significant impact on poverty: increased risk of corruption and inefficiency in the use of transfers for education and health spending increase poverty. On the contrary, transparent governance can reduce it; in fact, a one percentage point increase in transparency and relative efficiency indices reduces the poverty rate by 0.6 percentage points [12].

The risk of corruption tends to increase with the size of the local state and decrease with improvements in fiscal performance, tax collection and average years of education [13]. It is essential to implement strategies to mitigate this situation, one of the most promising of which is the use of analytics and data management in the public sector [14]. Technology in public administration (e-government) has enabled the delivery of public services over the internet, promoting efficiency in data collection, processing and reporting, and improving decision making [15].

Advances in smart technologies, better informed and connected citizens, and globally interconnected economies have created new opportunities [16]. Governments are taking the concept of e-government to a new level by recognizing the power of data and heuristic processing through artificial intelligence to improve their services, interact with citizens, develop policies and implement solutions for the well-being of the community - becoming a smart government [17].

However, the diverse sources and formats of public sector documents make the collection, processing and organization of these documents challenging from a data analytics perspective [18]. It is important to develop approaches to managing these data, as their analysis allows for greater citizen involvement by giving them more access to public decisions and spending [19], increasing transparency in the public sector, and giving citizens a greater sense of accountability by providing different views on government performance in meeting its public policy objectives. As an example, [15] presented the use of artificial intelligence (AI) as a promising tool for intelligent monitoring of the public sector and combating tax evasion. It is argued that the availability of this information could help stakeholders to make better informed decisions. The use of AI would work like any other decision support system, providing several alternatives and a wealth of information to enable the final decision to be made.

More generally, some of the advantages of artificial intelligence techniques relate to their ability to analyze any data, regardless of its distribution, size or format [20]. However, some limitations of AI-based decision making are related to the computational capacity available at the time. For instance, [21] showed how reinforcement learning based on Markov models and partially observable computations of the behavior of a contributing agent can refine the analysis of an audit policy. They showed that by synchronizing procedures and dynamically updating intelligent behaviors, a hybrid model can be built that combines bottom-up agent-based execution with various partially observable intelligent behaviors. This can serve as a platform for testing the effectiveness of an audit policy and for training an intelligent auditor. [22] developed a framework for the value of data analytics

in government oversight that is intended to serve as a tool for analysis and understanding. Future research can focus on developing this framework into a theory for prediction, design and action. In their study, they found that collective capacity building processes in data analytics and their link to capacity processes in an individual organization are very interesting but not yet well understood.

This paper contributes to the field of risk and financial management by outlining how effective data management and processing can significantly enhance financial control and auditing practices. Through a comprehensive literature review, current research trends and key focal points in data capture, storage, processing, interoperability, and visualization are identified and discussed. The study emphasizes the critical role of accurate data collection and secure storage as foundational elements for reliable fiscal oversight. It provides insights into advanced data processing techniques that enable proactive decision-making in public expenditure management. Moreover, the paper underscores the importance of interoperable systems that facilitate seamless data exchange across government entities, promoting efficiency and transparency. Finally, the article advocates for the adoption of appropriate data visualization strategies to improve communication of financial insights and support evidence-based policy-making.

The organization of the paper is as follows. Section 2 contains the bibliometric analysis. Section 3 provides an overview of data capture techniques, data capture tools, types of data capture, and data capture hardware architecture used in organizations for data collection. Section 4 describes data storage options, on-premise and cloud. Section 5 contains the three stages in data processing: data cleaning, data transformation, and data analysis. Section 6 refers to the architecture that facilitates effective interoperability. Section 7 provides an overview of the basic principles and advanced technologies related to data visualization in the modern era. Then, Section 8 contains the discussion, and finally, conclusions are presented in Section 9.

2. Literature Review

The study employed a systematic literature review methodology comprising three main stages: defining the research question, searching for relevant studies, and selecting them. The first stage involved defining the research question to delineate the review's aim and scope, guiding the subsequent search and selection of pertinent studies. The literature review focused on analyzing the impact and role of data collection, storage, processing, interoperability, and visualization in enhancing smart monitoring of public expenditure.

This review involved a comprehensive search for articles using the following predefined search equation: (("DATA CAPTURE" OR "DATA STORAGE" OR "PROCESSING" OR "DATA VISUALIZATION") AND "PUBLIC" AND "SUPERVISION"). This equation was applied to Scopus and ScienceDirect, two prominent engineering databases. For instance, this process yielded 547 papers in Scopus. These articles covered the period from 2014 to 2024, ensuring a ten-year search window. To focus the scope and select the most relevant literature, the review included articles, conference papers, journal sources, and publications in English published after 2019. Ultimately, 403 documents met these criteria for detailed analysis.

The quality of the articles was then assessed, and the quantity was reduced. Articles were systematically ranked according to their relevance to the research question, number of citations, and date of publication. Priority was given to the most recent articles directly related to the research question, while older articles with more than five citations were also considered. As a result of this process, 115 articles were selected for the study.

2.1. Bibliometric Analysis

The bibliometric analysis enables the extraction of pertinent information regarding trends and quantitative research data using mathematical tools. For metadata analysis, the Scopus tool was employed to visualize publication behaviors, while VOSviewer was used to analyze keywords and their relationships.

The selected tool is noted for its user-friendly interface in developing thematic maps and conducting analyses such as clustering, keywords, and sources, among others. All findings discussed in this section are derived from primary literature sources. Figure 1 illustrates the publication trends over a 10-year search period. It is evident that the number of publications focusing on data management and fiscal control techniques has increased over time, indicating a trend towards digital strategies to enhance control and reduce fraud.

The global expansion of data handling and management is impacting not only sectors where data analysis is critical, such as financial analysis, power generation, and industrial plants, but also those that historically managed data differently. With the emergence of machine learning and related tools, the tax sector has significant opportunities to continually monitor data and mitigate tax fraud. Through the collection, processing, manipulation, and visualization of data, trends, patterns, and outliers indicative of potential fraud can be identified, enabling more effective actions to be taken.

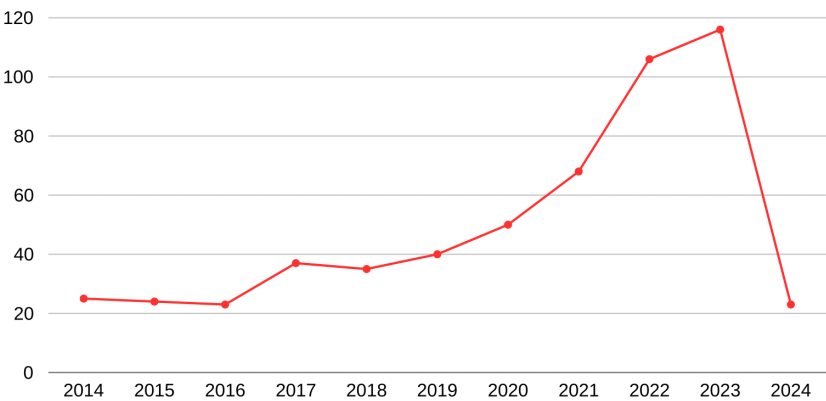


Figure 1. Publications over a 10-year search window.

Figure 2 shows a collection of common keywords extracted from the primary bibliography used for the bibliometric analysis conducted with VOSViewer. Two primary clusters were identified, interconnected, indicating coherence in the themes within the database. Prominent keywords are depicted by larger circles, highlighting terms such as 'fraud', 'information processing', 'crime', and 'machine learning'. Additionally, keywords like 'deep learning' and 'controlled study' are also visible. Connections are observed among issues such as fraud, crime, and specific government sectors like health, linked with data management and machine learning techniques. This underscores an increasing necessity to apply data management and analytics in the public sector to proactively address and mitigate fraud, reflecting current research interests in this area.

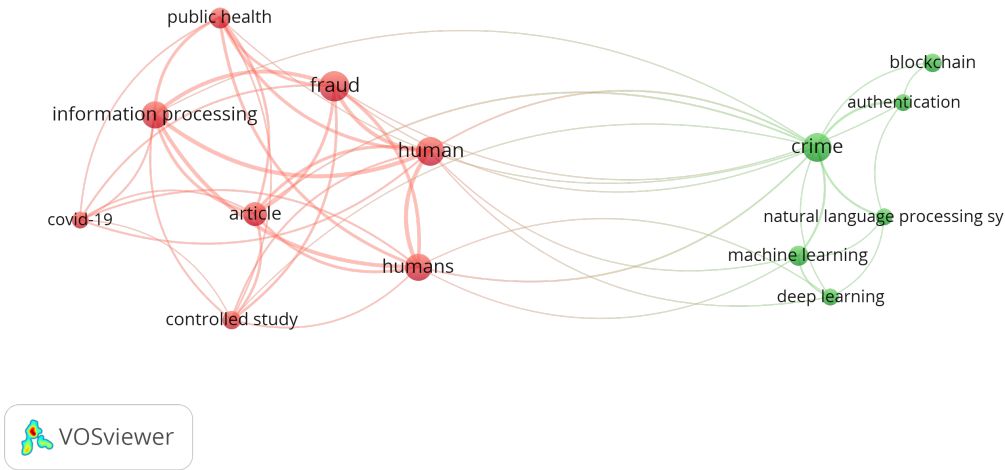


Figure 2. Clusters by keywords. Performed with VOSviewer 1.6.20.

Figure 3 illustrates the countries with the highest number of documents addressing these issues, with the United States notably leading, followed by China and the United Kingdom. This indicates a greater focus on research related to data management and the prevention, detection, and prosecution of corruption and fraud within these countries.

