systems

MDPI

Article

Secure Data Management Life Cycle for Government Big-Data Ecosystem: Design and Development Perspective

Reeba Zahid [1], Ayesha Altaf [1], Tauqir Ahmad [1], Faiza Iqbal [1,*], Yini Airet Miró Vera [2,3,4,5], Miguel Angel López Flores [2,6,7] and Imran Ashraf [8,*]

1 Department of Computer Science, University of Engineering & Technology, (UET), Lahore P.O. Box No. 54890, Pakistan; reeba.zahid1998@gmail.com (R.Z.); ayesha.altaf@uet.edu.pk (A.A.); tauqir_ahmad@uet.edu.pk (T.A.)
2 Research Group on Foods, Universidad Europea del Atlántico, Isabel Torres 21, 39011 Santander, Spain; yini.miro@uneatlantico.es (Y.A.M.V.); miguelangel.lopez@uneatlantico.es (M.A.L.F.)
3 Research Group on Foods, Universidad Internacional Iberoamericana, Arecibo, PR 00613, USA
4 Department of Project Management, Universidade Internacional do Cuanza, Cuito EN250, Bie, Angola
5 Fundación Universitaria Internacional de Colombia, Bogota 111311, Colombia
6 Department of Project Management, Universidad Internacional Iberoamericana, Campeche 24560, Mexico
7 Instituto Politécnico Nacional, UPIICSA, Mexico City 04510, Mexico
8 Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea
* faiza.iqbal@uet.edu.pk (F.I.); imranashraf@ynu.ac.kr (I.A.)

Abstract: The rapid generation of data from various sources by the public sector, private corporations, business associations, and local communities is referred to as big data. This large and complex dataset is often regarded as the 'new oil' by public administrations (PAs), and data-driven approaches are employed to transform it into valuable insights that can improve governance, transparency, digital services, and public engagement. The government's big-data ecosystem (GBDE) is a result of this initiative. Effective data management is the first step towards large-scale data analysis, which yields insights that benefit your work and your customers. However, managing big data throughout its life cycle is a daunting challenge for public agencies. Despite its widespread use, big data management is still a significant obstacle. To address this issue, this study proposes a hybrid approach to secure the data management life cycle for GBDE. Specifically, we use a combination of the ECC algorithm with AES 128 BITS encryption to ensure that the data remain confidential and secure. We identified and analyzed various data life cycle models through a systematic literature review to create a data management life cycle for data-driven governments. This approach enhances the security and privacy of data management and addresses the challenges faced by public agencies.

Keywords: big data; data life cycle; GBDE; secure data life cycle

Citation: Zahid, R.; Altaf, A.; Ahmad, T.; Iqbal, F.; Vera, Y.A.M.; Flores, M.A.L.; Ashraf, I. Secure Data Management Life Cycle for Government Big-Data Ecosystem: Design and Development Perspective. *Systems* **2023**, *11*, 380. https://doi.org/10.3390/systems11080380

Academic Editors: Lianyong Qi, Xiaokang Zhou and Xuyun Zhang

Received: 23 May 2023
Revised: 4 July 2023
Accepted: 20 July 2023
Published: 25 July 2023

1. Introduction

The concept of data originates from the Latin phrase "datum", referring to a distinct, unstructured, and unprocessed entity. Organizations process these data based on their specific needs to extract relevant information such as events, objects, facts, or concepts [1]. Big data, on the other hand, refers to data that possess characteristics like high volume, velocity, variety, and veracity. Effectively managing big data requires cost-effective and advanced approaches to processing, enabling organizations to gain insights and to facilitate decision-making [2–4]. In the digital marketplace, big data has emerged as a crucial strategic asset for data-driven organizations, offering them attainable advantages [5–7].

The proliferation of big data is evident across various sectors, both public and private. As highlighted by [8], big data is often referred to as the "oil" of the twenty-first century due to its value in driving success for non-profits, civil society, and government agencies. With the rapid generation of voluminous data from diverse smart information sources, private

and public corporations face the challenge of effectively managing big data throughout its life cycle to extract value from it [9,10]. Data security is also an essential part of its life cycle, and machine-learning-based encryption algorithms have been presented as well to safeguard data from security breaches and threats [11,12]. In recent years, data-driven organizations in the public and private sectors have developed data strategies and policies aligned with their goals and mission. These strategies encompass a range of skills and competencies related to data generation, storage, management, access, enrichment, publication, protection, analysis, sharing, use, privacy, and archiving. Within the government's big-data ecosystem, the primary tool for data management is the data life cycle, which encompasses all aspects of data from planning and collection to destruction, considering the inter-dependencies among these phases [13,14].

Maximizing public value through data utilization is a challenging endeavor that impacts technology, organizations, cultures, and individuals. The infrastructure for big data consists of interconnected elements and components across the entire data life cycle, including analytics, infrastructure, data models, and organizational structures [15–17]. Creating an effective data management environment is crucial for maximizing data value. The data life cycle encompasses planning, collection, distribution, use, reuse, and destruction stages, providing a top-level perspective for managing big data within the government's ecosystem and visualizing data flow and work processes [18].

This paper aims to design a secure data management life cycle tailored explicitly to the government's big-data ecosystem (GBDE). The objective is to address the challenges and limitations of existing data life cycle models, with a focus on maximizing the value of data, ensuring data privacy and security, and enhancing organizational performance within the government sector. Figure 1 provides an overview of the distribution of research on the data life cycle.

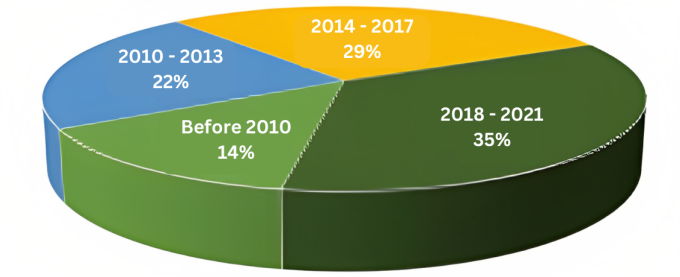


Figure 1. Distribution of research on the data life cycle.

Existing data life cycle models, while providing a high-level overview of data management stages, have certain limitations that are particularly relevant to the government's big data ecosystem. One major drawback is the insufficient attention given to security and privacy concerns associated with big data. Given the increasing frequency of data breaches, it is crucial to prioritize data protection throughout the entire life cycle, from data collection to disposal. Another limitation is the lack of adaptability of existing models to different contexts within the government sector. Current data life cycle frameworks often cater to specific industries or sectors, making them less suitable for managing data in the government's diverse landscape. Additionally, these models may not adequately account for the heterogeneity and diversity of big data sources, posing challenges in designing a comprehensive data management strategy.

Furthermore, existing data life cycle models often overlook the ethical considerations associated with big data, including the potential for bias and discrimination. These ethical concerns are of particular importance within the government sector, where decisions based on data can have significant consequences for individuals and society as a whole. The government's big-data ecosystem (GBDE) encompasses various stakeholders, including public administrations, government agencies, and potentially private corporations involved in data management. It involves the generation of diverse types of data, which may include

sensitive information related to citizens, public policies, and governmental operations. Effective data management within this ecosystem is essential to maximize the value of data, to enhance decision-making processes, and to ensure transparency and accountability in public administration.

