

## Harnessing the Power of Cloud-Based Big Data Analytics for E-Government Advancement in Morocco: A Catalyst for Development



Ahmed Amine Fariz<sup>1</sup>, Jaafar Abouchabaka<sup>2</sup>, Najat Rafalia<sup>3</sup>

LaRIT, Faculty of Sciences, IbnTofail University of Kenitra, Kenitra 14000, Morocco

Corresponding Author Email: [a.amine.fariz@gmail.com](mailto:a.amine.fariz@gmail.com)

<https://doi.org/10.18280/isi.280517>

### ABSTRACT

**Received:** 3 May 2023

**Revised:** 26 July 2023

**Accepted:** 23 August 2023

**Available online:** 31 October 2023

#### Keywords:

*E-government, big data egov, Morocco egov, big data analytics, cloud computing, digital government*

In response to economic, political, and technological stimuli, governments across the globe are progressively embracing digital transformation to devise innovative digital solutions. Despite these advancements, challenges persist in the integration of information resources, including deficiencies in government information systems and threats to network and information security. This paper investigates a novel algorithm for the filling and classification of big data within E-government systems, which comprises data management and governance, cultural and industrial shifts tied to human resource development, and data exchange protocols. A cloud computing environment serves as the infrastructure for constructing an E-government big data intelligence system. The system enables parallel data processing and classification via decision trees, thereby promoting the efficacious and sustainable employment of big data analytics in policy formulation and digital innovation. Additionally, the paper delineates the hurdles and issues that confront these agencies, and proposes potential solutions to augment citizen satisfaction and to deliver value within and beyond governmental sectors. The findings suggest that the integration of big data technologies in E-government presents an effective strategy for the provision of interactive services, thereby addressing citizens' demands for enhanced services.

## 1. INTRODUCTION

The burgeoning field of big data is emerging as a potent avenue for governmental investment. It is postulated that E-governments can harness big data to discern patterns and trends in societal behavior gleaned from social networking sites, thereby refining services and enhancing efficiency. The exploitation of big data is posited to engender novel opportunities for value creation and decision support within the realm of e-government activities.

By employing mobile and social networking data, such as browsing histories, purchasing records, and booking details, governments are equipped to gain insights into the habits and preferences of their citizenry. This, in turn, facilitates the prediction of citizen demands and the tailoring of advertising and programs to meet these needs. Furthermore, big data is instrumental in the creation of smarter, more efficient services for citizens, thereby fostering greater speed, transparency, and efficiency in public sector operations.

Despite these benefits, it is crucial to acknowledge that for e-government to deliver substantial value and instigate a revolution in ICT agencies, the right form of big data is necessary. E-government endeavors must acknowledge the pivotal role of effective big data management, comprehending not only its inherent advantages but also the analytical potential it brings along with the technologies intertwined. This acknowledgment serves as the linchpin in ushering in a new era of possibilities, fostering innovation, elevating productivity, enhancing competitiveness, and elevating the overall quality of services. The fundamental essence of big data revolves around the extraction of predictive insights, the refinement of operational capabilities, and the formulation of

data-driven rules that contribute significantly to fortifying data governance. Therefore, the integration of big data into the realm of e-government necessitates meticulous contemplation and strategic embedding within the broader framework.

## 2. MOROCCAN SMART GOVERNMENT: DIGITAL DEVELOPMENT AGENCY

The Digital Development Agency (ADD) (Figure 1) is a public organization having financial independence and legal personality. This organization works under the administration of the Ministry Delegate to the Head of Government Responsible for Digital Transition and Administrative Reform which handles the state level matters related to digital development strategy, spreading the awareness among citizens related to digital tools and their use. The Digital Development Agency has a lot of goals such as preparing a digital ecosystem, encouraging the appearance of trustworthy players in the digital economy, bringing the users and digital administration closer to each other, and making frameworks for digital services and products.

The responsibility of this institute also includes supporting the Industry 4.0 revolution to reduce the digital divide and leading a change management process in society through training and awareness-raising. ADD is also responsible for encouraging social innovation, motivating development and research, and ensuring continuous computerized involvement. It also involves all the stakeholders in this digital transformation. It is also responsible for making awareness among citizens, businesses, and administration about the use of advanced digital tools.



**Figure 1.** Digital development agency

## 2.1 Responsibilities of the add: Driving digital development

Several missions fall under the responsibility of the ADD, including:

- Ensuring, on behalf of the state, the implementation of the strategy for the development, and encouragement of investment in the field of digital development.
- Giving its opinion on all matters related to digital development that the government refers to it.
- Making any proposal and conducting any study necessary for the implementation of the digital development strategy.
- Developing an annual report on digital development.
- Ensuring monitoring in the field of digital development.
- Proceeding, as part of E-Gov programs, in coordination with relevant agencies, to the establishment of designs related to e-government projects and the development of digital public services, and ensuring their interoperability and integration.
- Establishing technical standards for digital products and services with relevant authorities and organizations, and ensuring their application.
- Contributing to the coherence and convergence of different public orientations and projects in the field of digital technology.
- Contributing to the encouragement, promotion, and support of digital projects and initiatives developed by local authorities.
- Undertaking with relevant organizations any action aimed at regulating, encouraging, and developing companies, especially small and medium-sized enterprises, operating in the field of digital economy.
- Contributing to the promotion and development of initiatives and entrepreneurship in the digital economy sector.
- Providing the necessary expertise to digital economy operators to strengthen their competitiveness.
- Encouraging companies operating in the field of digital economy to take an interest in research and development.
- Promoting the dissemination of digital tools and the development of their use among citizens.
- Ensuring the adequacy of training to meet the needs of actors in the field of digital development.
- Encouraging and promoting applied scientific research in the field of digital development.

The service offering of the Agency is structured around 3 roles:

1. Facilitator of the digital ecosystem
  - Proposal of strategic directions for digital development
  - Launching and implementation of catalytic projects for digital development
  - Provision of shared tools or infrastructures for the

benefit of public administrations to develop their digital projects

2. Accelerator of projects for public administrations
  - Support for the implementation of digital solutions according to the needs of partners
  - Contribution to the realization of structuring projects in the field of digital technology
3. Driver of promotion and awareness around Digital
  - Communication and awareness on a large scale to public administrations, businesses, and citizens on the use of digital technology
  - Promotion of employability in the digital field through training adapted to the needs of digital operators.

## 2.2 Digital transformation of Moroccan administration

The "Smart Government" component concerns the development of digital public services, their interoperability, integration, as well as the implementation of technical standards concerning digital products and services with the relevant authorities and agencies. Its main objective is to improve user experience (citizens and businesses) by providing a repository of services rendered by administrations, using the digital lever as a means to make the administration effective and efficient in serving citizens while:

- Securing data exchanges by controlling and limiting access to shared services without any data storage.
- Ensuring traceability of inter-administration exchanges.

Standardizing the structure of services using an interoperability repository that includes best practices, standards, rules, and recommended norms for exposed services.