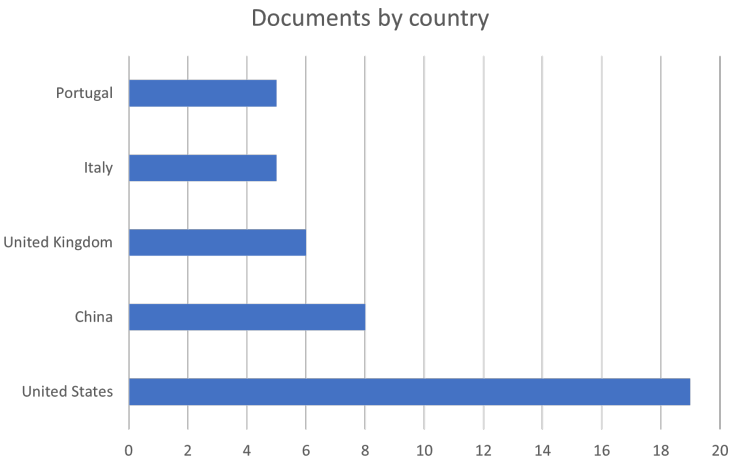


Figure 3. Documents by country.

3. Data Capture

This section provides an overview of data capture techniques, tools, types, and hardware architectures used in organizations for data collection. Before the advent of new technologies, data extraction was primarily performed manually, despite offering control and flexibility. However, manual data capture was susceptible to errors, lacked scalability, and often resulted in inconsistencies. Today, various technologies enable data extraction with varying levels of accuracy.

The primary goal of data capture is to enable processing, integration with other datasets, and subsequent transfer to databases. This extraction process assists organizations in identifying and leveraging valuable information for decision-making purposes. Nevertheless, like any operational activity, data capture presents challenges that need to be addressed. For instance, unstructured data requires standardization and formatting to be meaningful [23]. Data captured from diverse sources and

databases can introduce compatibility issues, necessitating the execution of an ETL (Extract, Transform, Load) process to ensure data compatibility [24].

3.1. Types of Data for Public Expenditure Monitoring

For effective monitoring of public spending, it is crucial to have access to specific data, which can often be challenging due to the dispersed nature of the data.

3.1.1. Public Budgets

Financial documents detail government revenues and expenditures for a given period, including allocations for different departments, programs, and projects. They are essential for assessing the efficiency of resource allocation and ensuring that funds are used according to the government's established priorities [25].

3.1.2. Government Contracts

These legal agreements between the government and suppliers define conditions, prices, delivery terms, and penalties for non-compliance. They play a key role in ensuring the transparency and integrity of procurement processes by verifying fair and competitive contract awards and ensuring compliance with contractual conditions. They also help prevent and detect possible cases of corruption and misappropriation of funds [26].

3.1.3. Public Bidding

Competitive processes in which the government solicits and evaluates bids from suppliers to provide goods or services. They include solicitations, bids received, evaluation criteria, and award results. Public bidding ensures the government obtains the best value for money by promoting competition and transparency [27].

3.1.4. Public Payments and Disbursements

These records encompass all financial transactions made by the government, including payments to suppliers, public employee salaries, transfers to other entities, and grants. They are crucial for monitoring the flow of public funds, ensuring payments align with approved budgets, and detecting possible irregularities or fraud [28].

3.1.5. Audits and Public Expenditure Control

Independent evaluations conducted by control agencies, such as comptrollers and external auditors, to review the legality, efficiency, and effectiveness of public resource use. They provide an objective assessment of the government's financial performance, identify areas for improvement, and prevent fraud and corruption [29].

3.1.6. Performance Indicators

These quantitative and qualitative metrics measure the performance of various programs and projects financed with public funds, including indicators of efficiency, impact, and beneficiary satisfaction. They help evaluate whether public resources achieve expected results and allow policy and program adjustments to enhance effectiveness [30].

3.1.7. Beneficiary Data

Information on individuals, communities, or organizations that directly benefit from publicly funded programs and projects, including demographic, socioeconomic, and geographic data. Essential for assessing equity and inclusiveness in the distribution of public resources to ensure they reach those most in need [31].

3.1.8. Transparency and Accountability Data

This encompasses public reports, performance indicators, and independent evaluations enabling citizens and oversight bodies to monitor and evaluate the use of public resources. Promotes citizen participation and accountability by providing accessible information on how public funds are utilized [32].

3.2. Data Capture Techniques

Full extraction is an ETL technique that extracts the entire dataset from the source system and loading it into the target database. This method ensures complete data transfer, making it suitable for scenarios where data integrity is critical. Nonetheless, this technique can be resource-intensive and time-consuming, and may impact system performance, especially with large datasets.

Incremental extraction, on the other hand, extracts only the data that has changed since the last extraction, typically triggered by discrete events such as a successful prior extraction or an update. This technique minimizes the volume of data transferred, reduces resource consumption, and speeds up the process. It is particularly efficient for handling large datasets and ensures that only new or modified data is processed.

Wrap extraction is a technique designed to manage and optimize the extraction process of large data volumes. It involves breaking down the data into smaller parts or batches to enhance efficiency and prevent overloading network or system resources. This method facilitates the handling of substantial datasets while maintaining performance and minimizing the risk of errors during data transfer.

3.3. Types of Data Capture

3.3.1. Batch Data Ingestion

Batch data ingestion involves capturing datasets from different sources at regular intervals. It is particularly suited for handling large volumes of data that do not require real-time processing. Data in batch ingestion processes can be grouped and sent to the system according to predefined criteria. One advantage of batch data ingestion is its cost-effectiveness. Additionally, it helps prevent network overload because data transmission occurs continuously but in controlled intervals [33].

3.3.2. Data Extraction from Documents

Data extraction from documents refers to the process of retrieving information from printed or digital documents. The objective is typically to identify and extract specific unstructured data, transforming it into structured data for storage and analysis. Automated document data extraction relies on a combination of techniques, tools, and algorithms to extract necessary data from complex documents. Key steps in this extraction process include [34]:

- Scanning high-quality documents or images improves OCR (Optical Character Recognition) output and extraction accuracy.
- Machine learning models should be updated and trained frequently using diverse and representative datasets, adapting them to new document layouts and formats, which improves extraction performance over time.
- Use a hybrid approach using rules for fields in structured data with predictable patterns in conjunction with machine learning algorithms to handle unstructured or complex data.
- Implement robust data validation to ensure data accuracy and integrity.
- Design the data extraction process to handle large volumes of documents without failure.

The automated document data extraction process enables organizations to efficiently process and retrieve data from multiple document types with minimal effort. It offers numerous advantages, including improved process efficiency, accuracy, accessibility, scalability, flexibility, and adaptability.

3.3.3. Web Data Extraction

Web data extraction, also known as web harvesting or web scraping, involves collecting data from websites using software that simulates HTTP or web browser interactions [35]. Web scraping is widely used by search engines to transform unstructured web data into structured data stored and analyzed in databases [35]. Often confused with web scraping, web crawling involves locating and storing different data from the internet, organizing it into a database based on relevant words or search keywords. Each word is associated with a hyperlink identifier, enabling traceability to its source [36]. This technique leverages data mining to analyze multiple statistical properties of collected information, facilitating the implementation of data monitoring services that generate user alerts. In essence, web data extraction seeks new or updated information [37].

3.3.4. Data Extraction from Enterprise Applications

The use of data mining and artificial intelligence has become indispensable for process analysis and generating smart recommendations within organizations. These tools enable accurate information extraction from processes and facilitate the collection and evaluation of behavioral patterns and preferences among specific groups [38]. Data mining in business applications involves searching, collecting, and analyzing data to uncover specific attributes, outcomes, or hidden patterns. For instance, in organizations, customer information is gathered through data mining, where raw data is stored and simultaneously transformed into actionable insights for both the organization and its customers [39].

3.3.5. Data Extraction from Logs and Records

Log files are plain text files typically reviewed during emergencies to analyze program or system errors, providing insights into system and user behaviors. Log data extraction tools enable an organization's IT teams to access all this information from a centralized location, eliminating the need to manage multiple software tools and simplifying troubleshooting tasks for swift issue detection and resolution. These tools facilitate determining which data to record, the format for storing data, the retention period for data storage, and optimal data deletion methods when data is no longer needed [40].

3.3.6. Data Extraction from Streaming Sources

Data extraction from streaming sources involves tools capable of transmitting data to generate basic reports and perform simple actions like issuing event alerts. These tools utilize complex algorithms for data extraction and stream processing, focusing on extracting data from smaller time intervals [41]. Key features of streaming data extraction include continuous querying and processing of data, logging of individual data or micro-batches consisting of a few records, low transmission latency in seconds or milliseconds, straightforward response functions, and incremental metrics.

3.3.7. IoT Data Extraction

The Internet of Things (IoT) has revolutionized how organizations interact with technology, necessitating efficient tools for managing various sources and types of data to automate collection from different IoT devices [42]. Given that IoT data is generated and transmitted in real-time, there is a crucial need for tools capable of automatically ingesting, processing, integrating, and transmitting this data .

3.3.8. Data Extraction from Social Networks

Tools like web scraping and data mining are employed to extract data from social networks based on predefined criteria [43]. Web scraping involves using bots to access various social network platforms and extract information using an unblocking API that acts as a proxy server. Data that can be captured include profile information, job details, URLs, images, videos, hashtags, geolocation data, timestamps,

comments, marketplace posts, engagement rates, and emerging trends. Data mining, another technique used, employs a combination of statistical, mathematical, and machine learning methods to extract and analyze information. Following data extraction, techniques such as classification, association, pattern tracking, predictive analysis, keyword extraction, sentiment analysis, and trend analysis are applied to derive insights from the extracted data.

