(IJTBM) 2024, Vol. No. 14, Issue No. II, Apr-Jun **e-ISSN: 2231-6868 p-ISSN: 2454-468X**

Developing A Smart Big Data Analytics Framework To Enhance The Efficacy Of Digital Government

Tejas Thakral
Vivekananda Institute of Professional Studies
Delhi

DOI:10.37648/ijtbm.v14i02.006

[1]*Received: 09 April 2024; Accepted: 06 June 2024; Published: 19 June 2024*

ABSTRACT

In response to challenges from the social, political, and technological spheres, governments worldwide are finding that finding creative digital solutions is a major challenge associated with digital transformation. In this paper, we present a framework that leads to the efficient and long-lasting use of big data analytics in policymaking and the development of digital innovations. The framework standardizes infrastructure architecture, cultural and industrial change with human resource development, data management and data governance, and data exchange protocols in a trusted and regulated environment. We also go over particular modifications made to the framework inside the framework of the Thai government and the adoption process, offering concrete proof of the framework's usefulness.

INTRODUCTION

Governments everywhere are realizing more and more that in order to respond to technological, political, or socioeconomic challenges, they must constantly innovate and change their digital solutions [1]. The data and data analytics that go into developing digital solutions are just as vital as the actual digital solutions. A government going through a digital transformation process can become a Smart and Open Government by using big data technologies methodically and creatively. This would allow the government to manage resources transparently and effectively, while also offering competitive advantages to its industries and effective public services.

In order for a government to successfully undergo a digital transformation, we provide the Big Data Analytics Framework in this article. Seven functional components make up the framework: laws and regulations, data governance, data exchange, data catalog, data governance, infrastructure, and smart and open governance (Fig. 1). Some of these components can be developed concurrently. Despite the fact that the framework was created especially for and accepted by the Thai government, we will attempt to generalize the models and concepts while providing evidence of the framework's applicability by outlining the government's acceptance process. The structure seeks to:

establish a framework for the digital transformation process, standardize infrastructure architecture, manage and govern data, develop human resources, manage and exchange data among government agencies in a safe and controlled environment, and pave the way for the efficient use of big data analytics in the formulation of public policy and the development of new digital innovations.

OVERVIEW

Similarities between the terms "smart" and "open government" and more well-studied terms like "digital government" and "e-government" may be found [1-4]. According to some research, creating a digital government can be seen as a multi-stage evolutionary process [1], [2]. A number of the phases that are discussed in these books bear similarities to the elements that make up our framework: infrastructure, cultural shifts and the establishment of trust through sound

[1] ***How to cite the article:*** Thakral T. (June 2024); Developing A Smart Big Data Analytics Framework To Enhance The Efficacy Of Digital Government; *International Journal of Transformations in Business Management,* Vol 14, Issue 2, 51-60, DOI: http://doi.org/10.37648/ijtbm.v14i02.006

governance, data exchange and sharing for information sharing and cooperation, and data-driven tailored public services.

A smart government's essential element is data governance, for which numerous models have already been put out, including in [5], [6]. In this study, we offer a data governance model that we have tailored for our framework using lessons learned from previous successful models.

This framework's data catalog concept entails creating metadata fields to go along with datasets that are kept up to date by government agencies. Although a detailed discussion of the metadata list is outside the purview of this study, we drew inspiration for our design from a number of established standards, including the Dublin Core Metadata Initiative [7] and the Statistical Data and Metadata Exchange (SDMX) [8].

ANALYTICS WITH BIG DATA SCHEMA FOR THE INTERNET OF GOVERNMENT

A. Infrastructure of the Government

The most essential component of a digital government is its data center and cloud infrastructure. High levels of availability (at least 99.9% [9]), network security, and facility security (at least ISO27001:2013 [10]) are requirements for a strong data center and cloud infrastructure. It offers a productive computer environment and permits methodical data interchanges.