To provide a more comprehensive review, it is essential to consider the relevant work related to the specific forms of big data, such as images. The following articles address image encryption algorithms and schemes: One is "Double Image Encryption Algorithm Based on Neural Network and Chaos". This article proposes a double image encryption algorithm that combines neural networks and chaotic systems. The algorithm aims to enhance the security and robustness of image encryption by leveraging the complex dynamics of chaotic systems and the learning capabilities of neural networks. It provides a novel approach for encrypting images, ensuring their confidentiality and integrity. Second is "Medical Image Encryption Scheme Based on Self-Verification Matrix". This article focuses on the encryption of medical images, which are crucial for patient privacy and data security. The proposed scheme utilizes a self-verification matrix to encrypt medical images, ensuring secure transmission and storage. The scheme provides a balance between security and efficiency, taking into account the specific requirements and characteristics of medical image data. By considering these studies, the review can encompass the encryption of specific forms of big data, particularly images. These articles contribute to the field by proposing novel encryption techniques tailored to the unique requirements and characteristics of image data, addressing challenges related to security, privacy, and integrity in image-based big data applications.

In summary, this paper aims to design a secure data management life cycle for the government's big-data ecosystem (GBDE) to address the challenges of existing models and to optimize data utilization, privacy, and security within the government sector. The scope of the GBDE includes multiple stakeholders, diverse data sources, and the imperative for effective data management to achieve organizational goals and public value. For the remainder of the paper, in Section 2, a literature review is presented. Section 3 outlines the secure data management life cycle for the government's big-data ecosystem. Their cross-cutting concerns are analyzed in depth in Section 4. In Section 4.8, the study is concluded.

2. Literature Review

The literature review discusses various prior art models for the data life cycle. While these papers were selected based on their engagement in formal research and alignment with industry standards, it is important to highlight their strengths and weaknesses:

In [19], a security data life cycle is presented with stages such as data production, editing, presentation, processing, transport, and storage. IBM's data life cycle model [20] includes stages like data creation, use, analysis, sharing, updating, protection, archiving, storage/retention, and destruction.

The open government data (OGD) life cycle model by Elmekki et al. [21] includes stages such as data collection, publication, transformation, quality assurance, and archiving. Raszewski et al. proposed a research data life cycle model [22], focusing on phases like planning research, data collection, processing, analysis, publishing, sharing, storing, and using and reusing data. The abstract data life cycle model (ADLM) by Knud Moller [23] covers stages like planning, creation, enrichment, access, storage, archiving, feedback, and termination.

The NIST data life cycle model [24] emphasizes data analytics and includes stages related to data planning and filtering. Jetten et al. defined the life cycle of scientific research data [25], including stages like design, creation, processing, analysis, use, preservation, and access. Kumar Rahul et al. referenced a data life cycle for commercial and healthcare initiatives [26], covering phases like creation, storage, analysis, utilization, distribution, archiving, and deletion. McKeever presented the web content management life cycle [27], including phases like creation, deployment, sharing, data control, content management, storage, archiving, use, and workflow.

The Research360 data life cycle by the University of Bath [28] covers steps such as design, collection, capture, interpretation, analysis, manipulation, publication, release, discovery, and reuse. Arass et al. proposed an intelligent information life cycle [29] that focuses on transforming raw data into intelligent data. Nawsher Khan et al. presented the Hindawi data life cycle [30], which includes stages like data collection, filtering and classification, analysis, storage, sharing, publication, and data retrieval and discovery.

Panah et al. introduced a data life cycle model for the Internet of Things (IoT) ecosystem [31], covering steps like collection, processing, storage, utilization, and dissemination. The lifecycle of informational data science by Gislaine P. F. et al. [32] includes stages such as data gathering, storage, visualization, and disposal, with a focus on GDPR criteria and blockchain management. It is important to consider these strengths and weaknesses when evaluating the applicability of these models to the government's big-data ecosystem. Table 1 provides a critical summary of discussed research works.

Table 1. Critical summary of the discussed research studies.

Ref.	Year	Stages	Strengths	Weakness
[28]	2003	Data creation, deployment, sharing, data control, content management, storage, archive, use, and workflow.	Centered on scientific studies data.	Limited discussion on other industry sectors.
[24]	2012	Data planning, data creation, data enrichment, access, storing, archive, feedback, and termination.	Emphasizes data analytics.	Limited coverage of other data life cycle stages.
[21]	2013	Creation, data use, analysis, sharing, updating, protection, archiving, storing/retaining, and disposal.	Specifically tailored for OGD.	Limited discussion on other industry sectors.
[31]	2014	Data collection, classification and filtering, analysis, storing, sharing and publishing, retrieval and discovery.	Addresses IoT-specific privacy protection.	Limited applicability beyond the IoT context.
[25]	2015	Collection, preparation, analysis, and action.	Focus on scientific research data.	Limited applicability beyond research contexts.
[26]	2015	Creation, selection, analysis, curation, publishing, discovery, exploration, storage, and exploitation.	Addresses large-scale data analytics.	Limited discussion on specific industry sectors.
[30]	2018	Data planning, collection, integration, filtering, analysis, enrichment, visualization, access, storage, destruction, archiving, quality, and security.	Focus on big data tools and technologies.	Limited discussion on specific industry contexts.
[32]	2019	Acquisition, processing, storage, use, and dissemination.	Incorporates GDPR and blockchain considerations.	Limited discussion on other aspects of the data life cycle.
[22]	2019	Collection, transformation, publication, quality, interoperability, use, sharing, consumer feedback, and archiving.	Addresses research data management.	Limited applicability beyond research contexts.
[29]	—-	Design, collect and capture, interpret and analyze, manipulate and preserve, release, publish, discover, and reuse.	Emphasizes data transformation.	Limited coverage of other data life cycle stages.
[27]	2020	Create data, store, analyze, use, share, archive, and destroy.	Focus on website content management.	Limited discussion on other aspects of the data life cycle.
[23]	2020	Research plan, collection, process, and analysis; share, store, publish, use and reuse of data.	Focus on the semantic web.	Limited discussion on specific industry contexts.
[19]	2020	Data production, editing, presentation, processing, transport, and storage	Emphasizes data protection.	Limited discussion on other aspects of the data life cycle.
[20]	2020	Creation, editing, process, display, transfer, and store.	Comprehensive coverage of data life cycle stages.	Limited focus on specific industry contexts.

3. Secure Data Management Life-Cycle

Access to accurate and timely information is essential for effective organizational planning and policy development. However, effectively managing organizational data poses significant challenges that require careful attention. While government sectors excel in managing their physical assets, data should also be treated as valuable assets and managed accordingly. Understanding the available data, its location, its precision, and its retrieval methods is crucial for proper management. It is important to ensure that necessary data are easily accessible to facilitate critical decision-making while being mindful of the cost involved in acquiring data.

Data management is a complex and multifaceted subject, making it challenging to grasp all its components. To address this complexity, we employed an iterative and cyclical approach to identify the useful components and applications of data within our organization's context. Through this process, we have developed a comprehensive framework that outlines the available information and distills it into overarching challenges and key areas within the secure data management life cycle (SDMLC). This life cycle provides a structured method for organizing data, tracking its value over time, and identifying potential political implications associated with data management challenges.

Figure 2 presents the SDMLC, illustrating the key stages involved in the management of data from its creation to its destruction. Additionally, it addresses the specific challenges that impact data at each stage of its life cycle. This framework builds upon existing literature and incorporates novel approaches to address the identified gaps in prior research. By integrating insights from existing data life cycle models and considering the specific needs and requirements of our organization, the SDMLC provides a tailored and secure approach to data management in our unique context. The SDMLC comprises the following seven phases

- Collecting;
- Processing;
- Storing and securing;
- Using;
- Sharing and communicating;
- Archiving;
- Destroying or reusing.