The objectives of the projects are as follows:

- Effectively supporting public administration in its digital transition by creating an environment conducive to the development of innovative solutions.
- Accelerating the development of digital products in favor of citizens and businesses, tailored to their needs, quickly deployed, and integrating user feedback in a logic of continuous improvement.
- Stimulating and driving agile digital transformations on a large scale.

## 2.3 The smart government-related ADD project

Some of the projects related to smart government are:

### 2.3.1 Digital Inbox

The Digital Inbox is a digital platform for submitting electronic mail. Who is this platform for? What are its advantages? This platform is for: Citizens; Companies; Associations; Public administrations. Its main advantages are:

- Provide a national online service for submitting and receiving mail to and from administrations and institutions;
- Securely manage and monitor exchanges between users and administrations;

- Notify the sender of the reception, processing and outcome of his request;
- Save time and costs associated with traditional mail.

### 2.3.2 Digital payment

This project aims to dematerialize the payment of fees and fines, and the collection of revenues and taxes. Its objectives are:

- To provide secure online payment methods, such as credit card, bank transfer, and mobile payment;
- To ensure that all transactions are processed in real-time with online tracking of transaction status;
- To improve transparency, traceability, and accountability in financial transactions;
- To provide real-time analytics and reporting on payments, revenue and tax collections.

### 2.3.3 Digital identity

This project aims to establish a secure and trusted digital identity for Moroccan citizens and residents, enabling them to authenticate themselves online, and to access a wide range of e-services offered by public and private organizations. Its objectives are:

- To provide a unique digital identity to every Moroccan citizen and resident;
- To ensure that the digital identity is secure, reliable, and tamper-proof;
- To provide an easy and seamless way to authenticate oneself online;
- To promote the use of e-services and online transactions.

### 2.3.4 Data exchange platform between administrations

This project aims to create a platform that enables the interconnection of the information systems of different public administrations and institutions for the benefit of citizens and businesses. The expected benefits are:

- Making public services more accessible to users
- Ensuring the effectiveness, transparency, and quality of public services rendered to users
- Improving the internal functioning of administrations through fully digitalized exchanges.

### 2.3.5 Digital Factory

This project aims to create a Digital Factory that works in an agile mode, responsible for the rapid digitization of public services through the development of two types of projects:

- Projects to accelerate the digital transformation of public services: by incubating projects driven by administrations, focusing on transferring skills and agile methodology to ADD partners.
- Structural projects for the digital ecosystem: Related to ADD projects and priorities.

Expected benefits:

- Reducing the time-to-market of digital solutions;
- Significantly improving the user experience;
- Creating the possibility of test & learn projects.

### 2.3.6 Digitalization of the investor journey

This project aims to digitize the entire investor journey,

starting with the business creation component.

Expected benefits:

- Attracting foreign investments and facilitating local investor investment;
- Reducing costs for the government by simplifying the process;
- Establishing a relationship of trust between the administration and the investor.

### 2.3.7 Digitalization of the import/export journey

This project aims to digitize the entire import & export journey, allowing for the generation of import/export titles and the execution of the entire customs clearance process.

Expected benefits:

- Online domicile of the import/export title;
- Reducing costs and delays in import/export procedures.

### 2.3.8 Citizen single portal

This project aims to create an evolving multichannel portal that enables the aggregation and dematerialization of administrative services while centralizing existing and future procedures and services for citizens.

Expected benefits:

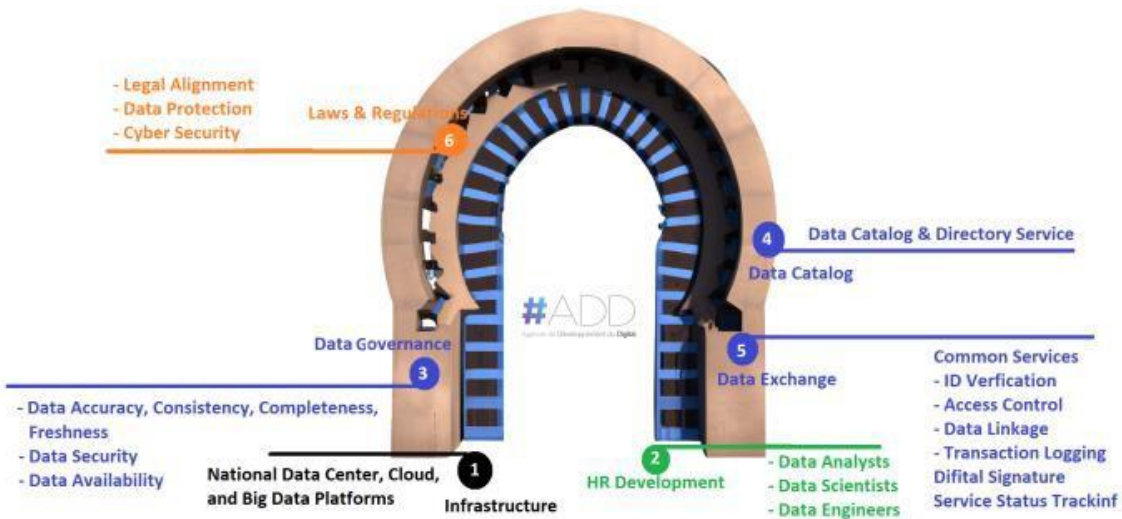
- Centralizing information from all public administrations (ministerial departments, public institutions, local authorities) for citizens;
- Simplifying administrative procedures for citizens;
- Optimizing the user experience.

In summary, the "Smart Government" component of the digital transformation of the Moroccan administration aims to develop digital public services, their interoperability and integration, and to establish technical standards for digital products and services. The main objective is to improve the user experience (citizens and businesses) by offering a repository of services provided by the administrations, using digital tools to make the administration efficient and effective in serving citizens. The projects of the ADD related to this component of the Smart Government include the creation of a platform for data exchange between administrations, the establishment of a Digital Factory for the rapid digitalization of public services, the digitalization of the investor journey, the digitalization of the import/export journey, and the creation of a citizen portal. Other projects include the Digital Inbox for secure submission of electronic mail, the digital payment platform for secure online payments, and the digital identity project for establishing a secure and trusted digital identity for Moroccan citizens and residents.

## 3. THE PROPOSED SOLUTION FOR E-GOV BIG DATA ANALYTICS FRAMEWORK

### 3.1 Government infrastructure

The basic ingredients for a digital government are cloud infrastructure and data centers. Both of them should provide a high level of security and availability. It grants structured data sharing and provides productive computing surroundings. The infrastructure is a hybrid cloud-computing environment, combining on-premises private clouds and public clouds.



**Figure 2.** A Comprehensive look at the framework for big data analytics in digital government

To accommodate different computing needs of government organizations, the framework (Figure 2) offers flexibility in infrastructure solutions, from on-premises data centers to public cloud providers or a hybrid solution managed by the government.

The government infrastructure architecture has four interconnected components: agency cloud, ministry cloud, data linkage and exchange services, and government cloud.

The government infrastructure architecture has four interconnected components.