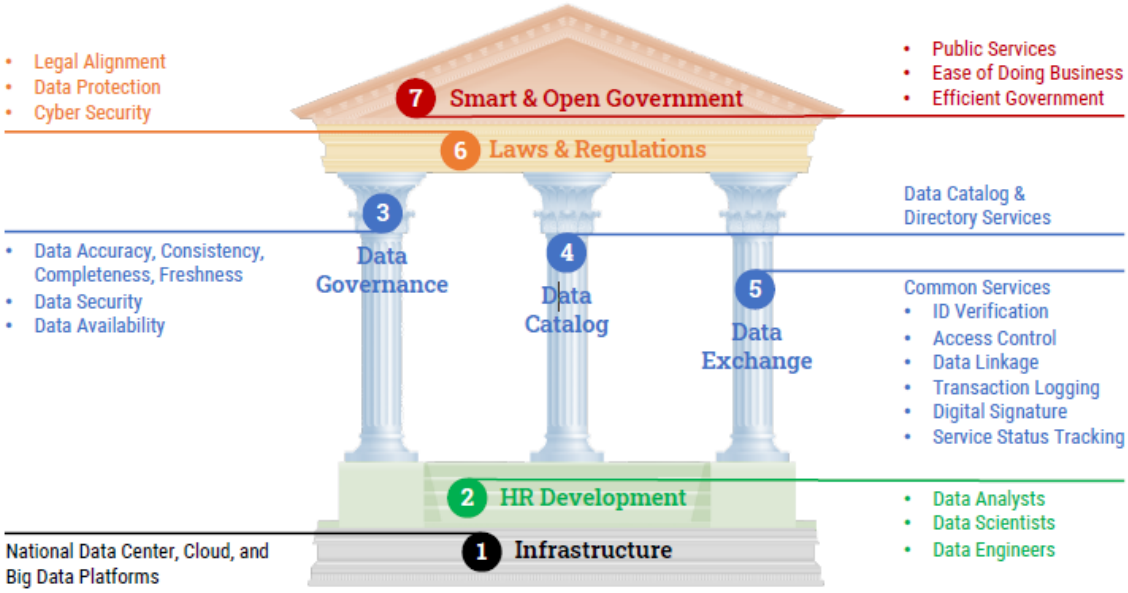

Fig 1 An overview of the Big Data Analytics framework designed for digital government.

physically represented by a distributed, hybrid cloud computing architecture that combines private and public clouds on-site, all operating under a common infrastructure.

Our framework allows flexibility in how each agency or ministry can implement its infrastructure solution, from an on-premises data center to a public cloud provider (where it is required to store and process public data only), to a hybrid solution, to a central solution provided by the government. This is because government organizations have different computing needs based on their missions and data classification levels.

The general architecture of government infrastructure is composed of four interrelated parts that work together.

1) Agency Cloud: To offer the greatest flexibility in modifying processing demands or storage needs, agency-level clouds are managed solely by an agency.

2) Ministry Cloud: Government ministries are in charge of managing ministry-level clouds. Any cloud infrastructure can be put into operation by either developing a physical data center that meets the aforementioned availability and security requirements, or by renting a public cloud and renting out services and infrastructure to lower-level organizations. As far as that government is concerned, it may require the public cloud service provider's data center to be physically situated inside its borders for security-related concerns. From high customizability to high convenience, a ministry can offer cloud services to subordinate agencies in three primary ways:

Colocation is the practice of a subordinate agency bringing its own hardware to the ministry's physical data center in order to benefit from the high facility security standard.

Subordinate agencies can share the same central software and hardware supplied by the ministry under the terms of Infrastructure-as-a-Service and Software-as-a-Service, respectively. The ministry supplies virtual machines (VM) to subordinate agencies.

3) Data Linkage and Exchange Services: These services are provided by at least one data center, which serves as a physical hub for data interchange between government agencies. In sections III-E, we go over the specifics of the data sharing system.

4) Government Cloud: Under a Service Level Agreement (SLA), Government Cloud delivers high availability and security standards-compliant, national-level cloud services to ministries or agencies that would like to lease a data center rather than construct their own. Network connection options such as an intranet, public, and virtual private network (VPN) should be offered by the Government Cloud. In order to ensure high availability in the event of a single data center outage, it should additionally install at least two physical data centers.