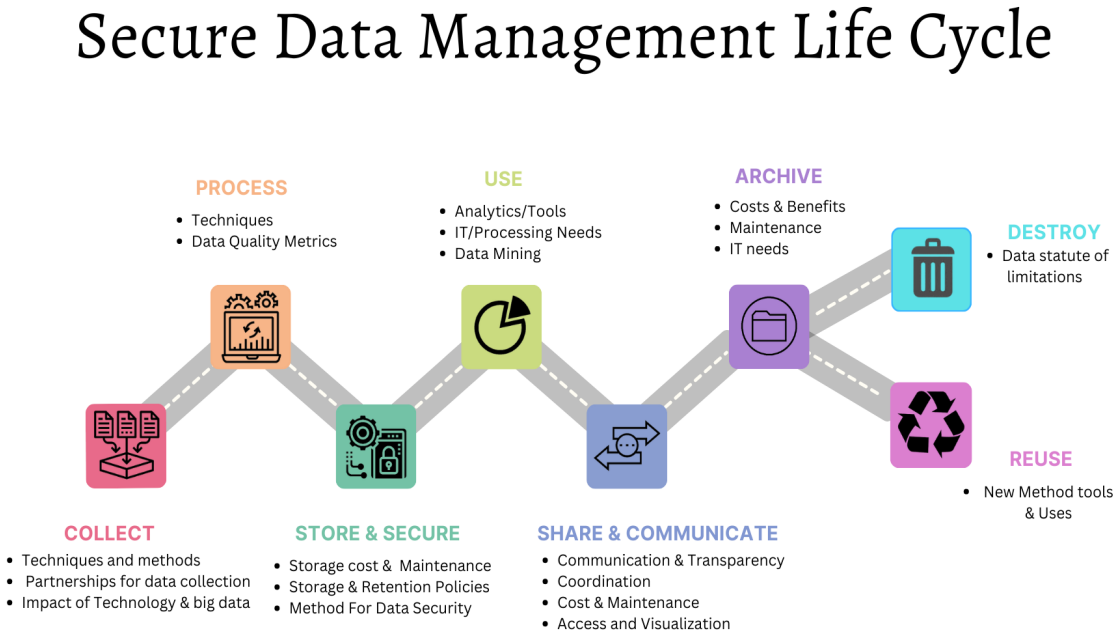


Figure 2. Architecture of the proposed secure data management life cycle.

In addition, we have identified seven concerns that impact each of the seven stages of the SDMLC and can propagate over its lifetime (as depicted in Figure 2). Each stage of the cycle is crucial to cross-cutting issues, and they all impact policy-making. The seven interrelated issues are as follows:

- Purpose and value;
- Privacy;
- Data ownership;
- Liability;
- Public perception;
- Security;
- Standards and data quality.

The interrelated issues mentioned in the SDMLC impact different stages of the data life cycle. Here is a breakdown of the interrelated issues and the stages they pertain to:

i. Purpose and value: This intersection issue relates to the collecting stage of the SDMLC. It involves identifying the purpose and value of the data being collected to ensure that it aligns with organizational goals and objectives.
ii. Privacy: The privacy issue spans multiple stages, including collecting, processing, storing, securing, using, sharing, communicating, and archiving. Privacy concerns involve protecting sensitive information and ensuring compliance with privacy regulations throughout the data life cycle.
iii. Data ownership: The issue of data ownership is relevant during the collecting stage. It pertains to determining who owns the collected data and establishing appropriate data ownership rights and responsibilities.
iv. Liability: The liability issue affects various stages of the SDMLC, including collecting, processing, storing and securing, using, sharing and communicating, and archiving. It involves identifying and addressing potential liabilities associated with the data, such as data breaches, unauthorized access, or misuse.
v. Public perception: The issue of public perception relates to the using stage of the SDMLC. It involves considering how the data are utilized and how they may be perceived by the public, stakeholders, or other entities. Maintaining public trust and addressing concerns related to data usage is crucial at this stage.
vi. Security: The security issue encompasses multiple stages, including collecting, processing, storing and securing, using, sharing and communicating, archiving, and destroying or reusing. It involves implementing measures to protect data against unauthorized access, breaches, and other security threats.
vii. Standards and data quality: The issue of standards and data quality is relevant throughout the entire SDMLC, influencing all stages. It involves adhering to data standards; ensuring data accuracy, reliability, and integrity; and maintaining data quality throughout its life cycle.

By addressing these intersections and their corresponding stages, the SDMLC aims to identify and mitigate potential challenges and issues that may arise during the data life cycle, enabling effective decision-making and policy development.

3.1. Data Collection

Data collection is the initial step of the SDMLC. Data are collected for a variety of purposes, including formulating plans and operations, measuring general performance, maintenance, and addressing a particular policy goal or objective. The essential components of this level are as follows:

3.1.1. Techniques and Strategies for Data Collection

Data collection methods and systems vary based on geographic and jurisdictional requirements and user needs. Hence, it is crucial for data collection systems to satisfy the requirements of both internal and external users, as well as to comply with legal standards.

As explained in [33], planning and designing a data collection system involves defining data goals and requirements, identifying data sources, developing plans and procedures, and documenting the data collection process and system design. It is important to consider data quantity, data volume, duration of data collection, research topics, and target audiences when selecting a data collection strategy. In addition, future research should investigate how public organizations utilize large datasets obtained from private sources.

3.1.2. Partnerships for Data Collection

Collaborations for information gathering to identify and collect accurate and useful data, technical knowledge, and advanced technologies are required. In this instance, a public–private partnership should facilitate data collection and maximize the capacity of agencies as practitioners and decision-makers in data-driven development. Currently, most public–private partnerships are concentrated on state infrastructure and building projects. The prospective data collection partnership is not being officially directed. Prior to engaging in a public–non-public collaboration, it is essential to comprehend that existing data ownership laws specify rights and responsibilities to safeguard data integrity.

3.2. Data Processing

The second stage of the SDMLC is data processing, which is crucial for transforming the first stage of collected data into meaningful information. Often, raw data collected through data collection are not immediately useful and require further processing. This phase begins with the identification of any inconsistencies or anomalies in the raw data, followed by data cleaning to improve the data's quality. Behavioral analysis should be used to generate information that can facilitate decision-making, lead to problem-solving, and improve existing conditions. Important aspects of this phase include the following.

3.2.1. Data Quality Metrics

Data quality measures are crucial for identifying inaccurate data and fragments, and evaluating the impact of data-driven activities. Evaluating data quality can assist government agencies in assessing the accuracy of their visitor and protection data, among other things. This assessment aids businesses in determining whether their data accurately represents the intended objects, events, and concepts. To ensure uniformity among states, AASHTO has developed seven fundamental data principles, as depicted in Figure 3.

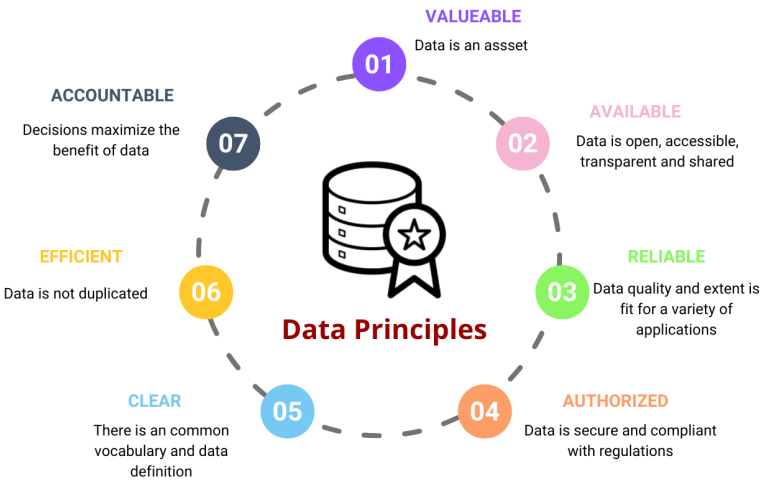


Figure 3. Fundamental principles for data.