3.3.9. Real-time Data Extraction

Real-time data is generated by various technologies, including IoT devices, which must be connected to multiple data sources for effective data capture, [44]. Real-time data capture tools have three primary characteristics: a single data source, a single channel, and minimal resource usage. These tools must also provide:

- High-speed response to a large number of events
- Management of large volumes of data
- Data capture from multiple sources and formats
- Data filtering and aggregation
- Management of a constant, asynchronous, persistent, and connection loss-resistant data transmission channel

3.3.10. Data Extraction from Emails

Unstructured data from the header, body, and attachments of emails is extracted and stored in appropriate databases [45]. Structured data is extracted using bots that validate the data's characteristics. Additionally, some tools analyze email activity to extract metrics such as the number of emails sent per day, emails opened, top senders, recipients, and average response time, which can be used to generate statistics for measuring productivity.

3.4. Challenges in Data Capture

Data capture is a critical process that can face various challenges to ensure the reliability of the data extracted from the source. The most common challenges are:

- **Data Quality:** refers to the accuracy, consistency, and timeliness of the captured data. Manual data entry can introduce errors affecting the accuracy and consistency of information. Data may often be incompletely captured due to a lack of standardized systems or clear procedures [46].
- **Data Integrity:** refers to the accuracy and reliability of data, ensuring it remains unaltered and trustworthy. Data can be intentionally manipulated to conceal fraud or errors. Additionally, data capture processes can be vulnerable to cyber-attacks, compromising the integrity of the information [47].
- **Data Accessibility:** refers to the ease with which data can be retrieved and used by authorized stakeholders. Restricted access policies may limit the availability of data to authorized users [48].
- **Data Security and Privacy:** crucial for protecting sensitive information and ensuring public confidence in the governance system [49].

4. Data Storage

To store large volumes of data in organizations, there are two main alternatives: on-premise storage and cloud storage. Each option has advantages and disadvantages depending on the specific needs of the organization. However, there is a strong trend towards cloud storage due to its flexibility and scalability [50]. A summary of the advantages and disadvantages of both types of storage is shown in Table 1.

The selection of a storage solution can be guided by the following aspects [51]:

- Redundancy: it is crucial to have copies or backups of data to prevent loss in case of failures. Therefore, data should be copied and stored simultaneously in different systems to ensure its persistence.
- Persistence and Preservation: over time, data can become inaccessible, so periodic reviews are necessary to detect and address any data that has been damaged.
- Transformation: as technology evolves, so do the methods for storing data. It is essential to migrate information to new storage systems without altering its semantic content.

Table 1. Advantages and disadvantages of on-premise and cloud storage.

Storage type	Advantages	Disadvantages	Ref.
On-premise	Full control over the hardware and the actions to be performed with the software. No access restrictions. They are reliable as they do not depend on internet connections.	Dependence on providers for hardware and software purchases. Security risk due to internet connection. Low flexibility and scalability. High cost of installation and maintenance.	[52]
Cloud	Fast implementation at start-up. Accelerated time to productive use of applications Lower initial and operating costs. No need for additional infrastructure for servers, networking, among others. No additional IT resources are needed to support both your infrastructure and applications. Cloud providers offer a best-class enterprise infrastructure with the appropriate servers, networks, and storage systems and are responsible for frequent upgrades of your application with each new release; regular backups, and necessary restores and compliance with the latest security and legal compliance requirements. Accessibility from anywhere in the world Backup, which avoids data loss.	Requires internet access and a good, stable, and fast connection. Risk in the security of the data since it can be hacked. Privacy can also be lost. Dependence on third-party hardware and software to operate.	[53–56]

4.1. On-premise Storage

On-premise solution is the name given to the storage systems installed in the organization that is going to use them. They contain all the servers and software necessary to provide the different services. The size and the economic activity of the organization determine the number of servers and the dimensioning of the installations [52]. Studies have shown the drawbacks of storing data at the desktop level. The problems increase with the increased cost of hardware and maintenance that involve high labor costs. Also, there are increased possibilities of data loss due to the low robustness of these systems [57]. However, because of the characteristics mentioned above, this type of storage is no longer being used as much. Gradually it is migrating to cloud storage [55].

4.2. Cloud Storage

Cloud storage systems consist of a series of devices connected by the network. These devices enable storage virtualization, which consists of abstracting physical storage from applications. The network of storage devices is used to access stored information regardless of locations or modes. Based on how the client accesses and interacts with the data, cloud storage can be classified into file, block, and object storage [58,59].

4.2.1. File Storage

Data is organized hierarchically in files, and the storage system stores all file information as metadata, accessible by determining the path to a specific file [59]. Some of the most well-known file storage systems are Hadoop Distributed File System (HDFS) and Google File System (GFS).

- Hadoop Distributed File System (HDFS): HDFS systems [60] can reliably store large amounts of unstructured data on commodity hardware. They facilitate fast data ingestion and bulk processing [61]. A large-scale file is divided into multiple data blocks, which are subsequently replicated across numerous nodes. HDFS divides each file into 128 MB blocks and replicates them with a factor of 3. A typical HDFS configuration includes an active name node and multiple data nodes, adhering to the master-slave architecture [62].
- Google File System (GFS): GFS is designed to organize and manipulate large files for various purposes as required by the user. The design goals of GFS are scalability (an increase in servers), reliability (the system's ability to store data), and availability (the system's ability to retrieve data under varying conditions) [63]. GFS generally comprises three key components: Masters, Clients, and Chunkservers. The master oversees all file system metadata. Chunkservers manage the chunks, which are the storage units in GFS. Clients engage with the master for metadata-related tasks and communicate directly with chunkservers when reading or writing content [64].

4.2.2. Block Storage

In block storage, the file is divided into blocks, and each block is assigned an address. The application can access and combine the blocks using their respective addresses, providing good performance. However, this setup does not ensure secure data transmissions [58]. The storage is raw, and the data is organized as an array of unrelated blocks [65].

4.2.3. Object Storage

Object storage stores virtual containers that encapsulate data, data attributes, metadata, and object identifiers [65]. The object can be of any type and geographically distributed. This setup allows direct and secure access to data by clients via metadata and provides excellent scalability to support Big Data applications [58].

4.3. Database for Storing Government Expenditure Data

Storing public expenditure data in an efficient, secure, and accessible manner is very important. Below are several database options that can be used to store and manage this data.

- Relational Databases (SQL): relational databases are database management systems that use a table-based model to organize data. They are ideal for structured data and complex transactions. Table 2 shows some examples of relational databases.
- NoSQL Databases: NoSQL databases are database management systems designed to handle unstructured or semi-structured data, with a focus on scalability and flexibility. Table 3 shows some examples of NoSQL databases.
- Data Warehouses: data warehouses are storage systems specifically designed for queries and analysis of large volumes of data [66].

Table 2. Relational Databases.

Name	Advantages	Application	Ref.
ySQL	Ease of use. Reliability. Scalability. Performance. High availability	Web applications and medium to large data management systems	[67,68]
PostgreSQL	Open-source. Highly extensible. Robust support for complex transactions and queries.	Applications requiring a high level of integrity and advanced data operations.	[69,70]
Microsoft SQL Server	Integration with other Microsoft products. Enterprise support. Advanced analysis and reporting tools.	Common in enterprise environments using Microsoft products	[71]

Table 3. NoSQL databases.

Name	Advantages	Application	Ref.
MongoDB	Open-source, flexible, dynamic scheme, high scalability	Applications that handle large volumes of unstructured data.	[72]
Cassandra	Open-source, horizontal scalability, high availability, fault tolerance, performance	Distributed applications requiring high availability and management of large volumes of data.	[70,73]

Ensuring data security is crucial for the effective and responsible management of public spending. It is imperative to implement robust security measures to safeguard sensitive information. Encryption practices, stringent access controls, continuous monitoring, regular backups, timely updates, and comprehensive staff training are essential steps for government entities to mitigate risks and protect their data from potential threats.

5. Data Processing

Data processing involves collecting, cleaning, transforming, and analyzing valuable information from diverse sources. Today, data processing has evolved into a specialized discipline, particularly in managing large-scale information or big data. The process typically encompasses three stages: data cleaning, data transformation, and data analysis [74].

5.1. Data Cleansing

Data cleansing, also known as preprocessing, is essential for obtaining a high-quality dataset that supports accurate and meaningful analysis [75]. This stage aims to mitigate the risk of extracting less valuable or erroneous information during subsequent data analysis processes [75,76]. Key techniques involved in data cleansing include:

- Integration: also referred to as data fusion, this process involves consolidating data from multiple sources into a unified and coherent data structure.
- Selection: alternatively termed filtering, selection focuses on identifying and retaining data that is relevant to the specific analysis or application at hand.
- Reduction: this can involve reducing the dimensionality or quantity of data while preserving its essential characteristics and integrity.
- Conversion: data conversion techniques transform data to enhance efficiency in analysis or improve interpretability of results. Common methods include normalization and discretization.

- **Imputation:** addressing missing data is crucial to prevent information loss. Imputation methods include averaging, predictive modeling (e.g., regression, Bayesian inference, decision trees), or other statistical approaches to estimate missing values.
- **Outlier Cleaning:** identifying and handling outliers—data points that significantly deviate from the norm—helps maintain data integrity. Outliers may be corrected if possible or treated as missing values depending on their impact on analysis outcomes.