**Agency cloud:** This framework provides elasticity to agencies in executing their infrastructure. The agency can choose a cloud, a data center, or any compound solution to process their data. So if an agency chooses clouds for its data processing, they will be entirely managed by that agency itself. So they can independently manage their storage requirements and computing requests.

**Ministry cloud:** These clouds are managed by the ministries of government. It is own choice of the ministry to make its own data center or to rent a public cloud. If they build their own physical data center, they have to manage the security and availability of resources on their own. Ministries have to help out their junior agencies. If the ministries choose to use a public cloud, then the data center of that public cloud service provider should locate in the same country for security reasons. A ministry can provide cloud services to junior agencies by using the following ways:

**Colocation:** In this structure, the agencies have to use their hardware to get high security.

**Infrastructure-as-a-Service:** The agencies will get the hardware from ministries.

**Software-as-a-Service:** The ministry will provide the central hardware and software. And all junior agencies will share them.

**Data linkage and Exchange services:** In this component, a data center is used for data sharing. All government organizations exchange their data through this data center. Data is shared between the users through a proper system. It supports public and private channels and provides user access rights. It also keeps track of all transactions.

**Government Cloud:** It is the biggest component that provides cloud services to all agencies and ministries on a national level. This will minimize the cost for all agencies and ministries because they don't have to build their data centers.

And the government cloud will also provide high-level security and availability through a Service Level Agreement (SLA). The government can provide different network connection choices for example VPN, intranet, and a general network. The government can also have two data centers so that the second data center can be used if the first gets some problem.

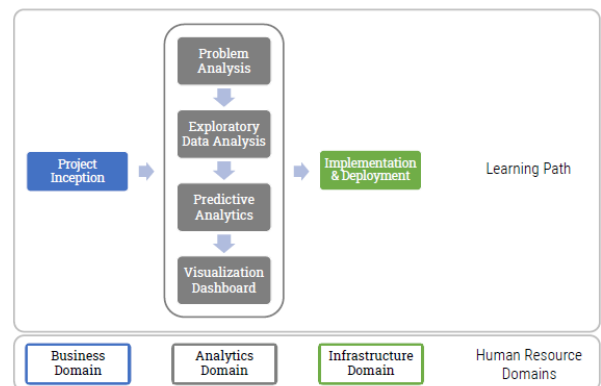
### 3.1 Human resource development

Human resources growth, specifically in the fields of data science, engineering, and analysis, is just as important as infrastructure development. The government should work to get expert manpower in these areas to support the shift towards data-driven decision-making and organization management.

Our proposed solution involves developing human resources in three domains: business, analytics, and infrastructure. To achieve this, we suggest a three-part solution that can be applied simultaneously:

**Short-term project-based training:** Standardized syllabus can be designed if government works by joining hands with academia. This syllabus should emphasize hands-on project-based learning for each of the three domains.

**Government consulting agency:** The government can form an agile consulting agency comprised of experienced data analytics professionals to train government officials and help them implement data-driven technologies and policies.



**Figure 3.** Data analytics learning in short-term training for human resource development



Open government data platform: An open platform that gives knowledge, data resources, and best practices to all collaborators, including citizens can be designed. This platform can increase continuous economic growth through public awareness and usage of big data analytics (Figure 3).

### 3.2 Government data governance

A huge amount of data needs to be processed during digital transformation. This data can be in physical form or digital form. During this procedure, we can achieve high data quality and many other benefits by using data governance protocols. The use of these protocols will provide regulation on data control and data sharing. Figure 4 is depicting the data governance model having four major components.

- **Data Governance Committee Building:** The committee is made up of project managers, executives, users, custodians, and stewards. The people in this committee make and analyze new policies that can clarify data ownership in departments and procedures of data management. They can get IT systems for data management and storage.
- **Data Management Policies:** These policies should be made by taking into account four things: Data life cycle, data privacy and security, data quality assurance, and data exchange. Following is the detail on these aspects.
- **Data life cycle:** Institutes can minimize costs and obey data privacy rules by knowing how data is collected, stored, used, and destroyed in its life cycle. To make the policies, the data must be organized in data sets with additional data (metadata), and saved in an institute-level data catalog for users.
- **Data privacy and Security:** There should be proper categorization of important data. And only authorized users should have access to this data with limited operations to ensure data integrity. A backup copy of data should always be available to stop the loss of data.
- **Data quality assurance:** Data set quality can be calculated by using some metrics such as completeness, accuracy, and consistency. Every involved user in the organization should play his/her role to maintain the quality of data.
- **Data Exchange:** The conditions to use data should be properly defined by the committee. These conditions can include laws, encryption keys responsibility, and encryption standards.

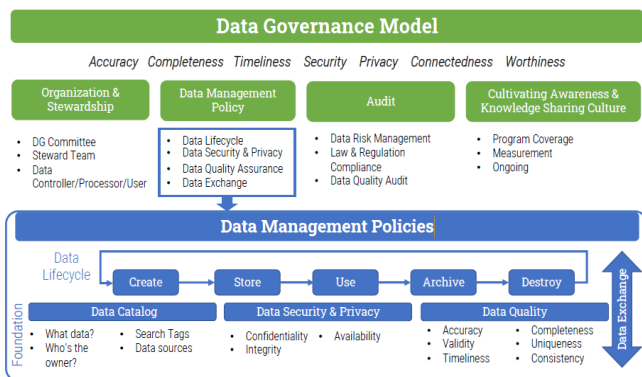


Figure 4. Data governance model

- **Auditing:** Through external and internal evaluations, compliance with risk management, laws, and quality assurance protocols should be observed.
- **Cultivating Awareness and knowledge-sharing culture:** Data governance success is highly dependent on effective communication: So, the agency should organize workshops to raise awareness about the advantages of data governance.

### 3.3 Government data catalog

With the advancement in technology, a huge amount of data is produced by several sources in different formats. So, it has become difficult to find the needed datasets. For the provision of datasets, the government should make a data catalog. This catalog should contain all the data sets with their additional data (metadata), so users can get desired data directly. This will increase the continuous growth of the economy. Also, government institutes will be able to get standardized data-sharing protocols. As depicted in Figure 5, the data catalog contains three parts; a metadata database, a data directory service portal, and a data linkage system.

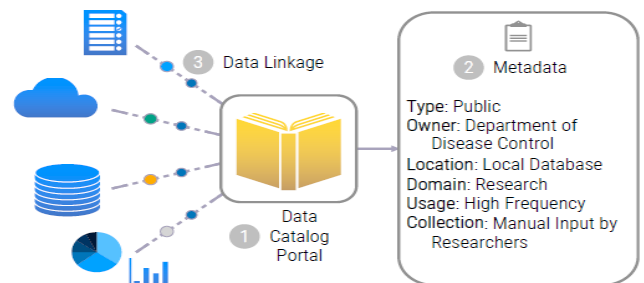


Figure 5. Data catalog architecture

The front of the data catalog is the data directory service portal. This portal allows the user to search the metadata of datasets. Users can apply filters to get their desired data such as tags, data owner names, business segments, etc. The Portal should also secure the sensitive data through an authentication layer. Metadata of the dataset is required for the catalog so that it can work properly; it doesn't need to store actual data. It stores data about tags, owners, collection methods, access request methods, attribute names, data types, and other descriptions.