3.2.2. Data Processing Techniques

The goal of public organizations is to harness big data's potential in order to make more informed decisions. However, traditional management and evaluation methods may not be sufficient to handle the complexity and volume of information assets. Advanced

data processing methodologies and tools are required to effectively extract and to mine vast amounts of data from multiple databases. This guide offers suggestions and ideas for data transformation and coding, dealing with missing data, generating estimates and projections, and evaluating and interpreting data. Utilizing advanced tools such as cloud computing can help stakeholders save time and improve the efficacy of real-time data processing. Cloud computing provides shared computer processing resources and data to multiple devices on demand, making it an excellent choice for government agencies [34].

3.3. Store and Secure

The third step of the SDMLC is data storage and security, which is crucial for ensuring trust and confidence in information use. It is essential to safeguard data from unauthorized access, regardless of whether it is false, malicious, or inaccurate. Government agencies at the federal, state, and local levels maintain a vast quantity of datasets that are typically stored in separate databases that may be difficult to locate or incompatible with one another. The essential aspects of this step include the following.

3.3.1. Storage Value and Maintenance

Due to primarily advancements in information technology and data collection, the amount of electronically stored data worldwide doubles every year. This exponential increase in government data necessitates economical storage options. As a result, an increasing number of organizations are adopting cloud storage alternatives or outsourcing storage services, which give customers access to computing power and storage space without requiring them to maintain them themselves. This allows customers to focus on information creation and analysis, while providers ensure the scalability, dependability, and performance of the computing environment [35].

3.3.2. Storage and Retention Policies

It is essential to consider the potential risks associated with cloud computing. These risks may include unauthorized access to data caused by cyber-attacks targeted at cloud service providers, security weaknesses within the cloud infrastructure, legal risks and compliance issues related to responsibility for data breaches, changes in pricing for essential features over time, and the risks associated with the unavailability of critical information due to cloud server downtime. Organizations must conduct a comprehensive risk assessment and develop a comprehensive risk management and mitigation strategy.

3.3.3. Method for Data Security

The proposed method basically includes the "upload" and "download" modules for this purpose.

Upload Module: It includes four steps.

- Authentication: With the help of special login and password, the user authenticates themselves to the cloud.
- Upload: This module enables a user to safely add their documents and to upload the data (document) in their encrypted form through this gateway to the user's cloud file location.
- Key generation: System time is the main factor in key generation.
- Encryption: Upon uploading, the data are initially stored in the server's temporary file within the cloud. Subsequently, to ensure security, the data must be encrypted using the consumer's public key and then saved in their files as the encrypted version. Finally, the temporary files should be detached or un-linked from the system.

Download Module: It includes two steps.

- Decryption: To download the secure data, the customer is required to provide a username as well as a confidential private key. The cloud decrypts the information using the consumer's private key.

- Download: The consumer receives their authentic information from the cloud after it has been decrypted.

3.4. Data Use

The fourth phase of the SDMLC is concerned with data utilization. Government systems are constructed and maintained using a variety of data-driven techniques that aid planners in comprehending user behavior and enable policymakers to enhance the system's efficacy and cost-effectiveness. Despite the wide range of practical applications for data, there are challenges that arise throughout the whole data life cycle, from data collection to deletion. The way data are collected, processed, and stored will determine how it is utilized. When reviewing data for whatever purpose, there are several considerations to keep in mind, including the following:

- Larger and more extensive data sources might provide challenging scenarios for researchers' analytical skills and processing tools, as well as difficulties in information exchange inside an organization or with partners.
- Users must balance increased access to and availability of information with their ability to process and interpret it.
- Striking a balance between legal information usage and data access security concerns.
- Privacy and proprietary rules for using the data that has been collected.

3.5. Share and Communicate

As government entities engage in collaboration with external stakeholders and partners for decision-making, planning, and operations, the demand for more data can increase. Expanding the sources of data and incorporating newer datasets from both internal and external resources can lead to improved decision-making. This can help agencies and researchers gain a more comprehensive understanding of the effects of their decisions, ultimately resulting in better outcomes. Shared data can also encourage decision-makers to demand a higher level of quality and clarity from the information they acquire, resulting in the use of fewer resources to make decisions that are more precise and timelier. The fifth stage of the SDMLC is data sharing, which must strike a balance between the need to limit the provision of sensitive, proprietary, and classified data and the advantages of open sharing and data release through applicable settlements. To effectively share and communicate information, it is essential to consider the following factors:

- Communication and transparency;
- Coordination within and outside the agency;
- Costs and maintenance of shared information;
- Access.

3.5.1. Communication and Transparency

Increasingly, governments are attempting to engage and involve the public in organizational decision-making processes, and promoting transparency is a crucial component of this effort. This is accomplished in part through the sharing of public datasets. Typically, state and federal government entities have websites with information about their public datasets. The website contains basic information about each dataset including its title, contact information, description, source, and update frequency.

3.5.2. Coordination

As organizations engage in decision-making, planning, and operations with stakeholders and external partners, the need to exchange information increases. Furthermore, there is a growing demand for businesses to achieve more with fewer resources. To meet this demand, state and local organizations have developed sophisticated management systems that integrate data from various public and private entities, enabling the coordination of data networks for an entire region across different modes, jurisdictions, and corpora-

tions. This type of local coordination complicates the purchase and design of information technology systems, as well as how information is shared between public and private entities [36].

3.5.3. Costs and Maintenance of Shared Data

Information sharing can be an effective strategy for addressing the rising costs of data storage, processing, and analysis. As data become more accessible, it can also assist with identifying cost-effective and efficient solutions. By sharing data, organizations can pool their resources and expertise to generate more complete and accurate insights, resulting in improved decision-making and outcomes. Moreover, sharing data can help organizations avoid the costs associated with developing and maintaining redundant systems by reducing duplication of effort and eliminating the need for redundant systems. Sharing information can provide organizations with a variety of benefits, including cost savings, increased efficiency, and improved decision-making.

3.5.4. Access

Sharing data is a crucial aspect of making information more accessible and reducing time demands on workers. It is important to ensure that users have access to the necessary data to perform their duties effectively. The widespread availability of well-packaged and processed data can improve decision-making performance and effectiveness, as well as provide prompt responses to data requests. The need to limit the disclosure of sensitive, proprietary, and confidential information must be weighed against the advantages of sharing and releasing data through public and private partnership agreements. There are significant policy implications regarding the sharing or withholding of organizational information, particularly in the areas of data ownership, security, privacy, and responsibility. For instance, the growing availability of new data in government systems that collect and transmit data may raise privacy concerns among users [37].

3.6. Archive

Archiving is the sixth stage of SDMLC, which is a process of identifying and transferring inactive information from active production systems to specialized long-term storage systems. This improves current performance by reducing the load on active systems and databases while ensuring that the information is still accessible [38]. For this purpose, an information archive, also known as an information bank or information center, is utilized. However, when implementing information archiving, state and federal governments must consider IT requirements, storage costs, information backup needs, cost–benefit analyses, and other data backup-related issues.

Data users have been accustomed to archiving data for quite some time. However, as data are produced at a faster rate and in greater quantities, the complexity and expense of archiving it increase. Data archiving necessitates the application of diverse software, database, and digital data storage technologies. In addition, a member of staff is required to manage these systems, to produce reports, and to provide IT and administrative support.