Each of these techniques plays a vital role in preparing data for analysis, ensuring that the dataset is accurate, complete, and suitable for deriving meaningful insights.

5.2. Data Transformation

It involves converting data into a specific format to align with your requirements, enabling efficient and easy management. This transformation is achieved through the application of rules or data merging. This step encompasses various manipulations, including moving, splitting, translating, merging, sorting, and classifying, among other actions [77].

5.3. Data Analysis

At this stage, data are transformed into useful information that supports decision-making. The collected data are explored to identify trends and patterns that offer valuable insights. Several recognized techniques can be applied [75]:

- **Descriptive statistical analysis:** these involve estimates or values calculated from a data sample that describe or summarize its intrinsic characteristics. Common estimates include mean, median, or standard deviation.
- **Supervised models:** these estimate a function or model from a set of training data to predict outcomes for new, unseen data. Training data sets consist of pairs of objects representing input data and desired results, which can be numerical values (regression) or class labels (classification).
- **Unsupervised models:** these involve adapting a model to given observations without prior knowledge of the data. Unlike supervised learning, there are no predetermined outcomes. Clustering methods are among the most widely used unsupervised learning techniques.

Processing large volumes of information demands dedicated computing resources, influenced by factors such as CPU speed, network capabilities, and storage capacity. Traditional processing systems find themselves outpaced when confronted with substantial data loads. Hence, cloud computing has emerged as an excellent solution for overcoming these limitations [78]. Considering the above information, Table 4 shows some of the most commonly used tools and algorithms in the cloud for data cleansing, Table 5 for data transformation, and Table 6 for data analysis.

Table 4. Tools and algorithms for data cleansing or preprocessing.

Name	Description	Ref.
Apache Spark	Open-source software that enables efficient large-scale data cleansing and transformation, including modules for data management such as: Spark SQL and Data Frames	[74]
Databricks	Cloud-based platform, widely used for large-scale data cleansing and preparation tasks	[79]
Talend	Open-source software that offers a wide range of data cleansing functions such as: validate, standardize, data deduplication, and delimited file cleansing	[80]

Table 5. Tools and algorithms for data transformation.

Name	Description	Ref
Apache Hadoop	Open-source framework for distributed storage and processing. It comprises three main components: Hadoop Distributed File System (HDFS), Hadoop MapReduce, and Hadoop Common System. Through the first two components, it facilitates data transformation. The framework is both scalable and fault-tolerant.	[81,82]
Apache Hive	It operates as a data storage application, presenting information in a tabular format. Tailored for low-latency operations, it facilitates script creation using the Hive Query Language (HQL). Recognized for its scalability and extensibility, this system is specifically crafted for Online Analytical Processing (OLAP) operations.	[83]
Apache Spark	Open source framework widely used in industries for data processing and transformation. It is recognized because data processing is in-memory. Therefore, it supports interactive queries and online streaming.	[74,83]

Table 6. Tools and algorithms for data analysis.

Name	Description	Ref
MapReduce	It is a distributed programming model that enables large-scale parallel data processing in a timely, faultless, and scalable manner. It has two stages: a mapping stage in which data is taken and key-value pairs are created; and a reduce stage that seeks to reduce the data output by joining data according to its key and value.	[81,84,85]
Machine learning algorithms	They are a set of techniques and methods used to learn patterns and perform specific tasks without explicit human intervention. Naive Bayes: estimate the probability that a piece of data belongs to a category. Decision Trees: is a tree-like structure that has leaves, denoting groupings and divisions, which thus establish the conjunctions of salient aspects that motivate those classifications. Random Forest: generates multiple decision trees from random subsets of data. It may have less error compared to other algorithms. Support Vector Machines (SVM): The model labels classes based on certain inputs, and the machine is trained to construct a predictive model for determining the class of newly introduced data. These algorithms undergo automatic improvement based on experience.	[86–88]
Linear Regression	A simple regression is the calculation of the equation corresponding to the line that best describes the relationship between the response and the variable that explains it.	[89]
Clustering Algorithms	These are classification algorithms. KNN (K-Nearest Neighbors): Recognized for its nearest neighbor approach, it operates on the premise that similar data points tend to cluster in space. Agglomerative Clustering: This algorithm organizes data into a tree or dendrogram structure. DBSCAN (Density-Based Spatial Clustering of Applications with Noise): Known for its capability to identify groups of varying densities within datasets.	[90,91]

Some of the most common applications in the area of fiscal control using the above techniques are:

- Pattern identification: which involves the identification of recurrent relationships and structures within the data. Correlation analysis can be used to identify relationships between different expenditure variables [92].
- Anomaly Detection: seeks to identify data that do not fit the expected pattern, which may indicate errors or suspicious activities. Different Machine Learning algorithms such as K-means clustering, Random Forest, Principal Component Analysis (PCA), and Neural Networks can be used for this purpose [93].
- Predictive Analytics: uses statistical and machine learning models to predict future spending behavior based on historical data. It uses linear regression or logistic regression models and Machine Learning algorithms: Random Forest, Gradient Boosting, and neural networks. This could be very useful to predict spending for specific periods based on historical data [94].
- Trend Analysis: involves examining data over time to identify directions and changes in spending behavior [95].

Various tools and platforms can be used to carry out these analyses:

- Statistical Software: R, SAS, SPSS.

- BI (Business Intelligence) tools: Tableau, Power BI, QlikView.
- Programming Languages: Python (with libraries such as pandas, NumPy, scikit-learn, statsmodels), R.

6. Architecture and Interoperability

In environments with large data volumes, establishing an architecture that supports effective interoperability is crucial. This capability is fundamental as it underpins accurate business understanding and informs future decision-making [96]. In today's digital landscape, achieving interoperability poses significant challenges across sectors, particularly in the public sector. Data collection now encompasses diverse types like images, text, audio, and video, spread across disparate databases, incompatible systems, and proprietary software. This fragmentation constraint efficient data exchange, analysis, and interpretation [96].

Recognizing this challenge, various industries have embraced solutions to enhance access, analysis, and communication between systems, devices, and applications, both internally and across organizations. Application Programming Interfaces (APIs) have emerged as a useful solution for managing data flow within internal applications. By enabling secure and straightforward data exchange and functionality sharing, APIs streamline information management processes [96].

Government agencies accumulate vast amounts of structured and unstructured data, offering potential to enhance services and policy-making processes. However, leveraging this data for informed decision-making, especially from large volumes of unstructured text, requires robust interoperability tools and architectures [97].

Interoperability stands out as a critical driver for future advancements across sectors, including the public sector. Among emerging technologies such as IoT, SaaS, and cloud computing, APIs are pivotal for optimizing performance in security, revenue generation, and customer insights [96,98]. APIs provide standardized means to access and exchange data between systems, empowering developers to integrate applications with legacy systems and unlock valuable data resources [96]. By abstracting the underlying technologies and data formats, APIs enable seamless integration across systems, making them indispensable for achieving interoperability goals [99,100].

While APIs solve many of the challenges related to data format and data capture, among others. There are a number of additional challenges that exist when talking about interoperability, which are outlined as follows [101].

- Managing data at scale: implementing interoperable systems requires a coordinated effort to consolidate data from multiple sources [102]. Legacy systems store information in databases with incompatible formats. Extracting, cleansing, transforming and loading massive amounts of data from disparate sources into a common repository requires specialised expertise, time and computing resources [101].
- Addressing privacy concerns: interoperability requires security measures to protect user information. These efforts are complicated when multiple systems exchange data through complex data channels. As a result, organizations will need to rely on appropriate security technologies and policies, which will incur additional costs [101].
- Enforce interoperability standards: organizations use systems that operate in the traditional way, with their own customized protocols and data storage structures. A common industry standard is needed to enable systems to communicate with a high degree of interoperability. Even when interoperability standards are introduced, organizations must upgrade their equipment, software and data infrastructure to enable data exchange between two or more systems [101].
- Cross-platform incompatibility: one of the biggest challenges is platform incompatibility. Each system has its own data format and structure, which can lead to significant inefficiencies when exchanging files. Data often needs to be converted, a process that can be time-consuming and error-prone [101].

- Data loss during transfer: when converting or transferring files between systems, critical information is often lost that can affect the integrity of the project. This can include fine design details that are incompatible between systems, requiring rework or manual adjustments [101].

When structuring systems to capture data from various sources, it is important to address these challenges to prevent future issues, inconsistencies, or data loss that could impact business operations or investigations. In the public sector, promoting interoperability in data management is especially vital for fiscal control and auditing processes, as it ensures that valuable information is not overlooked, potentially aiding in fraud prevention efforts.

7. Data Visualization

This section provides an overview of the fundamental principles and advanced technologies related to data visualization in the modern era. Technological advancements, particularly in information technology, have revolutionized how data is presented compared to traditional print media such as books, newspapers, and magazines [103].

Today, data visualization holds increasing significance as it plays a pivotal role in understanding business processes and facilitating informed decision-making [104]. Consequently, various theories and methodologies have emerged to promote best practices in creating data visualization dashboards that effectively communicate information and enhance user experience [105].

Entering the realm of data visualization necessitates grasping foundational theories, from understanding the workings of the human visual system to selecting appropriate visual metaphors. This involves principles like Gestalt psychology, knowledge of diverse data types, and proficiency in economics, statistics, color theory, graphic design, visual storytelling, and emotional intelligence skills [106,107].