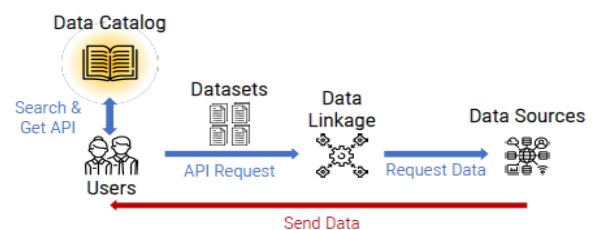


Figure 6. User journey of a data catalog platform

The intermediary component between the database and portal is the data linkage system. Because of this component, users can access actual data by sending API requests. But they should have access rights to send these requests. The system should be linked with the actual data sources. It should have

the ability to preprocess data in API requests, and to provide data management tools to data owners so they can categorize their datasets. Figure 6 shows the complete data linkage process of searching, identifying, and requesting datasets.

### 3.4 Government data exchange

Government-level data need to be shared between data owners and data users. The data exchange platform works behind the data catalog. This platform should be available for both public and private channels. It should only grant access rights when needed. It should be able to perform user authentication, data sharing between institutes, and save records of all transactions. It should be machine-readable without any additional software, and also machine independent. The appropriate formats for data exchange are XML, JSON, and RDF.

For data exchange, there should be assured data quality assurance by the data governance committee. Data quality assurance, authentication features, and data exchange formats are universally accepted, but still, government need to implement rules and opportunities for data exchange with different classification levels.

Figure 7 shows the data exchange architecture. Two types of data exchange platforms are proposed; Government data exchange (GDX) service, and open data sandbox. The first one is for non-public data and the second one is for public data. Both of these platforms coordinate with the government data catalog to provide services to users.

The sharing of government data requires data security such as transaction logging, access control, and user authentication. The dataset with its metadata is required in the data exchange workflow in the GDX platform. Users search datasets and send access requests through the government data catalog. GDX checks access rights and notifies the user who has requested the owner of that dataset. Data transfer options can be SPTF, email, API, and encrypted data in removable storage.

The sharing of public data operates like Kaggle and Sandbox. Where the public not only does data sharing, but also does data analysis, research, and collaboration. Open data sandbox benefits not only citizens but also government officials by increasing access to big data analytics to educational resources. It also increases data literacy.

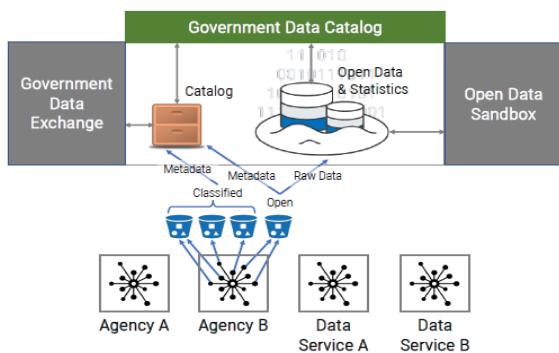


Figure 7. Data exchange architecture

### 3.5 Smart and open government

The complex problems of data analytics can be solved when the government sector (state enterprises, agencies, and ministries) realizes the full capability of storing, analyzing,

and managing the data systematically. It can be possible when government provides services to the public, solve their issues, and gives competition to the private sector. The smart and open government also needs to be cost-effective and transparent.

## 4. GOVERNMENT INFORMATION RESOURCES INTEGRATION SOLUTIONS

The convergence of big data and cloud computing provides fresh chances to integrate government information resources. Issues that can be solved with big data technology include; conflicting information standards and departmental silos. These were the issues because of which the previous E-government era was disturbed.

Morocco’s government management is trying to use big data for getting greater intelligence. They are planning to build a big data platform for the management of their information resources. By closely analyzing the situations and challenges in the integration of information resources, we have some solutions to address these challenges [1].

### 4.1 Principles of integration of government information

According to Figure 8, while integrating government information resources, the government must take an active leadership role in policy guidance, standardization, organization, and overall control to encourage the continued development of government information resources. In the study of Dixon [2]. Record management is integrated into government information resources of the U.S. In this development, the government played an active role. The government needs to keep records for business purposes. They also need the records for future use because they are also answerable to outer entities. So, record management is an essential thing that is needed in government resources. And these types of systems can only be integrated into government systems with the support of government. And the resource should be organized, qualitative, and made with the collaboration of the government. In the study of Lee and Kwak [3] water quality management is integrated into information resources of the U.S. where the government played the role of an active leader to support the project. The states are required to report water quality to the U.S. Environmental protection agency. Several water quality monitoring data sources were integrated by the developers in the management system.

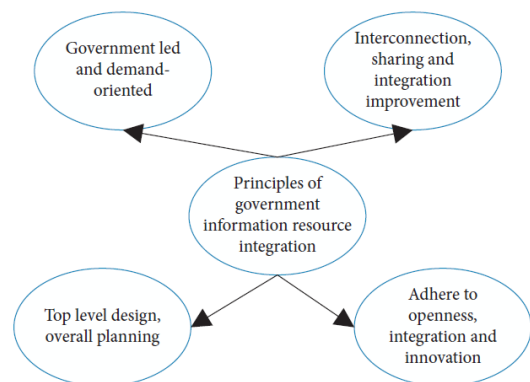


Figure 8. Principles of government information resources integration

These integrated sources should meet the requirements of the government and the general public. The public must be prioritized in this regard to meet their needs. While integrating resources, the most important departments should be government affairs, medical services, and transportation, because they are the most important areas to serve [4]. This will remove the departmental silos and the information will be shared across fields and industries. And the result will increase the efficiency of scientific decision-making and government information utilization.

The full potential of government information cannot be fulfilled by just using it within the government. The use of enterprises can help unlock its full value. And it will restore demand from customers. This will lead to the development of information industries, information technology, and the economy. The information sharing will be fluent [5].

#### 4.2 Developing a big data platform for government: framework and content for information integration

Figure 9 depicts the big data platform for government services. The construction of government affairs is composed of six primary components, which include a standardized specification and an information security system for the guarantee, along with the necessary infrastructure platform construction, application system, and database system.