3.7. Reuse/Re-Purpose

Re-purposing and reusing data is a valuable strategy that enables organizations to extract ongoing value from their data assets, thereby helping to justify the costs associated with collecting and managing large volumes of data. Many organizations have relied on their backup and archive data as their most comprehensive source of data, but they typically only utilize it for data recovery. Continuous reuse and re-purposing of data can result in the creation of new data products that are processed, discovered, analyzed, distributed, and preserved. As a result, the potential value of data can be mined and leveraged. By monetizing or productizing their data assets, businesses can realize their data's full potential and drive innovation and growth.

3.8. Destroy

When it comes to destroying data, the chosen method typically depends on how sensitive the data are or how much damage it could potentially cause if it were to be accidentally recovered or disclosed. If data owners decide to destroy the data, they must consider a few factors including when the data should be destroyed, who should make the decision, and the applicable laws governing document storage. In addition, they should weigh the importance of historical data against the need for current information sources and be aware of any "statute of limitations" for data.

4. Cross-Cutting Issues in Data Management

The research group has identified seven cross-cutting issues that affect all phases of the SDMLC. A summary of cross-cutting issues in SDMLC is presented in Figure 4.

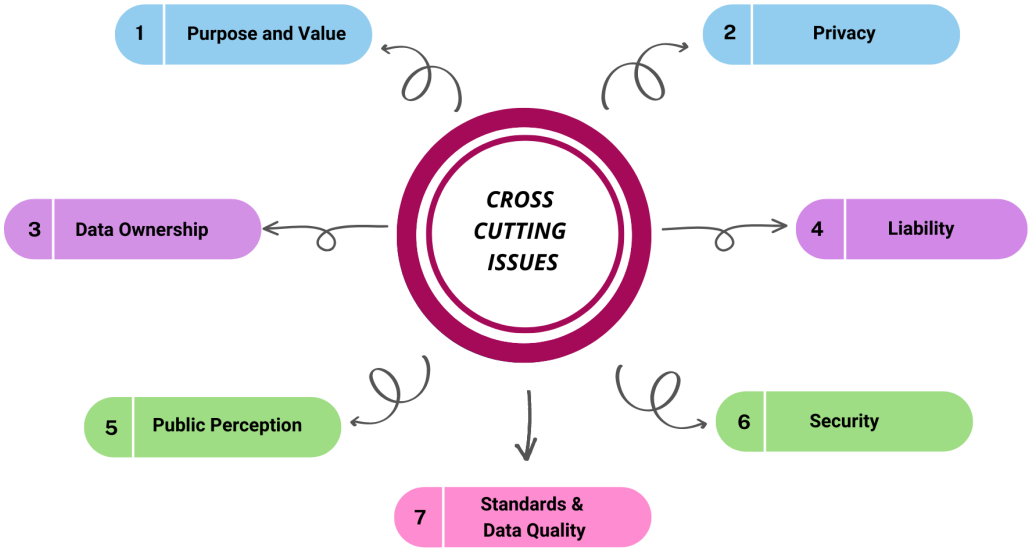


Figure 4. Cross-cutting issues in SDMLC.

4.1. Purpose and Value

Data are gathered, generated, and communicated with goals in mind. To ensure effective management, it is crucial to comprehend the objective, significance, and intended recipients of each data stage in the SDMLC. The significance and purpose of the data should be at the center of each stage, and each should have a justification for its necessity and significance. When considering the reuse, re-purposing, or monetization of data, it is essential to comprehend its purpose and value.

4.2. Privacy

The issue of data privacy impacts each phase of the SDMLC. Privacy is essential when determining compliance standards and levels, permitted uses, and procedures. Safety and privacy concerns regarding data collection are on the rise.

4.3. Data Ownership

At different stages of a company's data life cycle, ownership and control of the data vary, resulting in divergent perceptions. Individuals are the owners of the data they collect, whereas businesses own the data they collect. Even though some data may be associated with private citizens, they rarely have authority over it. The public may perceive that the individual owns the data, but ownership only encompasses privacy rights that govern how companies can use the data and prevent the disclosure of personally identifiable information.

4.4. Liability

The definition of "personally identifiable information" is inextricably linked to technological advancements, particularly in terms of liability. The management and protection of data assets against data breaches and cyber attacks, which can have significant security implications and affect the daily operations of the network, are the responsibility of both the public and private sectors. Damages caused by a data breach or cyber-security attack could be assessed against organizations responsible for collecting and controlling public information.

4.5. Public Perception

There is a close relationship between the public's perception of data management and various SDMLC stages. The Pew Research Center has discovered a strong correlation between public opinion on data management and the various SDMLC stages. Sixty-eight percent of internet users believe that the current legal framework does not adequately protect their online privacy. In addition, younger adults place a greater emphasis on privacy concerns than many of their elders and frequently take steps such as concealing their identities and removing their names from tagged images. In addition, approximately 75% of Americans believe that having control over their personal data is of the utmost importance [39].

4.6. Security of Data

Data security is a critical aspect of the secure data management lifecycle, aiming to protect private data assets from unauthorized access and to ensure the confidentiality, integrity, and availability of digital information. In the context of storage security, various measures and strategies are implemented to safeguard data stored in computers, databases, websites, and mobile devices.

Storage security involves the integration of storage, networking, and security disciplines, technologies, and methodologies. The objective is to protect digital assets stored in various storage systems, such as cloud storage, network-attached storage (NAS), or storage area networks (SANs). The Storage Networking Industry Association (SNIA) defines storage security as the practice of ensuring compliance with legal regulations, enabling e-discovery readiness, managing customer access, and physically securing data storage [40]. Compliance with legal regulations is a crucial aspect of storage security [41]. Organizations need to adhere to relevant laws and regulations pertaining to data protection, privacy rights, and confidentiality. This includes implementing measures to report data breaches, performing regular backups to ensure data recoverability, employing data masking techniques to anonymize sensitive information, securely erasing data when no longer needed, and applying encryption methods to protect data in transit and at rest.

Access authentication plays a vital role in storage security, ensuring that only authorized users and networks can access storage resources [42]. This involves implementing robust user authentication mechanisms, such as strong passwords, multi-factor authentication, or biometric authentication, to prevent unauthorized access to data storage systems.

Furthermore, physical protection of data storage is essential to safeguard against physical theft, damage, or unauthorized access to storage devices. This may include implementing access controls to data centers, surveillance systems, and environmental controls to maintain optimal storage conditions. By incorporating these information security elements into storage security strategies, organizations can effectively address data protection concerns; meet legal requirements; and ensure the confidentiality, integrity, and availability of their stored data. These measures are crucial in safeguarding sensitive information, maintaining customer trust, and mitigating the risks associated with data breaches and unauthorized access to data storage systems.

4.7. Standards and Data Quality

All phases of the SDMLC require high-quality data, as data standards guarantee reliable outcomes. Quality control and assurance aim to detect and correct data inconsistencies and errors to enhance data quality. This process is vital for both self-collected and privately acquired data. The primary goal of data quality control is to ensure that data are collected efficiently and produce meaningful results. However, data utilization and management can be challenging due to inconsistent accuracy and resolution levels and inappropriate data formats, despite the vast amounts of data being produced, as evidenced by various studies. As data volumes continue to grow, these problems worsen, resulting in underutilized data and increased data management expenses, which have a negative effect on data quality [43].

4.8. Proposed Methodology

A comprehensive approach to data security is crucial for effective data management. In the paper [44], a hybrid encryption technique combining RSA and AES is proposed to enhance cloud data security. Similarly, in this work, we adopt a hybrid strategy for data encryption. However, we have made significant improvements by replacing the RSA encryption algorithm with the more robust ECC encryption algorithm, known for its superior resistance to modern cracking techniques. While encryption plays a vital role in safeguarding data, it is important to note that it is just one aspect of a comprehensive data security framework. Encryption primarily focuses on protecting data at rest and in transit, ensuring its confidentiality. However, effective data management also entails addressing other critical aspects of security, such as access control, data integrity, and security protocols for data handling in various scenarios.