7.1. Basic Concepts for Data Visualization

User Experience (UX): commonly abbreviated as UX, focuses on all aspects related to the interaction between a user and a product or service. This perspective mainly applies to the relationship between people and websites or applications. Many elements come into play in this area, as numerous factors influence how products and services are perceived and their relationship with users [108]. If the design of a system does not align with user needs, it will ultimately lose user acceptance. UX and usability stand out as critical factors in ensuring that a system effectively meets users' needs [108].

User interface (UI): includes the various screens, pages, and visual components, including buttons and forms, through which individuals interact with a digital product or service. The UI is divided into three levels as shown in the Table 7 [109].

Color theory (contrast, colors associated with alerts): in the process of designing a user interface, the psychology of the user emerges as a critical determinant of its effectiveness. In this case color theory takes center stage, exploring which colors shape individuals' perceptions and govern their associations with visual elements [110]. As a result, the selection of a color palette for a user interface or graphic has a significant impact on the decisions made by the user [111]. Figure 4 highlights how colors affect the visibility of certain elements, most notably the birds. Their increased visibility occurs when they are assigned a vibrant color, as opposed to blending into the grayscale tones of their surroundings.

This color influence also extends to alarm systems. To create a clear contrast, a setting with neutral or muted colors, such as various shades of gray, is first established. When the alarm is activated, there is a deliberate color shift to a more attention-grabbing hue. This deliberate change is intended to increase detectability to the human eye and serve as a clear indicator of a significant change. Typically, green indicates a positive change, while red signals an alarm or potential danger [111,112].

Table 7. User interface levels.

Item	Description
Conceptual	Consists of the general idea of application use and operation.
Functional	It focuses on the input language, including what operations are available and how the user gets to use them.
Visual	Determines how the application and data are presented to the user.

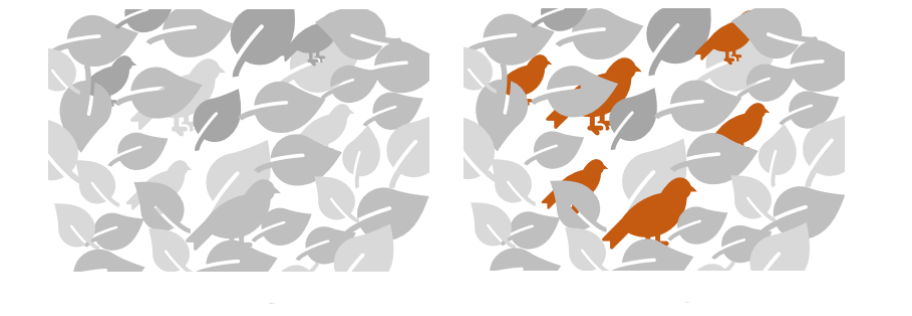


Figure 4. Ease of display and viewing the birds.

7.2. Principles of Data Visualization

According to [113] and [114], there are a number of general principles that apply to a complete presentation or analysis and serve as a guide for constructing new visualizations, as follows.

7.2.1. Structure Planning

Despite its non-technical nature, it is of fundamental importance. Before creating any visual element, it is imperative to prioritize the information to be communicated, visualized and designed. Although this premise may seem obvious, it is essential to focus first on the information and the message, before resorting to software tools that could somehow limit or distort the visual options available [113,114].

7.2.2. Choosing the Right Tool

To achieve effective visual effects, it is critical to master one or more specific software programs. In other words, it is unrealistic to expect complex, technical, and effective results using a simple spreadsheet or other software not designed to create complex and technical graphics [113,114].

7.2.3. Use Effective Geometries to Represent Data

Geometries refer to the shapes and characteristics commonly associated with a particular type of graph; for example, bar geometries are used to create bar graphs. Although it is tempting to go straight from a data set to one of the familiar geometries, it is important to recognize that geometries are visual representations of data in various forms. In most cases, it is necessary to consider more than one geometry to achieve a complete and effective representation of the data [113–115].

7.2.4. Color Always Conveys Meaning

Incorporating color into a visualization can be extremely powerful, and in most cases, there is no reason to avoid using it. A comprehensive study that examined the characteristics that contribute to memorable visualizations found that color visualizations tend to be easier to remember [112–114].

7.2.5. Incorporate Uncertainty

Uncertainty is an inherent part of understanding most systems. Omitting uncertainty from a visualization can lead to misinterpretation. It is important to include uncertainty in visual representations to avoid misrepresentations or misunderstandings related to this aspect. Uncertainty is often overlooked in figures, which results in the omission of part of the statistical message and may affect other facets of the message, such as inferences about the mean [113,114].

7.2.6. Distinction Between Data and Models

The information that is typically presented may be in the form of raw data, such as a scatter plot, or in summary form, such as a box plot, or even in the form of inferential statistics, such as a fitted regression line. Raw data and summary data are usually relatively easy to understand. However, when a model is presented graphically, it is necessary to provide a more detailed explanation so that the reader can reproduce the work in its entirety [113,114]. In a study, it is crucial to present any model comprehensively to ensure its reproducibility. In addition, any visual representation of a model should be accompanied by an explanation in the figure legend or a reference to another section of the paper where the reader can access all necessary details about the graphical representation of the model.

7.2.7. Seek External Opinions

While there are principles and theories about effective data visualization, the reality is that the most effective figures are those that readers can relate to. Therefore, authors of figures are strongly encouraged to seek external reviews of the figures. Often, figures are created quickly during the development of a study and, even if done carefully, are rarely subjected to objective review by people outside the creation process [113,114]. The fundamental goal of following these principles is to convey the message or information as effectively as possible. A better understanding of the design and presentation of data is undoubtedly a critical step in improving the communication of information and reducing misinterpretations.

7.2.8. Graphic Quality Indicators

Table 8 contains some of the most prominent and relevant indicators when designing graphs for data visualization [113,114]. These indicators can be used as a guide to assess whether good data visualization practices are being followed [116].

Table 8. Relevant indicators to data visualization.

Indicators	Description	Ref.
Graphical integrity	This indicator refers to the possibility of distortion or deception in the graphical representation of data. For example, the use of three-dimensional effects in a graph can create the illusion that one piece of data is larger than another when it is not. This indicator addresses any element that may lead to misinterpretation of the data.	[113,115]
Excessive chart elements	Chart junk refers to the presence of visual elements that do not correspond to variations in the data or that make the chart difficult to interpret. For example, the use of textures in bar or line charts, or the inclusion of unnecessary grids in some charts, can create unwanted visual effects.	[113]
Data Ink	This indicator measures the amount of ink used to represent the data compared to the total amount of ink used in the chart. For example, charts with very sharp borders can make it difficult to extract information efficiently.	[113]
Data Density	Data density refers to the ratio of the area occupied by the chart to the amount of data it contains. If a chart displays only four pieces of data, it does not need to take up a lot of space; space can be optimized by making the chart smaller. This makes it easier to understand the information and increases the efficiency of the visualization by making better use of the available space.	[113]

7.3. Public Sector Data Visualization

Governments around the world have worked hard to establish and embed a culture of open government. This approach involves the adoption of principles such as transparency and citizen participation in government decision-making and policy formulation. Open data not only represents a shift in government thinking towards transparency and accountability, but can also increase public engagement by fostering a culture of information sharing and collaboration through accessible data [97].

These cultural changes not only transform the way government operates, but can also generate economic innovation. They also strengthen government performance and ensure greater accountability of elected officials, all with the active participation of citizens [97]. However, a fundamental aspect lies in the way in which the visualization of the data provided is presented, as this influences the correct understanding of the information and consequently the subsequent decision making. The visual presentation of government data must be clear and accurate, following the principles of visualization theory to promote transparency [97,117].

In technical terms, it is essential to have the appropriate tools to carry out an adequate visualization of data for fiscal control, some of which are shown in the Table 9.

Table 9. Tools for data visualization.

Tool	Advantages	Disadvantages	Ref.
Tableau	Easy to use interface, with extensive community support and resources. It also has a variety of customization options.	Expensive, especially for features at the enterprise level.	[118]
Power BI	It has an easy to use interface. Provides drag-and-drop functionality that makes it easy to create compelling visualizations. Includes advanced dash boarding and reporting features that allow users to effectively monitor and analyze data.	Advanced functionality requires knowledge of Power Query and DAX.	[119]
Google Data Studio	It is free and user friendly.	Lack of sophisticated analytics and limited customization.	[120]
Plotly	Interactive, customizable, open-sourced libraries with active communities.	Fewer predefined chart types in comparison to other visualization platforms.	[121]
Grafana	Open source and compatible with a wide range of databases, it provides real-time monitoring.	Some dashboards can be complex to customize.	[122]
IBM Cognos Analytics	It has enterprise-level features and powerful protection and management tools.	Complex reporting requires learning, and large deployments require resources.	[123]

It is important to understand some data visualization tools and recognize their importance in the fields of financial reporting, auditing, and fiscal control. The choice of appropriate data visualization is essential as it can facilitate the identification of irregularities or suspicious patterns that may indicate fraudulent activity. This approach not only strengthens detection capabilities, but also promotes transparency in the public sector.

8. Discussion

The literature review examines essential components of effective data management crucial for enhancing oversight of public expenditure. Beginning with data capture, it underscores the importance of understanding the diverse types of data and the methodologies involved in their effective acquisition. The review distinguishes between traditional on-premise data storage and modern cloud-based solutions, highlighting how cloud databases offer scalability and accessibility advantages that are increasingly favored in contemporary public sector operations.

A key finding of the review is the critical role of interoperability in facilitating the exchange of information between various systems and organizations within the public sector. The integration of modern smart government processes, which involve financial reporting and auditing, can be strengthened through technologies such as the Internet of Things (IoT), Software as a Service (SaaS), cloud computing, and APIs. Additionally, ensuring information is comprehensible is essential, with

data visualization playing a pivotal role. When executed effectively and with appropriate tools, data visualization significantly enhances transparency and accountability in public spending.