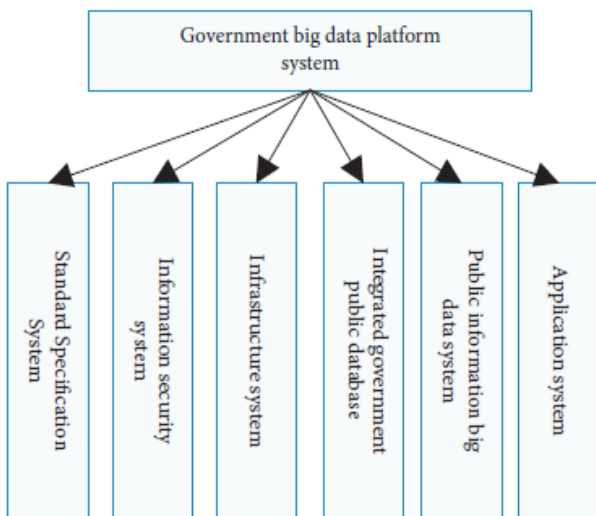


Figure 9. Big data system for government services

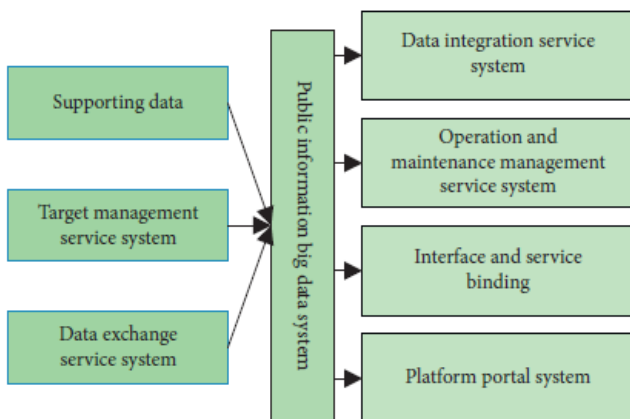


Figure 10. Big data system for public information

In the case of big data systems for public information, the major components are information collection, information integration, and data sharing as shown in Figure 10. Government information resources are integrated at three main levels; using safe application systems for infrastructure, systems, and relevant applications, all of which are developed upon standardization specifications.

### 5. BENEFITS AND OPPORTUNITIES

E-government has maximized its initiatives by investing in ICT and engaging with external and internal stakeholders. Investment in ICT helps to make the services better for people and these projects increase collaboration, efficiency, transparency, and e-participation [6]. E-government services are finding the potential of big data to enhance value, efficiency, and effectiveness. Big data contains people's information and e-government uses this data to make services for the betterment of citizen, E-government can provide certain services at less cost and better quality. Traditional government can be transformed into smart government with improved internal business decisions with the power of big data [2, 7, 8].

Big data benefits in e-government are given below [9, 10]:

- Merging efficient big data resource;
- Including important data in e-government for decision-making;
- Faster data generation;
- Bigger income;
- Increased storage volume;
- Enhanced quality life;
- Managed use of e-government resources;
- Improved transaction processing;
- Better transparency levels.

To achieve the above benefits, high-level resources, tools, and people engagement is needed. And this requires efficient use of big data, effective development, better technology, and effort. Policies should be developed for the sake of accuracy, and data security. For policy employment, big data analytics contains tools and applications for e-governments.

Big data can be a particular asset for e-government that can collect valuable information for citizens, businesses, and governments. In a report proposed by Riedad and Hawkins [5], the role of big data in boosting profit, efficiency, quality, and competition are elaborated.

Big data have improved the services and outcomes of many e-governments. These governments employed big data to improve their services and got remarkable results. U.S. government employed real-time analysis systems from big data to get real-time data from thousands of sources. Then, they employed data.gov for the accountability and clarity of government. The government of Michigan developed a warehouse to provide a single source of information for their citizens [11, 12].

In 2012, European Union used big data to know about the economic potential of the public. U.K government was the earliest to employ big data in improving their services. They created a public website (<http://data.gov.uk>) in 2009 by getting data from seven departments of government. South Korea used big data to join the public and private sectors to serve their citizens in a better way. The Australian government provided access to government data to the general public through a

website (<http://data.gov.au/>) to save their time and resources by providing automated tools [13-15].

## 6. CHALLENGES OF BIG DATA IN E-GOV

New expertise and methods become essential for optimizing the processing of large-scale data analytics and for the effective storage and analysis of data utilizing tools like Hadoop and Spark. As data volumes continue to expand, the demand arises for additional storage systems, novel environments, storage methodologies, and emerging technologies [16].

Efficient procedures are required to extract meaningful value from the big data revolution. However, deploying big data in e-government poses challenges when there's insufficient ICT infrastructure in place because it necessitates diverse processing capabilities and formats [17, 18]. The volume and breadth of data are continually on the rise, surpassing our capacity and ability for real-time modeling and analysis [19].

From a technological perspective, various hurdles emerge, encompassing IT and infrastructure capacities, data security and policy concerns, a shortage of human expertise and skills for big data analysis, limited control over the data, incongruity with existing IT systems, and the rapid growth of big data that outpaces modeling and analysis capabilities. Furthermore, the adoption of data processing tools employing Big Data technology, such as Hadoop and Spark, proves indispensable for effective big data analytics [20].

### 6.1 Technology perspective

New skills and techniques are required for process optimization of big data analytic and the ability to store and analyze data using data processing tools like Hadoop and Spark. As the volume of data increases, it requires additional storage systems, new environments, storage techniques, and new technologies [20].

Efficient processes are needed to derive meaningful added value from the big data revolution. However, applying big data in e-government is challenging without having enough ICT infrastructure because it requires several processing abilities, and formats [21, 22]. The amount and extension of data is increasing day by day surpassing the ability and capability to model and analyze it in real-time [3].

From a technology perspective, several challenges are identified, including the capacity of IT and infrastructure, data security and policy issues, lack of expertise and skills in human for the analysis of big data, less control over big data, less compatibility with current IT systems, and fast growth of big data that outpaces modeling and analysis capabilities. Additionally, using data processing tools that employ Big Data technology, such as Hadoop and Spark, is crucial for effective big data analytics [20].

### 6.2 People perspective

Online service providers collect and save all data that customers enter, browse, and click, providing them with information on who are their customers, their activities, location, and preferences. They can also sell users' data to third parties and advertisers to target advertisements. Consequently, people need education on what data can be

shared and what cannot be shared. Privacy cannot be entirely protected, and third parties managing big data can access all social networking activities.

Unfortunately, many people don't have an understanding about the usage of big data in companies, which poses a challenge from a people perspective [7, 23]. According to researchers, challenges from a people perspective include less human capital development, less learning skills, lack of capabilities, experience, cultural resistance, and trust in technology.

### 6.3 Business process perspective

**Table 1.** Challenges and proposed possible solutions

Challenge	Potential Solutions
Technology Perspective	<ul style="list-style-type: none"> <li>▪ Advanced storage systems should be made and installed to process big data. So that the huge data can be processed efficiently.</li> <li>▪ The capability of telecommunication equipment should be properly used and training should be organized for people to teach analysis skills.</li> <li>▪ Storage capacity should be raised and cloud can store and process big data. Because cloud computing uses new technical ways to save, search, divide, and mine huge amounts of data.</li> <li>▪ Security policies should be enforced regarding data privacy and protection.</li> <li>▪ New security models should be made by following the "Data Security as a Service (DaS)" technique.</li> <li>▪ Efficient big data resources should be offered and merged to ensure effective data processing and utilization.</li> </ul>
People Perspective	<ul style="list-style-type: none"> <li>▪ Government should provide secure technology based on encryption techniques that people can trust. So they can develop their skills upon it and can share their creativity on social networks.</li> <li>▪ Data analysis experts should be hired by the government to enhance human expertise and system performance.</li> <li>▪ For the security of data, people who are experts in security analysis and who have the experience to handle security problems should be included in the data management teams.</li> <li>▪ There should be programs to educate youth about the benefits of big data so they can collaborate with e-government.</li> <li>▪ Collaboration between e-government and big data should be promoted on the national level.</li> <li>▪ Citizens should be empowered to express their creativity and their thinking on social networks.</li> </ul>
Business Process Perspective	<ul style="list-style-type: none"> <li>▪ Bring up public-private partnerships to support big data initiatives.</li> <li>▪ A comprehensive roadmap should be made for implementing big data in e-government.</li> <li>▪ A network should be made that brings together government and community stakeholders to collaborate on big data projects.</li> <li>▪ Effective roadmaps should be made that help build a robust big data ecosystem.</li> <li>▪ Leaders of the community should motivate big data initiatives in e-government.</li> <li>▪ Valuable big data should be strengthened in formulating strategic plans and making decisions.</li> </ul>