In the proposed approach, we not only employ the advanced ECC encryption algorithm but also consider a holistic security framework for data handling. This includes implementing robust access control mechanisms to determine who has access to the data and under what conditions. We also emphasize the use of secure communication protocols for data transmission, ensuring data integrity and protecting against unauthorized interception or tampering. Furthermore, our data security strategy encompasses measures for secure data storage, such as employing encryption for data at rest, implementing secure backup and recovery procedures, and regularly auditing the data infrastructure for potential vulnerabilities. We also address the enforcement of security policies and regulations within the organization to ensure compliance with relevant data protection laws and standards.

By integrating encryption with a comprehensive data security framework, we aim to provide a robust and holistic approach to data protection. Our strategy goes beyond encryption algorithms alone and addresses the broader aspects of data security, including access control, data integrity, secure communication, and compliance with data protection regulations. Figure 5 shows the workflow of the proposed hybrid approach.

ECC uses significantly shorter key lengths than RSA but provides the same level of security. Therefore, brute-force attacks are less effective against ECC, especially for longer keys. ECC has a number of advantages over RSA including enhanced performance and scalability. ECC is more secure against modern cracking techniques due to its complexity and requires shorter key lengths, resulting in less network and CPU usage. This is particularly advantageous for devices with limited storage and processing capabilities. In addition, using ECC in SSL/TLS certificates can speed up SSL/TLS handshakes and improve website load times. The proposed approach executes the following steps

i. Registration: The registration process involves users providing their credentials and securely creating an account within the system. Specific mechanisms, such as password hashing and salting, can be employed to protect user passwords from unauthorized access.
ii. Sign in using TCP: Once registered, users can sign in to the system using a secure TCP connection. This process may involve secure authentication protocols, such as

challenge-response mechanisms, to verify the identity of the users and to ensure secure communication between the client and the server.

iii. ECC key exchange: After successful sign-in, the client and server engage in an ECC key exchange to establish a secure communication channel. The specific ECC parameters used, such as the elliptic curve type (e.g., NIST curves) and key length, should be clearly specified in the methodology. This information is crucial for evaluating the security and performance aspects of the proposed approach.

iv. Data encryption for uploading and downloading using AES: Once a secure channel is established, the client encrypts the data using the AES algorithm. The encryption process involves generating a random AES initialization vector (nonce) and combining it with the plaintext message. The AES encryption produces ciphertext. The ciphertext is then combined with the MAC code (authTag) obtained via GCM block mode, providing data integrity and protection against unauthorized tampering. Additionally, a temporary public key (ciphertextPubKey) is randomly generated and added to the encrypted message. This key allows for the retrieval of the AES symmetric key during decryption using the ECDH key exchange scheme.

v. A storage server is used to store and retrieve data: The encrypted data are uploaded to a storage server for secure storage. The storage server should employ measures such as encryption for data at rest, secure backup and recovery procedures, and regular auditing of the data infrastructure for vulnerabilities. These measures ensure the security and integrity of the stored data.

vi. Sign off: When the user is finished with their session, they can sign off securely, terminating the communication and ensuring that the session ends without leaving any vulnerabilities.

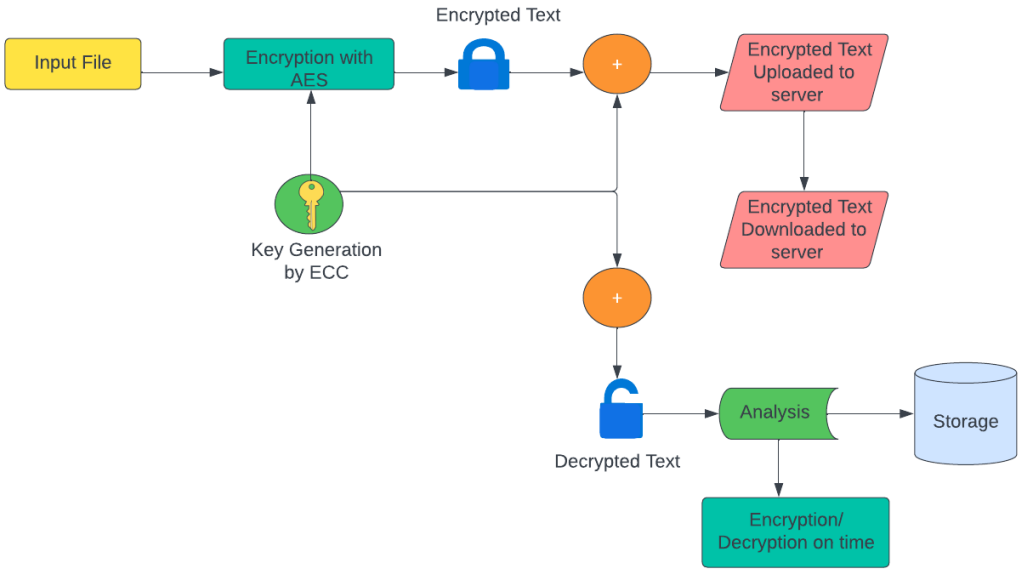


Figure 5. Hybrid approach comprising ECC and AES.

Following the generation of the ECC key pair for the message using the 'tinyec' library, the message is encrypted using "pubKey". This produces the outputs ciphertext, nonce, *authTag*, and *ciphertextPubKey*. AES-GCM encryption is combined with the nonce, which is a random AES initialization vector, and the authTag, which is the MAC code of the encrypted text obtained via GCM block mode, to extract the ciphertext. In addition, *ciphertextPubKey*, a temporary public key that can be added to the encrypted message and used to retrieve the AES symmetric key during decryption, is randomly generated. This utilizes the ECDH key exchange scheme, as previously demonstrated.

To decrypt the encrypted communication, we combine the decryption privateKey with the information generated during encryption including the ciphertext, nonce, *authTag*, and *ciphertextPubKey*. By using the decryption key or other relevant information, the decryption

process generates the plaintext message. However, if the decryption key or information is incorrect, the decryption process will fail and generate an exception.

5. Results and Benefits of Proposed System

This section contains the experimental results and discussions.

5.1. Experimental Setup

This study investigates the use of RSA and ECC for data confidentiality with three sets of 8-, 64-, and 256-bit data inputs, as well as randomly generated private keys. The experiments were conducted using MATLAB and 2 GB of RAM on Windows 10 platform.

5.2. Experimental Results

The experiments include different inputs including 8-bit, 64-bit, and 256-bit data inputs for ECC and RSA algorithms. Table 2 shows the results for 8-bit input for both algorithms regarding encryption and decryption time. The results indicate that ECC can respond faster for decryption as compared with the RSA technique.

Table 2. Encryption, decryption, and total time for 8 bits (in seconds).

Security Bit Level	Encryption Time		Decryption Time		Total Time	
	ECC	RSA	ECC	RSA	ECC	RSA
112	2.20	0.02	0.50	0.70	2.70	0.72
128	3.87	0.03	0.80	1.94	4.67	1.97
144	4.72	0.04	1.00	2.64	5.72	2.68

Similarly, Table 3 shows the encryption, decryption, and total time needed by the ECC and RSA techniques for 64-bit data input. The results indicate that the encryption time needed by the RSA approach is much less than that needed by the ECC techniques. However, the decryption time for the 112, 128, and 144 security bit levels is much shorter for the ECC algorithm. Consequently, the total time is also less for the ECC algorithm. Furthermore, the difference in time for ECC and RSA becomes wider as the security bit level is increased from 112 to 144. For example, the difference between ECC and RSA algorithms for the 112 security bit level is only 3.66 s, but this difference is increased to 49.2 s when the security bit level is increased to 144, indicating that RSA needs 49.2 s longer than ECC to process 64-bit data input.