Continual advancement in technologies that support the supervision of public spending is essential. This ongoing development aims to yield more accurate and advanced reviews, leveraging automation to reduce constraints on fiscal supervision capacity and optimize resource allocation.

However, it is important to acknowledge the inherent limitations of this review. While synthesizing existing literature and identifying current trends in data capture, storage, processing, interoperability, and visualization for the smart supervision of public expenditure, the findings are contingent upon the quality and scope of the studies included. The depth of analysis may vary based on available research and may not encompass emerging developments or specific contextual nuances in every instance.

9. Conclusion

The findings of this review, regarding data capture, storage, processing, interoperability, and visualization for the smart supervision of public expenditure, underscore the importance of adopting an integrated approach that combines several technologies. Such an approach can significantly enhance oversight of public spending by fostering more efficient, transparent, and accountable management of public resources.

In the public sector, analytics and data management are increasingly recognized as essential for continuous monitoring. Persistent challenges such as corruption, fund mismanagement, and inefficiency have highlighted the need for robust data analysis strategies. Implementing techniques that facilitate intelligent monitoring in the public sector is therefore recommended. Such techniques refer to advanced technologies like machine learning, artificial intelligence, and blockchain.

The integration of advanced technologies presents a transformative opportunity for enhancing financial reporting and auditing in public organizations in the transition from Industry 4.0 to Society 5.0. Such technologies offer unprecedented capabilities in detecting irregularities, optimizing resource management, and improving operational efficiency. Nonetheless, their effective implementation require training and cultural change within public sector institutions.

Author Contributions: Conceptualization, J.A.R.C., C.A.E., J.S.P., and R.E.V.; methodology, J.A.R.C., C.A.E., J.S.P., and R.E.V.; validation, J.A.R.C., J.C.Z., R.M.V., O.M., A.M.H., J.S.P., and R.E.V.; formal analysis, J.A.R.C., M.V., C.Z., C.A.E., J.S.P., and R.E.V.; investigation, J.A.R.C., J.C.Z., R.M.V., M.V., C.Z., O.M., A.M.H., C.A.E., J.S.P., and R.E.V.; writing—original draft preparation, M.V., C.Z., J.S.P., and R.E.V.; writing—review and editing, J.A.R.C., J.S.P., and R.E.V.; supervision, J.A.R.C., J.C.Z., and R.M.V. All authors have read and agreed to the published version of the manuscript.

Funding: This work was developed with the funding of Contraloría General de la República (CGR) and Universidad Nacional de Colombia (UNAL) in the frame of Contract CGR-373-2023, with the support of Universidad Pontificia Bolivariana (UPB).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available on request from the authors.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
API	Application Programming Interface
CGR	Contraloría General de la República (Office of the Comptroller General of the Republic)
DIARI	Directorate of Information, Analysis, and Immediate Reaction of the CGR
DT	Digital Transformation
ETL	Extract, Transform, and Load
GFS	Google File System
HDFS	Hadoop Distributed File System
IoT	Internet of things (IoT)
OCR	Optical Character Recognition
PCA	Principal Component Analysis
SaaS	Software as a Service
SQL	Structured Query Language
UI	User Interface
UX	User Experience

References

1. Lone, K.; Sofi, S.A. Effect of Accountability, Transparency and Supervision on Budget Performance. *Utopía y Praxis Latinoamericana* **2023**, *25*, 130–143.
2. Bharany, S.; Sharma, S.; Khalaf, O.I.; Abdulsahib, G.M.; Al Humaimeedy, A.S.; Aldhyani, T.H.H.; Maashi, M.; Alkahtani, H. A Systematic Survey on Energy-Efficient Techniques in Sustainable Cloud Computing. *Sustainability* **2022**, *14*. doi:10.3390/su14106256.
3. Kubina, M.; Varmus, M.; Kubinova, I. Use of Big Data for Competitive Advantage of Company. *Procedia Economics and Finance* **2015**, *26*, 561–565. doi:10.1016/s2212-5671(15)00955-7.
4. Spiekermann, S.; Novotny, A. A vision for global privacy bridges: Technical and legal measures for international data markets. *Computer Law & Security Review* **2015**, *31*, 181–200. doi:10.1016/j.clsr.2015.01.009.
5. Fang, C. Taxation with Information: Impacts of Customs Data Exchange on Tax Evasion in Pakistan. *Journal Pre-proof*, doi:10.1016/j.ecosys.2024.101243.
6. Marcel, M.; Sangines, M.; Martinez, J.P. La mejora del gasto público a través del presupuesto basado en desempeño, 2010.
7. Alsaadi, A. Financial-tax reporting conformity, tax avoidance and corporate social responsibility. *Journal of Financial Reporting and Accounting* **2020**, *18*, 639–659. doi:10.1108/jfra-10-2019-0133.
8. Petropoulos, T.; Thalassinou, Y.; Liapis, K. Greek Public Sector’s Efficient Resource Allocation: Key Findings and Policy Management. *Journal of Risk and Financial Management* **2024**, *17*. doi:10.3390/jrfm17020060.
9. Gao, S. An Exogenous Risk in Fiscal-Financial Sustainability: Dynamic Stochastic General Equilibrium Analysis of Climate Physical Risk and Adaptation Cost. *Journal of Risk and Financial Management* **2024**, *17*. doi:10.3390/jrfm17060244.
10. Dammak, S.; Jmal Ep Derbel, M. Social responsibility and tax evasion: organised hypocrisy of Tunisian professionals. *Journal of Applied Accounting Research* **2023**, *25*, 325–354. doi:10.1108/jaar-12-2022-0320.
11. Adam, I.; Fazekas, M. Are emerging technologies helping win the fight against corruption? A review of the state of evidence. *Information Economics and Policy* **2021**, *57*, 100950. doi:10.1016/j.infoecopol.2021.100950.
12. Muñoz, L.P.C. La corrupción y la ineficiencia en el gasto público local y su impacto en la pobreza en Colombia, 2021.
13. Valle-Cruz, D.; García-Contreras, R. Towards AI-driven transformation and smart data management: Emerging technological change in the public sector value chain. *Public Policy and Administration* **2023**, *0*, 09520767231188401. doi:10.1177/09520767231188401.
14. Thomas, M.A.; Cipolla, J.; Lambert, B.; Carter, L. Data management maturity assessment of public sector agencies. *Government Information Quarterly* **2019**, *36*, 101401. doi:10.1016/j.giq.2019.101401.

15. Valle-Cruz, D.; Fernandez-Cortez, V.; Gil-Garcia, J.R. From E-budgeting to smart budgeting: Exploring the potential of artificial intelligence in government decision-making for resource allocation. *Government Information Quarterly* **2022**, *39*, 101644. doi:10.1016/j.giq.2021.101644.
16. Oliveira, T.A.; Oliver, M.; Ramalhinho, H. Challenges for Connecting Citizens and Smart Cities: ICT, E-Governance and Blockchain. *Sustainability* **2020**, *12*. doi:10.3390/su12072926.
17. Kankanhalli, A.; Charalabidis, Y.; Mellouli, S. IoT and AI for Smart Government: A Research Agenda. *Government Information Quarterly* **2019**, *36*, 304–309. doi:10.1016/j.giq.2019.02.003.
18. Bendre, M.R.; Thool, V.R. Analytics, challenges and applications in big data environment: a survey. *Journal of Management Analytics* **2016**, *3*, 206–239. doi:10.1080/23270012.2016.1186578.
19. Aftabi, S.Z.; Ahmadi, A.; Farzi, S. Fraud detection in financial statements using data mining and GAN models. *Expert Systems with Applications* **2023**, *227*, 120144. doi:10.1016/j.eswa.2023.120144.
20. Parycek, P.; Schmid, V.; Novak, A.S. Artificial Intelligence (AI) and Automation in Administrative Procedures: Potentials, Limitations, and Framework Conditions. *Journal of the Knowledge Economy* **2023**. doi:10.1007/s13132-023-01433-3.
21. Yahyaoui, F.; Tkouat, M. Partially observable Markov methods in an agent-based simulation: a tax evasion case study. *Procedia Computer Science* **2018**, *127*, 256–263. doi:10.1016/j.procs.2018.01.121.
22. Rukanova, B.; Tan, Y.H.; Slegt, M.; Molenhuis, M.; van Rijnsoever, B.; Migeotte, J.; Labare, M.L.; Plecko, K.; Caglayan, B.; Shorten, G.; van der Meij, O.; Post, S. Identifying the value of data analytics in the context of government supervision: Insights from the customs domain. *Government Information Quarterly* **2021**, *38*, 101496. doi:10.1016/j.giq.2020.101496.
23. Maté Jiménez, C. Big data. Un nuevo paradigma de análisis de datos. *Anales de mecánica y electricidad* **2014**, *91*, 10–16.
24. Muñoz, L.; Mazon, J.N.; Trujillo, J. ETL Process Modeling Conceptual for Data Warehouses: A Systematic Mapping Study. *IEEE Latin America Transactions* **2011**, *9*, 358–363. doi:10.1109/tla.2011.5893784.
25. Cretu, C.; Gheonea, V.; Talaghir, L.; Manolache, G.; Iconomescu, T. Budget - Performance Tool in Public Sector. Proceedings of the 5th WSEAS International Conference on Economy and Management Transformation, 2013.
26. Dawar, K.; Oh, S.C. The role of public procurement policy in driving industrial development. Technical report, United Nations Industrial Development Organization (UNIDO), 2017.
27. Adam, I.; Hernandez-Sanchez, A.; Fazeka, M. Global Public Procurement Open Competition Index. Technical report, Government Transparency Institute, 2021.
28. CABRI. Value for money in public spending. Technical report, CABRI, 2015.
29. Popov, M.P.; Prykhodchenko, L.L.; Lesyk, O.V.; Dulina, O.V.; Holynska, O.V. Audit as an Element of Public Governance. *Studies of Applied Economics* **2021**, *39*, 1–9. doi:10.25115/eea.v39i5.4965.
30. Abdullah, A.S.B.; Zulkifli, N.F.B.; Zamri, N.F.B.; Harun, N.W.B.; Abidin, N.Z.Z. The Benefits of Having Key Performance Indicators (KPI) in Public Sector. *International Journal of Academic Research in Accounting, Finance and Management Sciences* **2022**, *12*, 719–726. doi:10.6007/ijarafms/v12-i3/15064.
31. European Commission. Guidelines for disaggregated data on interventions and beneficiaries. Technical report, European Commission, 2021.
32. Han, Y. The impact of accountability deficit on agency performance: performance-accountability regime. *Public Management Review* **2020**, *22*, 927–948. doi:10.1080/14719037.2019.1679237.
33. Yaqoob, I.; Hashem, I.A.T.; Gani, A.; Mokhtar, S.; Ahmed, E.; Anuar, N.B.; Vasilakos, A.V. Big data: From beginning to future. *International Journal of Information Management* **2016**, *36*, 1231–1247. doi:10.1016/j.ijinfomgt.2016.07.009.
34. Eito-Brun, R. Gestión de contenidos, 2014.
35. Hernandez, A.T.; Vazquez, E.G.; Rincon, C.A.B.; Montero-García, J.; Calderon-Maldonado, A.; Ibarra-Orozco, R. Metodologías para analisispolitico utilizando web scraping. *Research in Computing Science* **2015**, *95*, 113–121.
36. Kumar, N.; Gupta, M.; Sharma, D.; Ofori, I. Technical Job Recommendation System Using APIs and Web Crawling. *Computational Intelligence and Neuroscience* **2022**, *2022*, 1–11. doi:10.1155/2022/7797548.
37. Puñales, E.M.; Salgueiro, A.P. Aplicación de minería de datos a la información recuperada de la intranet para agrupar los resultados relevantes. Jornada Científica ICIMAF-2015, 2015.