Big Data can enhance e-government services by creating valuable insights for public help. However, its implementation



requires government support in terms of research and partnerships. Big Data can improve competitiveness, performance, and decision-making capabilities, but governments must use e-participation to achieve a knowledge economy and enhance their competitive advantage [3, 7, 20].

Investing in Big Data presents complex challenges from a business perspective that need to be addressed to get high-quality results from it. These challenges include changes in business strategy, transformation and management, partnership and collaboration, community and network creation, and leadership roles [24].

Table 1 proposes possible solutions to address the critical challenges of applying Big Data in e-government, categorized according to the aforementioned perspectives [25, 26].

### 7. BIG DATA SYSTEM FOR E-GOVERNMENT IN CLOUD COMPUTING ENVIRONMENT

In the current era of big data and cloud computing, the amount of energy consumed by big data in E-government's data centers is significant. Unfortunately, data loss often occurs due to equipment failures, power outages, and other unstable factors, leading to information gaps and damage, resulting in losses. Traditionally, this problem is controlled by using the method of rough set theory; this theory is boring in process and only can control little amounts of data [4].

To address this challenge, this paper proposes a complete compatibility theory, which extends the compatibility relations theory [27]. According to Figure 11, the management architecture contains three main things: a cluster monitoring module, sensors, and a data center. Cluster monitoring is used to get the raw data set. And then for data categorization and data filling, this raw dataset gives data. The data centers of E-government face the problem of missing data. To solve this problem, an algorithm is proposed in this study. This algorithm is based on the theory of compatibility and completeness.

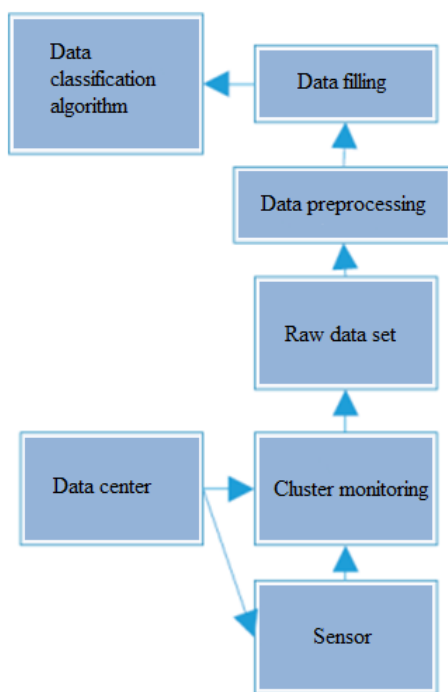


Figure 11. E-government big data system management architecture

The main steps of this algorithm are:

- (1) The attribute values of data are discretized.
- (2) Break the flow of this data.
- (3) Selection of missing data details so that they can be separated.
- (4) Sorting of attribute values for further processing.
- (5) Put on inverted indexing processing.
- (6) Distinguish whether missing data is perfectly compatible or not.
- (7) If the data is perfectly compatible, we applied the rule of minimum value.
- (8) The resulting attribute is then used to fill in the missing data.
- (9) In case the missing data is not perfectly compatible, then the missing attribute is filled with the biggest frequency of the attribute.

In the above process, the Double clustering method is used to break down the data set. And it is classified into small parts according to the divergence of data, and each part contains data with different attributes. Due to the use of the clustering method, cluster data have more similarity between them if there is a smaller and average residual in it. The minimum and average residuals in each cluster are then changed into quadratic shapes. And then the quadratic minima are used for solving missing data values.

The specific algorithm is given below:

The data set is denoted by B, and its related set of expression attributes is denoted by C. The subgroups of B and C are I and J respectively. And  $A_{ij}$  is the data elements in matrix D. And the average residual of I are calculated as given below:

$$Z(I, J) = \frac{1}{|I||J|} \sum_{i \in I, j \in J} a_{ij} (a_{ij} - a_{i.} - a_{.j} + a_{..})$$

where,  $a_i$  and  $a_j$  are the averages of row  $i$  and column  $j$ , respectively, and  $a$  is the overall average.

Consider a given  $m \times n$  matrix  $A$  and let  $\delta$  be a fitted value. Let  $A_{ij}$  be a submatrix of  $A$ , where  $i$  and  $j$  are subsets of the set of row and column indices, respectively. Let  $a_{i.}$  be the mean of row  $i$  of the submatrix,  $a_{.j}$  be the mean of column  $j$  of the submatrix, and  $a_{..}$  be the mean of the submatrix. If the submatrix  $A_{ij}$  satisfies  $Z(I, J) \leq \delta$  and  $0 \leq \delta$ , where  $Z(I, J)$  is the absolute deviation of  $a_{ij}$  and  $a_{i.}$  from  $a_{.j}$ , The data will be more similar in the related submatrix if there is the smaller value of  $\delta$ .

Let's use a bicluster matrix  $S$  that has only one missing data value denoted by  $X$ . And the total number of columns and rows in  $S$  are  $n$  and  $m$ . And the columns of missing data are denoted by  $q$  and rows of missing data are denoted by  $p$ . SUM is used to denote the sum of all values in  $S$  excluding the missing value.  $(1, 2, \dots, p-1, p+1, \dots, n)$  are the values of  $P$ , and  $(1, 2, \dots, q-1, q+1, \dots, n)$  are values of  $q$ . And the average residual is calculated as given below:

$$Z(m, n) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n Z_{ij}$$

where,

$$Z_{ij} = (a_{ij} - a_{i.} - a_{.j} + a_{..})^2, a_{i.} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n a_{ij} = \frac{1}{mn} (x + SUM)$$

The average of data in  $S$  is represented by  $A_p$  and  $B_p$ , where

$B_p$  is for column  $j$  and  $A_p$  is for row  $i$ . And the average residual of data clusters is calculated as follows:

$$\begin{cases} A_p = \overline{A_p} + \frac{x}{n} \\ B_p = \overline{B_p} + \frac{x}{m} \end{cases}$$

And then we got the following expression by using the above equation where  $j$  represents  $q$  and  $i$  represent  $p$ .

$$Z_{ij} = \left( x - \overline{A_p} - \frac{x}{n} - \overline{B_p} - \frac{x}{m} \right) + \frac{(x + SUM)^2}{mn}$$

The quadratic function for the missing data  $x$  can be calculated using the following formula:

$$Z_{ij} = c_{ij2}x^2 + c_{ij1}x + c_{ij0}$$

Here,  $c_{ij2}$ ,  $c_{ij1}$ , and  $c_{ij0}$  are constants. Now we need to calculate the minimum value. And it is calculated by examining the attributes of minimum value and the characteristics of the quadratic function. A submatrix has more similarity between the data in it if the value of  $Z(m, n)$  is smaller. If  $Z(m, n)$  is minimum, the formula to find missing data is as follows:

$$x = \frac{1}{(m-1)(n-1)} \sum_{i=U} \sum_{j=V} a_{ij}$$

By using the above-described method, Missing data can be filled very efficiently. The efficiency of this algorithm can be further improved by the optimization of this algorithm.