Table 3. Encryption, decryption, and total time for 64 bits (in seconds).

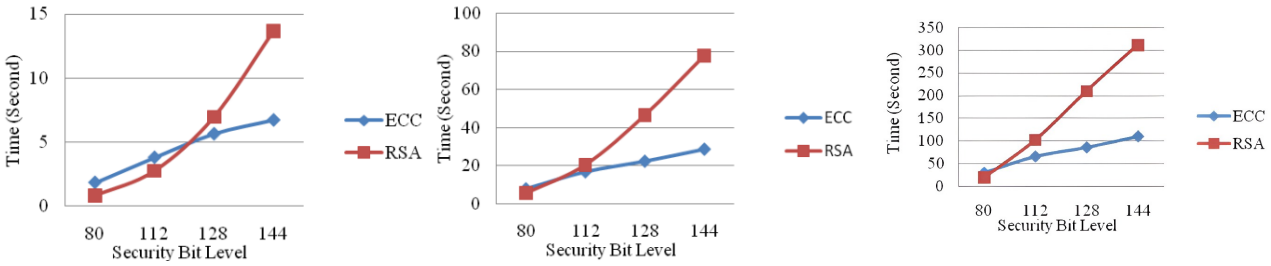
Security Bit Level	Encryption Time		Decryption Time		Total Time	
	ECC	RSA	ECC	RSA	ECC	RSA
112	9.98	0.16	6.93	20.41	16.91	20.57
128	15.08	0.16	7.35	46.47	22.44	46.64
144	20.23	0.13	8.47	77.76	28.70	77.90

Similar behavior is observed for 256-bit data input, as shown in Table 4. ECC is a more secure and efficient algorithm than RSA, providing robust protection with shorter key lengths. This results in lower network and computing power requirements, which improves the user experience. RSA can process 450 requests per second with an average response time of 150 milliseconds, whereas ECC can respond to the same number of requests per second in only 75 milliseconds.

Table 4. Encryption, decryption, and total time for 256 bits (in seconds).

Security Bit Level	Encryption Time		Decryption Time		Total Time	
	ECC	RSA	ECC	RSA	ECC	RSA
112	39.70	0.58	26.33	102.03	66.03	102.61
128	58.43	0.56	27.40	209.60	85.84	210.16
144	77.50	0.57	32.15	311.06	109.65	311.63

The experiments demonstrated that RSA is fast at encryption but slow at decryption, while ECC is slow at encryption but fast at decryption. Based on the results, it can be concluded that in general, ECC is more efficient and secure compared with RSA. This has been demonstrated in Figure 6.



(**a**) (**b**) (**c**)

Figure 6. Execution time (in sec) for ECC and RSA, (**a**) total time for 8-bit input, (**b**) total time for 64-bit input, and (**c**) total time for 256-bit input.

5.3. Discussion

The objective of the experiment was to measure the time required for encryption and decryption using RSA and ECC on three distinct input patterns of 8, 64, and 256 bits with randomly generated keys according to NIST guidelines. ECC outperformed RSA in terms of operational efficiency and security with fewer parameters, as demonstrated by the results. ECC is ideally suited for systems with limited resources.

ECC is significantly more secure than RSA for current key sizes, and this is its primary advantage. A 2048-bit RSA key is 10,000 times weaker than a 256-bit standard ECC key, which corresponds to a 3072-bit RSA key. The length of RSA keys must be increased to outperform an adversary's processing ability. ECC is also faster due to a variety of factors. First, smaller keys have the advantage of reducing the amount of data that need to be transmitted from the server to the client during an SSL handshake. In addition, ECC uses significantly less CPU and memory, resulting in dramatically improved response times and throughput on Web servers when it is active. Perfect forward secrecy (PFS) is a third significant advantage of using ECC. PFS is not an advantage of ECC, but the cypher suites supported by current Web servers and browsers that enforce PFS also enforce ECC. Web servers that favor ECDHE utilize cypher suites that provide both PFS and ECC benefits.

6. Conclusions and Future work

Effective data management is the key element for large-scale data analysis. Despite the existing solutions for big data life cycle management, no standard solution is available and security is a primary concern for such solutions. In this study, we proposed the use of SDMLC as a model for the big data life cycle. The main characteristics of SDMLC include the evaluation of 15 data life cycle models over the past 20 years that form the basis of the data life cycle. Additionally, complying with all of the suggested stages is not required. The life cycle stays applicable and significant in diverse government fields which include education, transportation, health, and production at the same time as it covers distinct varieties of data, which include open government data, commercial enterprise

data, scientific and research facts, citizen data, etc. A hybrid approach for data security is proposed and comprises the ECC and RSA algorithms. The experimental results indicate that ECC can provide better security with reduced execution time and is most suitable for devices with limited computational processing.

Author Contributions: Conceptualization, R.Z. and A.A.; Data curation, A.A. and T.A.; Formal analysis, R.Z. and T.A.; Funding acquisition, M.A.L.F.; Investigation, T.A. and F.I.; Methodology, F.I. and Y.A.M.V.; Project administration, Y.A.M.V.; Resources, Y.A.M.V. and M.A.L.F.; Software, M.A.L.F.; Supervision, I.A.; Validation, I.A.; Visualization, F.I.; Writing—original draft, R.Z. and A.A.; Writing—review and editing, I.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the European University of Atlantic.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Pathak, A.R.; Pandey, M.; Rautaray, S. Construing the Big Data Based on Taxonomy, Analytics and Approaches. *Iran. J. Comput. Sci.* **2018**, *1*, 237–259. [CrossRef]
2. Höchtl, J.; Parycek, P.; Schöllhammer, R. Big data in the policy cycle: Policy decision making in the digital era. *J. Organ. Comput. Electron. Commer.* **2016**, *26*, 147–169. [CrossRef]
3. Nobubele, A.; Mtsweni, J. *Big Data Privacy and Security: A Systematic Analysis of Current and Future Challenges*; University of South Africa: Pretoria, South Africa, 2016.
4. Khalouf, H.; Abouelmehdi, K.; Beni-hssane, A.; Saadi, M. Security Model for Big Healthcare Data Lifecycle. *Procedia Comput. Sci.* **2018**, *141*, 294–301. [CrossRef]
5. Immonen, A.; Kalaoja, J. Requirements of an energy data ecosystem. *IEEE Access* **2019**, *7*, 111692–111708. [CrossRef]
6. Lukoianova, T.; Rubin, V.L. Veracity roadmap: Is big data objective, truthful and credible? *Adv. Classif. Res. Online* **2014**, *24*, 4. [CrossRef]
7. Faroukhi, A.; El Alaoui, I.; Gahi, Y.; Amine, A. Big Data Monetization throughout Big Data Value Chain: A Comprehensive Review. *J. Big Data* **2020**, *7*, 3. [CrossRef]
8. Becker, M.J. The consumer data revolution: The reshaping of industry competition and a new perspective on privacy. *J. Direct Data Digit. Mark. Pract.* **2014**, *15*, 213–218. [CrossRef]
9. Allard, S. DataONE: Facilitating eScience through Collaboration. *J. EScience Libr.* **2012**, *1*, 4–17. [CrossRef]
10. Arass, M.E.; Tikito, I.; Souissi, N. Data lifecycles analysis: Towards intelligent cycle. In Proceedings of the 2017 Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 17–19 April 2017; pp. 1–8. [CrossRef]
11. Man, Z.; Li, J.; Di, X.; Sheng, Y.; Liu, Z. Double image encryption algorithm based on neural network and chaos. *Chaos Solitons Fractals* **2021**, *152*, 111318. [CrossRef]
12. Man, Z.; Li, J.; Di, X. Medical image encryption scheme based on self-verification matrix. *IET Image Process.* **2021**, *15*, 2787–2798. [CrossRef]
13. Group DSR; Structural Reform Group. Overview of the DDI Version 3.0 Conceptual Model. 2004. Available online: http://opendatafoundation.org/ddi/srg/Papers/DDIModel_v_4.pdf (accessed on 21 May 2023).
14. Michener, W.K.; Jones, M.B. Ecoinformatics: Supporting ecology as a data-intensive science. *Trends Ecol. Evol.* **2012**, *27*, 85–93. [CrossRef]
15. Demchenko, Y.; de Laat, C.; Membrey, P. Defining architecture components of the Big Data Ecosystem. In Proceedings of the 2014 International Conference on Collaboration Technologies and Systems (CTS), Minneapolis, MN, USA, 19–23 May 2014; pp. 104–112. [CrossRef]
16. Dawes, S.S.; Vidiasova, L.; Parkhimovich, O. Planning and designing open government data programs: An ecosystem approach. *Gov. Inf. Q.* **2016**, *33*, 15–27. [CrossRef]
17. Magalhaes, G.; Roseira, C.; Manley, L. Business models for open government data. In Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance, Guimaraes, Portugal, 27–30 October 2014; pp. 365–370. [CrossRef]
18. Arass, M.E.; Tikito, I.; Souissi, N. An audit framework for data lifecycles in a big data context. In Proceedings of the 2018 International Conference on Selected Topics in Mobile and Wireless Networking (MoWNeT), Tianjin, China, 15–16 August 2018. [CrossRef]
19. Shameli-Sendi, A. An efficient security data-driven approach for implementing risk assessment. *J. Inf. Secur. Appl.* **2020**, *54*, 102593. [CrossRef]
20. IBM. *Wrangling Big Data: Fundamentals of Data Lifecycle Management*; IBM Managing Data Lifecycle: Armonk, NY, USA, 2013.

