



Article Driving Excellence in Official Statistics: Unleashing the Potential of Comprehensive Digital Data Governance

Hossein Hassani^{1,*} and Steve MacFeely²

- ¹ Research Institute of Energy Management and Planning (RIEMP), University of Tehran, Tehran 1417466191, Iran
- ² Data and Analytics Division, World Health Organization, 1201 Geneva, Switzerland; macfeelys@who.int
- Correspondence: hassani.stat@gmail.com

Abstract: With the ubiquitous use of digital technologies and the consequent data deluge, official statistics faces new challenges and opportunities. In this context, strengthening official statistics through effective data governance will be crucial to ensure reliability, quality, and access to data. This paper presents a comprehensive framework for digital data governance for official statistics, addressing key components, such as data collection and management, processing and analysis, data sharing and dissemination, as well as privacy and ethical considerations. The framework integrates principles of data governance into digital statistical processes, enabling statistical organizations to navigate the complexities of the digital environment. Drawing on case studies and best practices, the paper highlights successful implementations of digital data governance in official statistics. The paper concludes by discussing future trends and directions, including emerging technologies and opportunities for advancing digital data governance.

Keywords: data governance; digital; official statistics; framework; privacy; ethics; data quality; international organizations; advanced technology; IoT; big data



Citation: Hassani, H.; MacFeely, S. Driving Excellence in Official Statistics: Unleashing the Potential of Comprehensive Digital Data Governance. *Big Data Cogn. Comput.* 2023, 7, 134. https://doi.org/ 10.3390/bdcc7030134

Academic Editor: Carson K. Leung

Received: 7 June 2023 Revised: 13 July 2023 Accepted: 21 July 2023 Published: 29 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

1. Introduction

Official statistics play a critical role in providing accurate, reliable, and objective information for decision making, policy formulation, and monitoring socio-economic trends (see, for example, [1–10]). In the digital era, the availability and accessibility of vast amounts of data present both opportunities and challenges for official statistics. The proliferation of digital technologies, such as the internet, mobile devices, and social media platforms, has resulted in an exponential increase in the production and availability of data. This digital transformation has brought about significant changes in the way data are collected, processed, analyzed, and disseminated.

Figure 1 illustrates the amount of data generated annually from 2016 to 2025. Each bar corresponds to a specific year, and the height of the bar represents the volume of data generated in zettabytes. The chart showcases the significant growth in data generation over recent years. It is evident that the amount of data generated annually has been consistently increasing since 2016. In fact, it is remarkable to note that approximately 90% of the world's data has been generated in just the last two years. By 2023, data generation reached 120 zettabytes. The chart also indicates a projected increase of over 180% by 2025, with an expected volume of 218 zettabytes.

These statistics emphasize the rapid expansion of data production and highlight the exponential growth rate witnessed in recent years. It showcases the ever-increasing scale of the digital universe and the ongoing significance of data in our modern world.

The significance of official statistics in the digital era lies in their ability to provide authoritative and trustworthy information amidst the abundance of data sources and information sources of varying quality and reliability. Official statistics serve as a benchmark for measuring economic performance, social indicators, demographic trends, and environmental factors. They are essential for evidence-informed decision making, policy evaluation, and monitoring progress towards national and international development goals, including the UN's sustainable development goals (SDGs).



Data Produced from 2016 to 2025

Figure 1. Annual data generation trend from 2016 to 2025 (forecasted values: 2023-2025).

However, the digital era also poses challenges for official statistics. The exponential growth in digital data sources, such as administrative records, sensor data, and online platforms, requires the adaption of methodologies, infrastructure, and good data governance practices to ensure data quality, privacy protection, and data integration. Additionally, the digital era brings opportunities for leveraging innovative technologies, such as big data analytics, machine learning, and artificial intelligence to enhance the accuracy, timeliness, and granularity of official statistics.

Understanding the background and significance of data governance for official statistics in the digital era is crucial for statistical organizations, policymakers, researchers, and other stakeholders involved in data-driven decision making. This emphasizes the need to embrace digital transformation, adopt robust data governance frameworks, and leverage emerging technologies to harness the potential of the vast amount of digital data available while upholding the core principles of official statistics—relevance, impartiality, reliability, and confidentiality.