38. Güemes, V.L. Business intelligence para la toma de decisiones estratégicas: un caso de aplicación de minería de datos dentro del sector bancario. Master's thesis, Universidad de Cantabria, 2019.
39. Olson, D.L.; Lauhoff, G. *Descriptive Data Mining*; Springer Singapore, 2019. doi:10.1007/978-981-13-7181-3.
40. He, S.; Zhu, J.; He, P.; Lyu, M.R. Experience Report: System Log Analysis for Anomaly Detection. 2016 IEEE 27th International Symposium on Software Reliability Engineering (ISSRE). IEEE, 2016. doi:10.1109/issre.2016.21.
41. Bahri, M.; Bifet, A.; Gama, J.; Gomes, H.M.; Maniu, S. Data stream analysis: Foundations, major tasks and tools. *WIREs Data Mining and Knowledge Discovery* **2021**, *11*, e1405. doi:10.1002/widm.1405.
42. Krishnamurthi, R.; Kumar, A.; Gopinathan, D.; Nayyar, A.; Qureshi, B. An Overview of IoT Sensor Data Processing, Fusion, and Analysis Techniques. *Sensors* **2020**, *20*. doi:10.3390/s20216076.
43. Bazzaz Abkenar, S.; Haghi Kashani, M.; Mahdipour, E.; Jameii, S.M. Big data analytics meets social media: A systematic review of techniques, open issues, and future directions. *Telematics and Informatics* **2021**, *57*, 101517. doi:10.1016/j.tele.2020.101517.
44. Patnaik, S.K.; Babu, C.N.; Bhawe, M. Intelligent and adaptive web data extraction system using convolutional and long short-term memory deep learning networks. *Big Data Mining and Analytics* **2021**, *4*, 279–297. doi:10.26599/BDMA.2021.9020012.
45. Gangavarapu, T.; Jaidhar, C.D.; Chanduka, B. Applicability of machine learning in spam and phishing email filtering: review and approaches. *Artificial Intelligence Review* **2020**, *53*, 5019–5081. doi:10.1007/s10462-020-09814-9.
46. Taleb, I.; Serhani, M.A.; Dssouli, R. Big Data Quality: A Survey. 2018 IEEE International Congress on Big Data (BigData Congress). IEEE, 2018. doi:10.1109/bigdatacongress.2018.00029.
47. World Economic Forum. Data Integrity. Technical report, World Economic Forum, 2020.
48. OECD. *Data accessibility: Open, free and accessible formats*; OECD, 2019. doi:10.1787/8ccf5c38-en.
49. Nikiforova, A., Data Security as a Top Priority in the Digital World: Preserve Data Value by Being Proactive and Thinking Security First. In *Springer Proceedings in Complexity*; Springer International Publishing, 2023; pp. 3–15. doi:10.1007/978-3-031-19560-0_1.
50. Cabello, S. Análisis de los modelos de gobernanza de datos en el sector público: una mirada desde Bogotá, Buenos Aires, Ciudad de México y São Paulo. Documentos de proyectos (lc/ts.2023/71), Comisión Económica para América Latina y el Caribe (CEPAL), 2023.
51. Blumzon, C.F.I.; Pănescu, A.T. Data Storage. In *Good Research Practice in Non-Clinical Pharmacology and Biomedicine*; Springer International Publishing, 2019; pp. 277–297. doi:10.1007/164_2019_288.
52. Perez-Acosta, C.A. Despliegue de aplicación On-Premise en Cloud Computing utilizando servicios de AWS, 2020.
53. Aliaga, L. Disyuntiva: ¿Modelo on-premise o cloud? Technical report, SCL, 2016.
54. Sriramaju, S. A Comprehensive Review on Data Storage. *International Journal of Scientific Research in Science and Technology (IJSRST)* **2019**, *6*, 236–241. doi:10.32628/ijrst.
55. Muñoz-Calderón, P.F.; Zhindón-Mora, M.G. Cloud computing: infrastructure as a service versus the On-Premise model. *Ciencias técnicas y aplicadas* **2020**, *6*, 1535–1549.
56. Antolinez-Diaz, R.O.; Cleves, J.L.R. Almacenamiento en la nube, 2014.
57. Sen, R.; Sharma, A. Optimization of cost: storage over cloud versus on premises storage. 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT). IEEE, 2020. doi:10.1109/csnt48778.2020.9115736.
58. Nachiappan, R.; Javadi, B.; Calheiros, R.N.; Matawie, K.M. Cloud storage reliability for Big Data applications: a state of the art survey. *Journal of Network and Computer Applications* **2017**, *97*, 35–47. doi:10.1016/j.jnca.2017.08.011.
59. Saadoon, M.; Ab-Hamid, S.H.; Sofian, H.; Altarturi, H.H.; Azizul, Z.H.; Nasuha, N. Fault tolerance in big data storage and processing systems: a review on challenges and solutions. *Ain Shams Engineering Journal* **2022**, *13*, 101538. doi:10.1016/j.asej.2021.06.024.
60. Shvachko, K.; Kuang, H.; Radia, S.; Chansler, R. The Hadoop Distributed File System. 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST). IEEE, 2010. doi:10.1109/msst.2010.5496972.
61. Strohbach, M.; Daubert, J.; Ravkin, H.; Lischka, M. Big Data Storage. In *New Horizons for a Data-Driven Economy*; Springer International Publishing, 2016; pp. 119–141. doi:10.1007/978-3-319-21569-3_7.