21. Lmekki, H.; Chiadmi, D.; Lamharhar, H. Open Government Data. In Proceedings of the ArabWIC 6th Annual International Conference Research Track on—ArabWIC, Rabat, Morocco, 7–9 March 2019; pp. 1–6.
22. Raszewski, R.; Goben, A.H.; Bergren, M.D.; Jones, K.; Ryan, C.; Stefen, A.D.; Vonderheid, S.C. A survey of current practices in data management education in nursing doctoral programs. *J. Prof. Nurs.* **2021**, *37*, 155–162. [CrossRef]
23. Heimstädt, M.; Saunderson, F.; Heath, T. Conceptualizing Open Data Ecosystems: A timeline analysis of Open Data development in the UK. In Proceedings of the International Conference for E-Democracy and Open Government (CeDEM204), Krems, Austria, 21–23 May 2014; pp. 1–11.
24. NIST Big Data Public Working Group; Subgroup. *NIST Big Data Interoperability Framework: Volume 2, Big Data Taxonomies*; Technical Report; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2015. [CrossRef]
25. Jetten, M.; Simons, E.; Rijnders, J. The role of CRIS's in the research life cycle. A case study on implementing a FAIR RDM policy at Radboud University, the Netherlands. *Proc. Comput. Sci.* **2019**, *146*, 156–165. [CrossRef]
26. Rahul, K.; Banyal, R. Data life cycle management in big data analytics. *Proc. Comput. Sci.* **2020**, *173*, 364–371. [CrossRef]
27. McKeever, S. Understanding web content management systems: Evolution, lifecycle and market. *Ind. Manage. Data Syst.* **2003**, *103*, 686–692. [CrossRef]
28. Ball, A. Review of Data Management Lifecycle Models. 2012. Available online: http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf (accessed on 18 May 2023).
29. El Arass, M.; Souissi, N. Data Lifecycle: From Big Data to SmartData. In Proceedings of the 2018 IEEE 5th International Congress on Information Science and Technology (CiSt), Marrakech, Morocco, 21–27 October 2018; pp. 80–87. [CrossRef]
30. Khan, N.; Yaqoob, I.; Hashem, I.A.T.; Inayat, Z.; Mahmoud Ali, W.K.; Alam, M.; Shiraz, M.; Gani, A. Big data: Survey, technologies, opportunities, and challenges. *Sci. World J.* **2014**, *2014*, 1–18. [CrossRef]
31. Soltani Panah, A.; Yavari, A.; van Schyndel, R.; Georgakopoulos, D.; Yi, X. Context-driven granular disclosure control for internet of things applications. *IEEE Trans. Big Data* **2019**, *5*, 408–422. [CrossRef]
32. Freund, G.P.; Fagundes, P.B.; de Macedo, D.D.J. An Analysis of Blockchain and GDPR under the Data Lifecycle Perspective. *Mob. Netw. Appl.* **2021**, *26*, 266–276. [CrossRef]
33. Wikipedia. Cloud Computing. 2017. Available online: https://en.wikipedia.org/wiki/Cloud_computing (accessed on 10 April 2023).
34. LaChapelle, C. The Cost of Data Storage and Management: Where Is It Headed in 2016. *Data Cent. J.* **2016**. Available online: https://www.datacenterjournal.com/cost-data-storage-management-headed-2016/ (accessed on 10 April 2017).
35. Lei, H.; Xing, T.; Taylor, J.D.; Zhou, X. Monitoring Travel Time Reliability from the Cloud. *Transp. Res. Rec. J. Transp. Res. Board* **2012**, *2291*, 35–43. [CrossRef]
36. Miller, M. *Telephone Discussion with Stephen Lockwood*; PB Consult: Nurnberg, Germany, 2015.
37. Chen, M.C.; Chen, J.L.; Chang, T.W. Android/OSGi-based Vehicular Network Management System. *Comput. Commun.* **2011**, *34*, 169–183. [CrossRef]
38. EMC, D. Dell EMC Glossary: Data Archiving. Available online: https://www.emc.com/corporate/glossary/data-archiving.htm (accessed on 4 April 2017).
39. Olmstead, K.; Smith, A. *Americans and Cybersecurity*; Pew Research Center: Washington, DC, USA, 2017.
40. Hibbard, E.A.; Austin, R. *Storage Security Professional's Guide to Skills and Knowledge*; SNIA: Santa Clara, CA, USA, 2008.
41. ComputerWeekly.com. Data Storage Security: What It Is and the Key Components of a Storage Security Strategy. 2010. Available online: http://www.computerweekly.com/feature/Data-storage-security-What-it-is-and-the-key-components-of-a-storage-security-strategy (accessed on 17 May 2023).
42. Texas State Legislature. Unauthorized Use of Identifying Information. 2007. Available online: http://www.statutes.legis.state.tx.us/Docs/BC/htm/BC.521.htm (accessed on 17 May 2023).
43. Ahn, K.; Rakha, H.; Hill, D. *Data Quality White Paper*; Technical Report FHWA-HOP-08-038; US Department of Transportation, Federal Highway Administration: Washington, DC, USA, 2008.
44. Mahalle, V.S.; Shahade, A.K. Enhancing the data security in Cloud by implementing hybrid (Rsa & Aes) encryption algorithm. In Proceedings of the 2014 International Conference on Power, Automation and Communication (INPAC), Maharashtra, India, 6–8 October 2014; pp. 146–149.