In Thailand, CAT Telecom Public Company Limited, a state corporation overseen by the Ministry of Digital Economy and Society, is responsible for the construction and upkeep of the Government Data Center and Cloud Service (GDCC).As the Government Cloud previously mentioned, GDCC plans to launch the service in September 2020.

B. Development of Human Resources

The development of human resources is just as crucial as infrastructure, and the government needs to make plans in advance to establish a highly skilled workforce of data scientists, engineers, and analysts. A trained workforce is essential to changing the work culture of governments and corporations to one that emphasizes value creation from data, data-driven policymaking and organization management, and more Public-Private Partnerships (PPP).

Three different human resource development sectors are taken into consideration by our framework:

Business domain, comprising managers, directors, and executives who choose the organization's course covers the data analytics learning path and identifies specific problems that can be resolved with data analytics, data scientists, analysts, and visualizers working in the data mining and data visualization fields comprise the analytics domain.

Infrastructure professionals involved in the creation and upkeep of big data infrastructure include data engineers, data architects, and IT security operators.

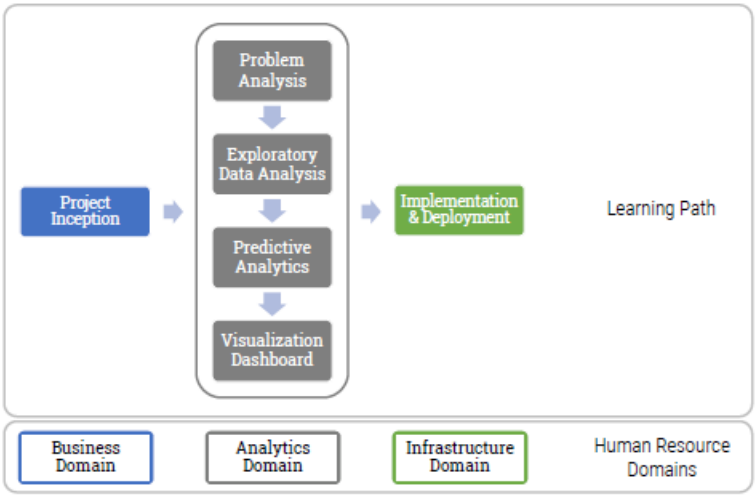

Fig. 2. A short-term training program for human resource development

We offer a three-part approach that can be applied concurrently to enhance human resources in these areas:

brief project-based instruction. The government can create standardized curricula with an emphasis on practical, project-based learning by partnering with academic institutions in the infrastructure, analytics, and business areas. With this method, government employees study in an environment analogous to a sandbox, going through the whole data analytics process from project conception, problem analysis, data analytics, and data visualization, to implementation and deployment (Fig. 2).

establishment of a consultancy firm for the government. Since it can be difficult to develop experts in complex fields like data science and data engineering in the early stages of HRD, the government may think about creating an agency that is structured like an agile team with current high-skilled data analytics professionals on board to advise and mentor government officials in particular. A few government organizations can initiate the digital transformation process by developing in-house data analytics expertise and utilizing early big data implementation and deployment as proofs of concept and hands-on learning.

establishment of an open government data platform for residents and businesses, offering information, best practices, and data resources to all parties involved—including regular people. The platform has the potential to raise public knowledge of and use of big data analytics, which could result in increased domestic and international Public-Private Partnerships (PPP) and sustainable economic growth from digital technologies.

The Government Big Data Institute (GBDi) was founded by the Thai government in particular to act as a government consulting body for big data analytics and technologies. In addition to supporting interdisciplinary partnerships with government agencies to deploy new data-driven technologies and policies, GBDi's primary missions encompass all three elements of the above-mentioned human resource development model (section III-G).

C. Data Governance in Government

A significant amount of data, both digital and physical, will be generated and handled by government organizations and ministries during the digital transformation process. A government organization can oversee data exchange, sharing, and control over data quality throughout the data lifecycle by implementing data governance procedures as part of its business goals. We suggest implementing the four components of the data governance paradigm listed below (Fig. 3):

1) The creation of the data governance committee, which will include project managers, executives, users, stewards, and custodians of data. In addition to purchasing IT solutions for data management and storage, the committee works together to establish and evaluate high-level policies that specify data management procedures, data ownership, and departmental accessibilities.