62. Shin, H.; Lee, K.; Kwon, H.Y. A comparative experimental study of distributed storage engines for big spatial data processing using GeoSpark. *The Journal of Supercomputing* **2021**, *78*, 2556–2579. doi:10.1007/s11227-021-03946-7.
63. Verma, C.; Pandey, R. Comparative Analysis of GFS and HDFS: Technology and Architectural landscape. 2018 10th International Conference on Computational Intelligence and Communication Networks (CICN). IEEE, 2018. doi:10.1109/cicn.2018.8864934.
64. Wang, M.; Li, B.; Zhao, Y.; Pu, G. Formalizing Google File System. 2014 IEEE 20th Pacific Rim International Symposium on Dependable Computing. IEEE, 2014. doi:10.1109/prdc.2014.32.
65. Izquierdo, C. Almacenamiento de Datos en AWS, 2023.
66. Abello, A.; Curto, J.; Gavidí, A.R.; Serra, M.; Samos, J.; Vidal, J. Introducción al Data Warehouse, 2016.
67. Oracle. What is MySQL?, 2024.
68. Oracle. MySQL Documentation, 2024.
69. IBM. ¿Qué es PostgreSQL?, 2024.
70. The PostgreSQL Global Development Group. PostgreSQL 16.3 Documentation, 2024.
71. Microsoft. SQL Server technical documentation, 2024.
72. MongoDB, I. Mongo Documentation, 2024.
73. The Apache Software Foundation. Cassandra's documentation, 2024.
74. Malla-Valdiviezo, R.O.; Gorozabel, O.A.L.; Arévalo-Indio, J.A.; Tóala-Briones, C.H. Mecanismos para el procesamiento de big data. Limpieza, transformación y análisis de Datos, 2023.
75. Subirats-Maté, L.; Pérez-Trenard, D.O.; Calvo-González, M. Introducción a la limpieza y análisis de los datos, 2019.
76. García, S.; Ramírez-Gallego, S.; Luengo, J.; Herrera, F. Big Data: Preprocesamiento y calidad de datos. *Big Data monografía* **2016**, *NA*, 17–23.
77. Arias, W. Big data: extraer, transformar y cargar los datos, 2016.
78. Yang, C.; Huang, Q.; Li, Z.; Liu, K.; Hu, F. Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth* **2016**, *10*, 13–53. doi:10.1080/17538947.2016.1239771.
79. Databricks. Introduction to data preparation in Databricks, 2023.
80. Talend Inc. Data cleansing, 2023.
81. Dolev, S.; Florissi, P.; Gudes, E.; Sharma, S.; Singer, I. A Survey on Geographically Distributed Big-Data Processing Using MapReduce. *IEEE Transactions on Big Data* **2019**, *5*, 60–80. doi:10.1109/tbdata.2017.2723473.
82. Sansores-Cuevas, J.E. Big data: conceptos básicos, tecnologías y aplicaciones, 2020.
83. Sharma, M.; Kaur, J. A comparative study of big data processing: hadoop vs. spark. 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), 2019, pp. 690–701.
84. Fajardo-Orsorio, L. Implementación de la técnica de BigData MapReduce para la caracterización del consumo de energía de clientes residenciales y comerciales., 2021.
85. Hernández-Domínguez, A.; Hernández-Yeja, A. Acerca de la aplicación de MapReduce + Hadoop en el tratamiento de Big Data. *Revista Cubana de Ciencias Informáticas* **2015**, *9*, 49–62.
86. Ariyaluran-Habeeb, R.A.; Nasaruddin, F.; Gani, A.; Hashem, I.A.T.; Ahmed, E.; Imran, M. Real-time big data processing for anomaly detection: a survey. *International Journal of Information Management* **2019**, *45*, 289–307. doi:10.1016/j.ijinfomgt.2018.08.006.
87. Rodríguez-Serrano, A. Clasificación supervisada de documentos mediante el algoritmo Naive Bayes, 2013.
88. Murillo Castañeda, R.A. Implementación del método máquinas de soporte vectorial en bases de datos espaciales para análisis de clasificación supervisada en imágenes de sensores remotos. *Revista Cartográfica* **2021**, *NA*, 27–42. doi:10.35424/rcarto.i102.830.
89. Carrasquilla-Batista, A.; Chacón-Rodríguez, A.; Núñez-Montero, K.; Gómez-Espinoza, O.; Valverde-Cerdas, J.; Guerrero-Barrantes, M. Regresión lineal simple y múltiple: aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal. *Revista Tecnología en Marcha* **2016**, *29*, 33. doi:10.18845/tm.v29i8.2983.
90. Bansal, M.; Goyal, A.; Choudhary, A. A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. *Decision Analytics Journal* **2022**, *3*, 100071. doi:10.1016/j.dajour.2022.100071.

91. Gao, C.X.; Dwyer, D.; Zhu, Y.; Smith, C.L.; Du, L.; Filia, K.M.; Bayer, J.; Menssink, J.M.; Wang, T.; Bergmeir, C.; Wood, S.; Cotton, S.M. An overview of clustering methods with guidelines for application in mental health research. *Psychiatry Research* **2023**, 327. doi:10.1016/j.psychres.2023.115265.
92. Psycharis, Y., Public Spending Patterns. In *Contributions to Economics*; Physica-Verlag HD, 2008; pp. 41–71. doi:10.1007/978-3-7908-2086-7_3.
93. Aliyu, G.; Umar, I.E.; Aghiomesi, I.E.; Onawola, H.J.; Rakshit, S. Anomaly Detection of Budgetary Allocations Using Machine-Learning-Based Techniques. Engineering Innovation for Addressing Societal Challenges. Trans Tech Publications Ltd, 2021, ICASCIE 2020. doi:10.4028/www.scientific.net/ast.107.174.
94. Wolniak, R.; Grebski, W. Functioning of predictive analytics in business. *Scientific Papers of Silesian University of Technology Organization and Management Series* **2023**, 2023, 631–649. doi:10.29119/1641-3466.2023.175.40.
95. Bhikaji, V.; Abdul, S. Trends of public expenditure in India: an empirical analysis. *International Journal of Social Science and Economic Research* **2019**, 4, 3307–3318.
96. Mishra, R.; Kaur, I.; Sahu, S.; Saxena, S.; Malsa, N.; Narwaria, M. Establishing three layer architecture to improve interoperability in Medicare using smart and strategic API led integration. *SoftwareX* **2023**, 22, 101376. doi:10.1016/j.softx.2023.101376.
97. Hagen, L.; Keller, T.E.; Yerden, X.; Luna-Reyes, L.F. Open data visualizations and analytics as tools for policy-making. *Government Information Quarterly* **2019**, 36, 101387. doi:10.1016/j.giq.2019.06.004.
98. Ramadhan, A.N.; Pane, K.N.; Wardhana, K.R.; Suharjito. Blockchain and API Development to Improve Relational Database Integrity and System Interoperability. *Procedia Computer Science* **2023**, 216, 151–160. doi:10.1016/j.procs.2022.12.122.
99. Borgogno, O.; Colangelo, G. Data sharing and interoperability: Fostering innovation and competition through APIs. *Computer Law & Security Review* **2019**, 35, 105314. doi:10.1016/j.clsr.2019.03.008.
100. Platenius-Mohr, M.; Malakuti, S.; Grüner, S.; Schmitt, J.; Goldschmidt, T. File- and API-based interoperability of digital twins by model transformation: An IIoT case study using asset administration shell. *Future Generation Computer Systems* **2020**, 113, 94–105. doi:10.1016/j.future.2020.07.004.
101. AWS. ¿Qué es la interoperabilidad?, 2024.
102. Ramadhan, A.N.; Pane, K.N.; Wardhana, K.R.; Suharjito. Blockchain and API Development to Improve Relational Database Integrity and System Interoperability. *Procedia Computer Science* **2023**, 216, 151–160. doi:10.1016/j.procs.2022.12.122.
103. Chen, J.X. Data Visualization and Virtual Reality. In *Handbook of Statistics*; Elsevier, 2005; pp. 539–563. doi:10.1016/s0169-7161(04)24017-2.
104. Chandra, T.B.; Dwivedi, A.K. Data visualization: existing tools and techniques. In *Advanced Data Mining Tools and Methods for Social Computing*; Elsevier, 2022; pp. 177–217. doi:10.1016/b978-0-32-385708-6.00017-5.
105. Zhang, G. Packaging Big Data Visualization Based on Computational Intelligence Information Design. *Computational Intelligence and Neuroscience* **2022**, 2022, 1–10. doi:10.1155/2022/4558839.
106. Bina, S.; Kaskela, T.; Jones, D.R.; Walden, E.; Graue, W.B. Incorporating evolutionary adaptations into the cognitive fit model for data visualization. *Decision Support Systems* **2023**, 171, 113979. doi:10.1016/j.dss.2023.113979.
107. Ryan, L. Data visualization as a core competency. In *The Visual Imperative*; Elsevier, 2016; pp. 221–242. doi:10.1016/b978-0-12-803844-4.00011-x.
108. Zaki, T.; Islam, M.N. Neurological and physiological measures to evaluate the usability and user-experience (UX) of information systems: A systematic literature review. *Computer Science Review* **2021**, 40, 100375. doi:10.1016/j.cosrev.2021.100375.
109. Lindholm, M.; Sarjakoski, T. Designing a Visualization User Interface. In *Visualization in Modern Cartography*; Elsevier, 1994; pp. 167–184. doi:10.1016/b978-0-08-042415-6.50016-8.
110. Ryan, L. The importance of visual design. In *The Visual Imperative*; Elsevier, 2016; pp. 153–175. doi:10.1016/b978-0-12-803844-4.00008-x.
111. Ware, C. Color. In *Information Visualization*; Elsevier, 2021; pp. 95–141. doi:10.1016/b978-0-12-812875-6.00004-9.
112. Ware, C. Foundations for an Applied Science of Data Visualization. In *Information Visualization*; Elsevier, 2021; pp. 1–29. doi:10.1016/b978-0-12-812875-6.00001-3.
113. Tufte, E.R. Beautiful Evidence, 2006.

114. Midway, S.R. Principles of Effective Data Visualization. *Patterns* **2020**, *1*, 100141. doi:10.1016/j.patter.2020.100141.
115. Ware, C. Images, Narrative, and Gestures for Explanation. In *Information Visualization*; Elsevier, 2021; pp. 331–358. doi:10.1016/b978-0-12-812875-6.00009-8.
116. Zhang, Y.; Cao, T.; Li, S.; Tian, X.; Yuan, L.; Jia, H.; Vasilakos, A.V. Parallel processing systems for big data: a survey. *Proceedings of the IEEE* **2016**, *104*, 2114–2136. doi:10.1109/jproc.2016.2591592.
117. Pazzi, S.; Svetlova, E. NGOs, public accountability, and critical accounting education: Making data speak. *Critical Perspectives on Accounting* **2023**, *92*, 102362. doi:10.1016/j.cpa.2021.102362.
118. Tableau. Tableau, 2024.
119. BI, P. Power BI, 2024.
120. Google. Google Data Studio, 2024.
121. Plotly. Plotly, 2024.
122. Grafana. Grafana, 2024.
123. IBM. IBM Cognos Analytics, 2024.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.