## 8. EXPERIMENT

The above section describes the proposed method and its specific steps with equations. To test the accuracy of this proposed algorithm, experimental analysis is performed on it. The whole experiment process is described in the following sections.

### 8.1 Experimental setup

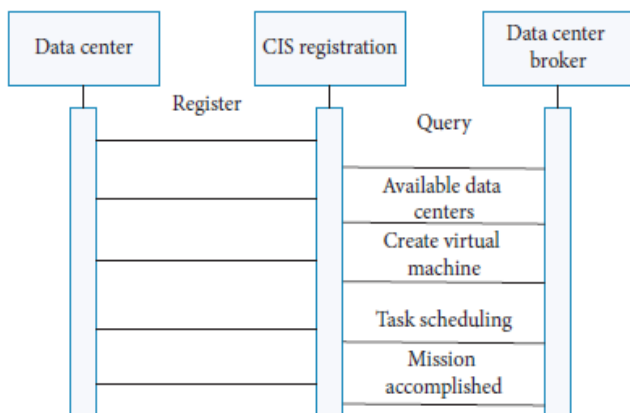


Figure 12. Simulation flow

Table 2. UCI data set contents

Serial Number	Data Set	Number of Sample	Attribute	Catalogue
1	Rabat	2,023	6	4
2	Casa	638	6	3
3	Fes	5,648	6	3
4	Tanger	2,033	7	5
5	Dakhla	720	4	4

Table 3. PC configurations in HAD platform

Component	High-Performance PC	Ordinary PC
CPU processor	I7-2620M	I7-2260M
Hard disk (GB)	1,024	512
Memory (GB)	16	8
Operating system	WIN10	WIN10

Five datasets are used to conduct the experiment: Rabat, Casa, Fes, Tanger, and Dakhla which are obtained from the UCI machine-learning database. They were stored in ARFF (Attribute-Relation File Format) for system testing [28].

Table 2 presents basic information on the UCI datasets utilized. And Table 3 contains the PC specifications used in the experiment. In this experiment, distributed cloud computing was applied to the HAD platform. For testing the system, the CloudSim simulator was our basic platform [29, 30]. The specifications of this simulator regarding its initialization and installation are depicted in Figure 12.

### 8.2 Experimental indicators

Two types of metrics are used in this study: fill accuracy and classification accuracy.

#### 8.2.1 Fill accuracy

As we know that the missing data that is being processed in this study are diverse in nature. So, different matches of approaches are needed to get the filling accuracy. The true value of the filled data is deemed to be identical to the value before replacement, and correctness is determined if the true value is equivalent to the filled value.

The calculation formula is given below:

$$P = \frac{\left( t_i + \sum_{j \in N} ang_j \left( |u_i - u| - \gamma \sqrt{S_j} \right) \right)}{n_i}$$

where,

$t_i$  is the period required for filling,  $a$  donates the total number of fills,  $g_j$  represents the strength factor for detecting missing data,  $u$  represents the complete data size of the system before filling,  $u$  represents the complete data size of the system after filling,  $\lambda$  represents the variance,  $S_j$  represents the margin of error between the filled and real values,  $N$  is the set of E-government data,  $n_i$  represents the size of the data to be filled.

#### 8.2.2 Classification accuracy

This metric is considered very efficient in classification algorithms. The formula to find the classification accuracy is given below:

$$L = \frac{\sum_{i=1} b_i}{N_a}$$

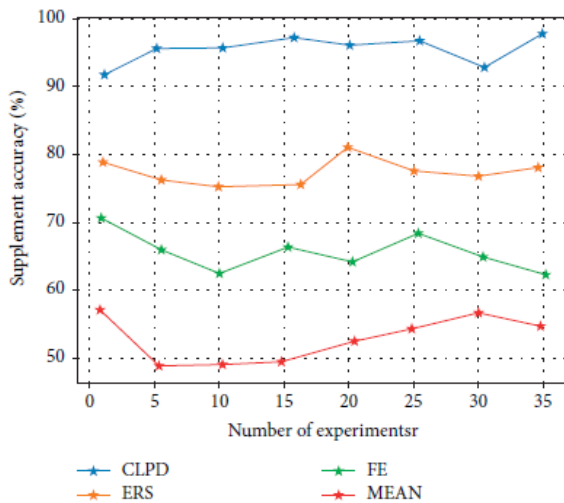
Here, 'b' refers to the number of classifications that were exact matches with the target, and 'N<sub>a</sub>' represents the total target classifications.

## 9. COMPARISON WITH PREVIOUS METHODS

In the realm of data analysis and handling, various conventional techniques have historically been employed to address the issue of missing values within datasets. Among these methods, one can mention the utilization of rough set theory, the Mean method, the FE method (which involves Discrete Random Forest), and the ERS method (an acronym for Weakly Correlated Random Forest method). These approaches have served as valuable tools in the quest to fill in the gaps in datasets.

However, in the context of this particular study, a pioneering solution emerges in the form of a cloud computing-based system known as CLPD, specifically designed to address the challenges presented by big data in the domain of E-government. This innovative system represents a significant departure from traditional methods and offers a fresh perspective on handling large-scale data.

To gauge the effectiveness and performance of the proposed CLPD algorithm, a comprehensive comparison was conducted with the aforementioned conventional techniques, as documented in Figure 13. The results of this comparative analysis reveal a compelling narrative. Notably, the proposed CLPD algorithm outshines its predecessors, boasting an impressive accuracy rate of 96%. This outcome is a testament to the algorithm's ability to leverage the complete information within the dataset, resulting in a substantial enhancement of accuracy.



**Figure 13.** Comparison of the accuracy between different methods for filling in missing data sets

In essence, this study demonstrates the pivotal role that CLPD, the cloud computing-based system, can play in advancing data analysis within the E-government sector. It not only surpasses traditional methodologies in terms of accuracy but also highlights the potential for innovative solutions to revolutionize the way we approach and address data challenges in the era of big data.

The proposed algorithm is far better in accuracy and reasonableness than the rough set theory mean method, ERS

method, and FE method. It is because the predicted algorithm has taken into account the wide amount of information, and data dimension, and an efficient decision strategy is used in processing. Also, the proposed algorithm has more speed and quality as compared to previous methods.