2) Data management policies should be examined from four angles:

Data life cycle. Organizations may reduce data storage costs and ensure compliance with relevant requirements, like data privacy laws, by having a clear understanding of how data is gathered, used, kept, archived, and deleted throughout its life cycle. Organizing current data into datasets with metadata completed and gathered in an organization-level data catalog—a platform that allows data users to find datasets for business objectives—is a necessary step before creating life cycle management policies. Section III-D goes into great length on the data catalog.

Privacy and security of data. To preserve data integrity, confidential data must be appropriately labeled, accessible by authorized users only, and limited to the read, write, process, exchange, and share operations that are permitted. As stated in section III-A, data should always be accessible to data consumers and backed up to prevent data loss from unforeseen events. In order to ensure data security and privacy, the committee may collaborate with data custodians to identify and create IT solutions.

Assurance of data quality. A number of measures, including correctness, consistency, and completeness (Fig. 3), can be used to assess a dataset's and its metadata's quality. For the benefit of the entire company, the committee and data stewards may collaborate to create and uphold good data quality.

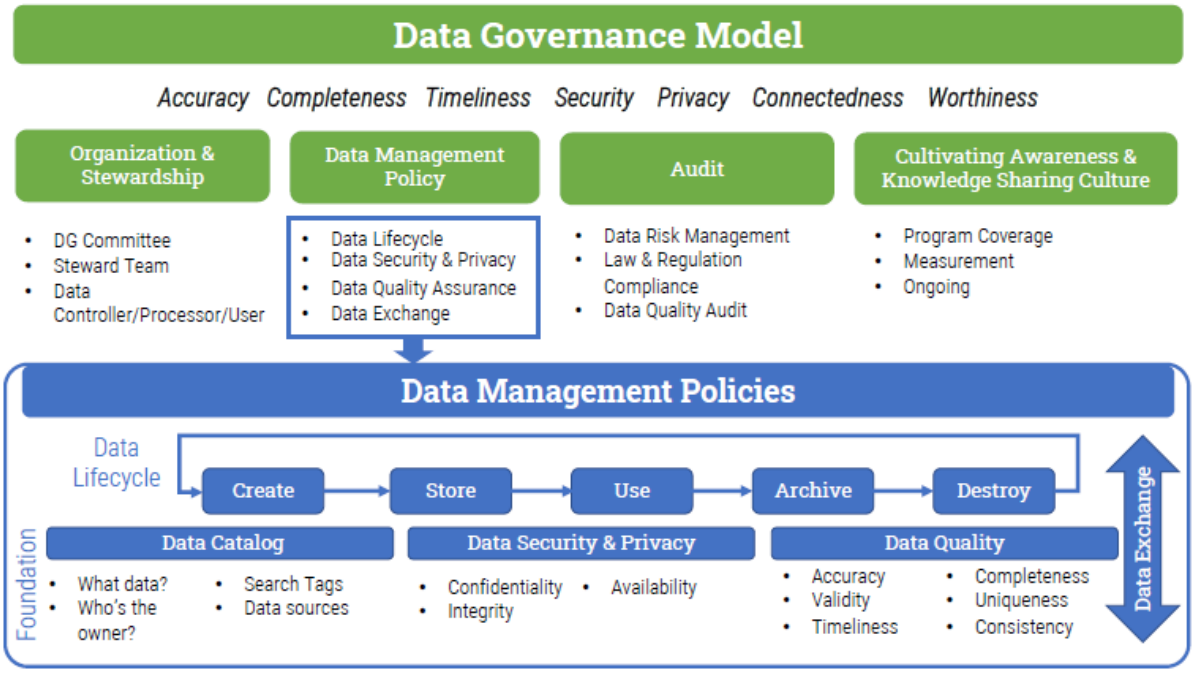

Fig 3: Model of data governance.

Information sharing. In accordance with existing laws and regulations, the committee must specify the terms and conditions of data usage for both importing and exporting data, encryption standards, and the individuals in charge of keeping encryption keys. The committee must also establish rules about the movement of data among various staff members and internal departments, the length of time that copies of the data are held at various points in the data's life cycle, and the methods for logging and preventing tampering with the data. Section III-E will cover a central government-level data exchange platform.

3) Data governance audits: conducted by an external entity as well as an internal team to check for adherence to data quality assurance guidelines, risk management, and relevant laws and regulations.

4) Fostering awareness and a culture of information sharing: Since good communication is essential to the success of data governance [11], it is imperative to cultivate awareness of the advantages of data governance through a persistent and consistent approach. For instance, a government agency might schedule frequent data governance workshops for the entire organization and use quantifiable indicators to monitor its progress.

The Digital Government Development Agency (DGA) in Thailand has created and published official government-wide data governance rules. Additionally, DGA has been working in conjunction with the National Data Catalog platform development team (NECTEC), the National Statistics Office (NSO), GBDi (section III-B), and the National Electronics and Computer Technology Center (NECTEC) to co-design metadata specifications. These collaborations are covered in more detail in the section that follows.

D. Catalog of Government Data

These days, a lot of data may be produced fast, in a variety of forms, and from various data sources. As a result, data users are finding it more and more difficult to search for and identify the datasets they need.

In order for data users to create value and contribute to a sustainable economy, the government should offer a Government Data Catalog, which functions as a platform akin to the yellow pages for searching and locating datasets owned by the government. Data users can often find datasets they want to request access to by searching through their metadata; they don't always need to access the data directly first. Additionally, the data catalog can let government agencies communicate data via established methods.

A metadata database, a data linking system, and a data directory service portal make up a data catalog architecture (Fig. 4).

A data directory service portal gives users of data a front-end interface for searching and perusing the metadata of datasets. Users have the option to search using tags, business sector, data owner, and other parameters. In order to shield sensitive or classified information from public view, the portal should additionally have an authentication layer.

Details about the data itself are kept in a metadata database. The database might include attribute-level data dictionaries such attribute names, data types, and descriptions, as well as dataset-level metadata like a dataset owner, tags, collecting techniques, and instructions on how to obtain access to the data. Be aware that this database of metadata does not need to save the actual data; all that is required for a data catalog to function is dataset information.

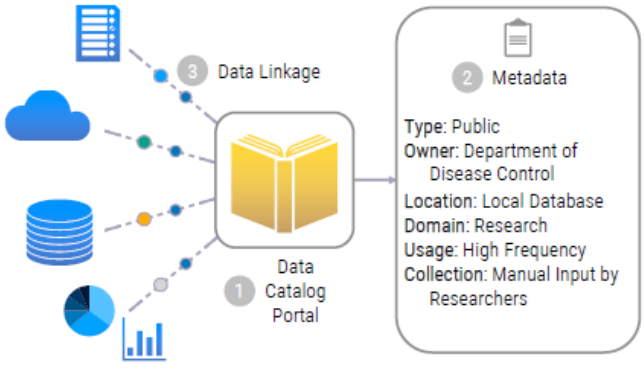

Fig. 4. Data catalog architecture

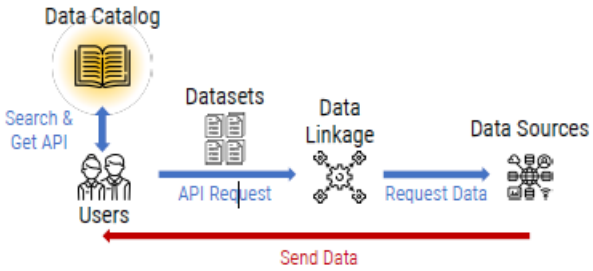

Fig. 5. User journey of a data catalog platform

In the backend, a data connection system stands in between the database and the portal. If a user has the necessary access rights, this system enables them to send API queries through it to obtain the actual data. As a result, the system needs to be able to establish connections with real data sources, enable database queries that are delivered as part of an API request for preparing data, and give data owners administrative tools to arrange their datasets according to access rights and data classification levels.