## 10. CONCLUSIONS

Big data is becoming an important investment field for every government. E-governments are using big data to know about the behaviors of their citizens on different platforms to provide them with better services. This data can be used for decision-making. Big data can improve efficiency, revenue, standard, and competence. So it creates huge benefits for citizens. But at the same time, governments face a lot of challenges in implementing big data in e-governments. In this study, these challenges are categorized into three perspectives: Knowledge, people, and business. And then a detailed description of these challenges and their solutions is provided. Previously, big data is applied to many projects of e-government, such as disaster management, merging vehicles and smart device data into e-government systems to save time, fuel, and cost for companies, etc. This study is providing a thorough description of the Moroccan Digital Development Agency. A solution framework for e-government big data analytics is presented. And then an algorithm is proposed for filling missing attribute values in E-government data centers to ensure a complete dataset. This algorithm got remarkable results and achieved a higher accuracy of 96%. This algorithm is compared with some state-of-the-art methods, and it is far better in terms of speed, quality, and accuracy. So, this algorithm has to ability to enhance the processing capability of e-government data centers in terms of big data.

## REFERENCES

- [1] Gandomi, A., Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2): 137-144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- [2] Dixon, B.E. (2010). Towards E-government 2.0: An assessment of where e-government 2.0 is and where it is headed. *Public Administration and Management*, (2): 418. <https://hdl.handle.net/1805/4334>.
- [3] Lee, G., Kwak, Y.H. (2012). An open government maturity model for social media-based public engagement. *Government Information Quarterly*, 29(4): 492-503. <https://doi.org/10.1016/j.giq.2012.06.001>
- [4] Al Nuaimi, E., Al Neyadi, H., Mohamed, N., Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, 6(1): 1-15. <https://doi.org/10.1186/s13174-015-0041-5>
- [5] Piedad, F., Hawkins, M. (2001). *High availability: Design, techniques, and processes*. Prentice Hall Professional.
- [6] Bertot, J.C., Choi, H. (2013). Big data and e-government: Issues, policies, and recommendations. In *Proceedings of the 14th Annual International Conference on Digital Government Research*, pp. 1-10. <https://doi.org/10.1145/2479724.2479730>
- [7] Anshari, M., Lim, S.A. (2017). E-government with big data enabled through smartphone for public services:

- Possibilities and challenges. *International Journal of Public Administration*, 40(13): 1143-1158. <https://doi.org/10.1080/01900692.2016.1242619>
- [8] Hopwood, P. (2008). Data governance: One size does not fit all. *Information Management*, 18(6): 16.
- [9] Schweizerische, S.N.V. (2013). Information technology-Security techniques-Information security management systems-Requirements. ISO/IEC International Standards Organization.
- [10] Data, S.S. (2014). Metadata exchange. SDMX Standards Version.
- [11] Pavlichev, A., Garson, G.D. (2004). Digital government: principles and best practices. Igi Global.
- [12] Initiative, D.C.M. (2006). Dublin core metadata element set, version 1.1. <https://hdl.handle.net/10421/3401>.
- [13] Joseph, R.C., Johnson, N.A. (2013). Big data and transformational government. *It Professional*, 15(6): 43-48. <https://doi.org/10.1109/MITP.2013.61>
- [14] Wende, K. (2007). A model for data governance—Organising accountabilities for data quality management. In *ACIS 2007 Proceedings*.
- [15] Khatri, V., Brown, C.V. (2010). Designing data governance. *Communications of the ACM*, 53(1): 148-152. <https://doi.org/10.1145/1629175.1629210>
- [16] Janowski, T. (2015). Digital government evolution: From transformation to contextualization. *Government Information Quarterly*, 32(3): 221-236. <https://doi.org/10.1016/j.giq.2015.07.001>
- [17] Kshetri, N. (2014). The emerging role of Big Data in key development issues: Opportunities, challenges, and concerns. *Big Data & Society*, 1(2): 2053951714564227. <https://doi.org/10.1177/2053951714564227>
- [18] Fariz, A.A., Abouchabka, J., Rafalia, N. (2020). Improving MapReduce Process by Mobile Agents. In *Software Engineering Perspectives in Intelligent Systems: Proceedings of 4th Computational Methods in Systems and Software 2020*, Springer, Cham, pp. 851-863. [https://doi.org/10.1007/978-3-030-63319-6\\_79](https://doi.org/10.1007/978-3-030-63319-6_79)
- [19] Zainal, N.Z., Hussin, H., Nazri, M.N.M. (2016). Big Data initiatives by governments--issues and challenges: A review. In *2016 6th International Conference on Information and Communication Technology for the Muslim World (ICT4M) Jakarta, Indonesia*, pp. 304-309. <https://doi.org/10.1109/ICT4M.2016.068>
- [20] Olshannikova, E., Olsson, T., Huhtamäki, J., Kärkkäinen, H. (2017). Conceptualizing big social data. *Journal of Big Data*, 4(1): 1-19. <https://doi.org/10.1186/s40537-017-0063-x>
- [21] West, D.M. (2005). *Digital government: Technology and public sector performance*. Princeton University Press.
- [22] Melitski, J., Holzer, M., Kim, S.T., Kim, C.G., Rho, S.Y. (2005). Digital government worldwide: A E-government assessment of municipal web sites. *International Journal of Electronic Government Research (IJEGR)*, 1(1): 1-18. <https://doi.org/10.4018/jegr.2005010101>
- [23] Boyd, D., Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5): 662-679. <https://doi.org/10.1080/1369118X.2012.678878>
- [24] Gudivada, V.N., Baeza-Yates, R., Raghavan, V.V. (2015). Big Data: Promises and problems. *Computer*, 48(3): 20-23.
- [25] Kache, F., Seuring, S. (2017). Challenges and opportunities of digital information at the intersection of Big Data analytics and supply chain management. *International Journal of Operations & Production Management*, 37(1): 10-36. <https://doi.org/10.1108/IJOPM-02-2015-0078>
- [26] Sivarajah, U., Kamal, M.M., Irani, Z., Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70: 263-286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- [27] Glossary, G.I. (2014). Answering big data's 10 biggest vision and strategy questions. <https://www.gartner.com/doc/2822220?refval=&pcp=mpe#a-1319868613>.
- [28] Fariz, A., Abouchabaka, J., Rafalia, N. (2015). Using multi-agents systems in distributed data mining: A survey. *Journal of Theoretical & Applied Information Technology*, 73(3): 427-440.
- [29] Morabito, V., Morabito, V. (2015). Big data and analytics for government innovation. *Big Data and Analytics: Strategic and Organizational Impacts*, 23-45. [https://doi.org/10.1007/978-3-319-10665-6\\_2](https://doi.org/10.1007/978-3-319-10665-6_2)
- [30] Al-Shboul, M., Rababah, O., Ghnemat, R., Al-Saqqa, S. (2014). Challenges and factors affecting the implementation of e-government in Jordan. *Journal of Software Engineering and Applications*, 7(13): 1111. <https://doi.org/10.4236/jsea.2014.713098>