The data catalog architecture facilitates the easy search, identification, and request of datasets by leveraging the data connection (Fig. 5). On the other hand, a government agency might decide to start a data catalog project with just two elements: a database that contains real data access instructions in a metadata field, and a directory portal.

E. Exchange of Government Data

Data users and data owners can systematically access and exchange data by using a government-level data exchange platform in conjunction with the government data catalog. The platform should promote data interchange inside or between companies, support both public and private channels, offer user authentication features and access privileges only when needed, and log all data transactions.

Machine and domain independence are desirable qualities in a data interchange format [12]. It should ideally be machine-readable without requiring any proprietary software. Better yet, the format should minimize the amount of preprocessing or parsing required to use the data. Given these standards, formats

Data formats like RDF, XML, and JSON are better suited for data communication than formats like Excel, Word, PDFs, pictures, or even paper data (Fig. 6).

The organization should make sure that datasets meet the Data Governance Committee's (section III-C) standards for Data Quality Assurance, which include high levels of Accuracy and Conformance, before exchanging data.

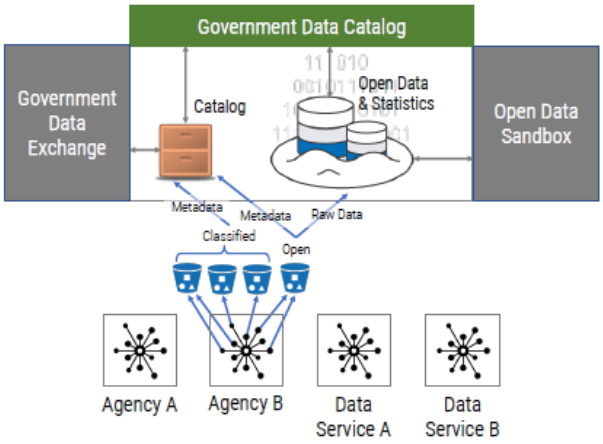

Fig. 6. Data exchange architecture

While data exchange formats, data quality assurance, and access control and user authentication features are widely used, the sharing of data with varying classification levels (public and non-public) results in various implementation opportunities and restrictions. We suggest two different kinds of data exchange platforms: an Open Data Sandbox for public data exchange and a Government Data Exchange (GDX) service for non-public data exchange. In order to provide data services to various data users and demands, both platforms collaborate with the Government Data Catalog (Fig. 6).

1) Classified Data Exchange via the Government Data Exchange (GDX):

It is necessary to share private or restricted datasets in a methodical manner that includes transaction logging, access right management, and user authentication. The following steps can be used to characterize the data exchange workflow:

Data owners submit dataset listings and metadata to the Government Data Exchange (GDX) platform; no actual data is uploaded.

A data user uses the central Government Data Catalog to search for and identify datasets before submitting an access request.

After deciding on permissions and access privileges, GDX notifies the owners and users of the data.

To prevent a bottleneck, actual data transmission can take place through the GDX system or a custom channel between the data owners and the data users. Options for transferring data can include email, SFTP, APIs, and encrypted data stored on a portable drive.

2) Open Data Sandbox: It's simple to exchange and work together on public datasets. As a result, the public data exchange platform we suggest functions more like a sandbox, like to Kaggle, where users can share public data but including conducting cooperative research, analysis, and knowledge sharing.

This one-stop open data analytics platform is very beneficial to both government employees and regular citizens, and it also helps the government provide its citizens with better access to big data analytics resources and data literacy.

F. Statutes and Rules

The government must create and support pertinent laws and rules that guarantee the proper use of digital technology, control data exchange services, and foster public trust in security and privacy norms.

Three pertinent Acts have been passed in Thailand between 2017 and 2019 to further the aforementioned goals:

The Digital Development for Economy and Society Act (2017) establishes a committee to supervise the creation of the digital government's architecture and formally codifies the roadmap for digital transformation.

The Personal Data Protection Act of 2019 safeguards the rights of individuals who are data owners when they deal with the public or private sectors. These rights include the ability to consent or withdraw it, the right to know how their personal data is gathered, used, disclosed, or stored, and the right to request that their data be updated, deleted, or copied.

The Cybersecurity Act of 2019 allows the government to monitor and track digital data in order to protect the nation from cyberthreats. It also enforces legal safeguards against cybercrime and requires critical infrastructure entities and essential government organizations to implement strong security measures in accordance with government guidelines.

G. Informed and Transparent Governance

Government will fully realize its potential as a Smart and Open Government once ministries, agencies, and state enterprises are able to store, manage, and analyze data in a systematic manner; develop the workforce required to support this process; and foster a shift in culture toward making decisions based on data; and use data exchange mechanisms to solve complex problems with data analytics. In addition to offering competitive advantages to the

private sector and efficiently resolving citizen complaints, a smart and open government may also maintain cost-effectiveness and transparency.

Here, we give the Thai government concrete instances of big data analytics being applied widely across many government subsectors. The Government Big Data Institute, or GBDi (section III-B), is the lead investigator for the following projects and works with directly accountable government agencies to enable and co-develop digital solutions:

Smart Tourism uses digital technology to disperse visitors around the nation rather than just a few well-known locations, ensuring that national tourism is sustainable on both an economic and environmental level. Platforms, offering individualized one-stop shopping throughout the client experience.

Data Integration System for Healthcare gives patients the right to access their own medical records in compliance with data privacy laws, enabling physicians to appropriately evaluate and treat emergency patients based on their medical profile, which is synchronized throughout a nationwide hospital network by citizen ID.

Sentiment analysis from social networks combined with data-driven policymaking enables the government to quickly and effectively determine public opinion or response to proposed policies by analyzing large-scale, intricate data sets, such as social media.

CONCLUSION

The Big Data Analytics Framework for Digital Government, which formalizes the digital transformation process, is what we have described in this article. We explain the status of active framework adoption in Thailand's environment and offer tailored models and protocols for each of the seven components required for the framework's implementation. The Thai government is extending the use of big data analytics to additional government subsectors into education, agriculture, workforce and employment, public finance, public housing, environment, and more as it continues to embrace this framework in phases. We are confident that the framework will be strong in guiding a government toward a smart and open government that can withstand future social, economic, political, and technological pressures because it is built on previously successful digital government models, standards, and architecture, with an emphasis on human resource development, laws, and regulations.

REFERENCES

[1] T. Janowski, “Digital government evolution: From transformation to contextualization,” Government Information Quarterly, vol. 32, no. 3, pp. 221 – 236, 2015.

[2] D. M. West, Digital government: Technology and public sector performance. Princeton University Press, 2005.

[3] A. Pavlichev and G. D. Garson, Digital government: principles and best practices. Igi Global, 2004.

[4] J. Melitski, M. Holzer, S.-T. Kim, C.-G. Kim, and S.-Y. Rho, “Digital government worldwide: A e-government assessment of municipal web sites,” International Journal of Electronic Government Research (IJEGR), vol. 1, no. 1, pp. 1–18, 2005.

[5] K. Wende, “A model for data governance – organizing accountabilities for data quality management,” in ACIS 2007 Proceedings, no. 80, 2007.

[6] V. Khatri and C. V. Brown, “Designing data governance,” Communications of the ACM, vol. 53, no. 1, pp. 148–152, 2010.

[7] D. C. M. Initiative et al., “Dublin core metadata element set, version 1.1,” 2012.

[8] S.-S. Data, “Metadata exchange,” SDMX Standards Version, vol. 2, 2014.

[9] F. Piedad and M. Hawkins, High availability: design, techniques, and processes. Prentice Hall Professional, 2001.

[10] ISO/IEC, "Iso/iec 27001: 2013: Information technology – security techniques – information security management systems – requirements," 2013.

[11] P. Hopwood, "Data governance: One size does not fit all," Information Management, vol. 18, no. 6, p. 16, 2008.

[12] F. C. Billingsley, "General data interchange language," ISPRS Archives, vol. 27, pp. 80–91, 1988.