The importance of data governance with official statistics has become increasingly evident in recent years, as organizations across national statistical systems recognize the critical role it plays in ensuring the effective and responsible management of data (see, for instance, [11–15]). The significance of data governance is reflected in the worldwide search volume on the topic, as demonstrated by the analysis of the search volume data over time. Figure 2 depicts the trend in search volume related to data governance, revealing a noticeable peak in the recent months of 2023. This surge highlights the increasing attention and recognition given to data governance as a crucial issue in today's digital landscape.

While data governance is gaining more attention, there is currently no worldwide figure available to comprehensively evaluate the current state of this vital issue and project where it might lead. Here, we have developed an estimate, Figure 3, for worldwide data governance based on factors such as data quality, data availability, timeliness, and data protection laws. By considering these different aspects of data governance, we aimed to provide an initial approximation. Figure 3 highlights the amount of work to be done conceptually and definitionally on data governance.



Figure 2. Data governance worldwide search volume. (Data source: Google Trends).



Figure 3. Data governance worldwide map.

It is important to note that this estimate is a work in progress, and further components need to be incorporated to make it more accurate and comprehensive. Nonetheless, it serves as a first step until we have a more concrete and detailed understanding of data governance on a global scale.

The increased attention to data governance is driven by several factors. First, the exponential growth in data, both in volume and variety, necessitates comprehensive data governance to enable organizations to harness the value and insights that data can offer. Second, regulatory frameworks, for example guideline on data confidentiality, statistical laws adopted by national or international organizations long before the recent General Data Protection Regulation (GDPR), emphasize the need for organizations to implement robust data governance practices to ensure compliance and protect individuals' privacy rights.

Organizations are also becoming more aware of the potential risks associated with data mismanagement, such as data breaches, privacy violations, and inaccurate decision making. Consequently, they are actively seeking guidance and knowledge on how to establish effective data governance frameworks that align with their specific industry, regulatory requirements, and organizational goals.

Table 1 provides an overview of important data governance milestones and references. Categorized by initiator, summary of subject, and year, this table showcases the involvement of diverse organizations and entities in shaping global data governance practices.

The milestones in Table 1 highlight key initiatives and positions taken by influential entities in the field of data governance. These milestones serve as significant reference points in understanding the recent evolution and development of data governance frameworks and policies.

The table covers a range of topics and perspectives related to data governance. It includes endorsements of position papers, calls for global consensus on data, emphasis on regulatory frameworks and interoperability, recognition of the importance of human rights and safeguards in data storage, and the need for inclusive and equitable distribution of digital data benefits.

Initiator	Summary of Subject	Year	
CEB	Endorsed position paper "International Data Governance—Pathways to Progress"		
HLCP	Approved the position paper "International Data Governance—Pathways to Progress"		
World Bank	Call for a new social contract for data and a global consensus on data	2021	
OECD	Regulatory data frameworks and the need for technical and legal interoperability	2022	
G7 leaders	Identification of "future frontiers" related to data governance and endorsement of a roadmap for cooperation		
G20	Review of AI strategies on data governance and publication of policy briefs	2019	
African Union	Data policy framework and data governance	2022	
UNCTAD	Call for multilateral consensus on digital and data governance	2021	
CCSA	Call for a new global consensus on data	2021	
African Union	Comprehensive data policy framework with dedicated chapter on data governance	2022	
UN	Consultation for proposed UN global digital compact highlighting the link between digital and data governance	2022	
Swiss Internet Governance Forum	Need for a code of conduct for dataspace operators	2022	
Amnesty International	Role of legal frameworks and safeguards in protecting human rights in data storage and manipulation	2023	
G77	Importance of inclusive global data governance and equitable distribution of digital data benefits	2023	
European Union	Regulation on European data governance, voluntary framework for data collection and processing for altruistic purposes	2022	
OECD Working Party on Data Governance and Privacy	Report on data free flow with trust and policy priorities for cross-border data transfers	2023	
HLAB	Report on the need for digital and data governance, recommendation for a global data compact	2023	
WHO	Various reports, summit, and calling for the institutionalizing of good data governance practices along all stages of the health data value chain	2021	
Google	Google submission stressed the importance of data portability as the key to innovation, competition, and data protection		

Table 1. A selection of data governance milestones/references by content and year.

Notable contributors to the data governance landscape include international organizations such as CEB, HLCP, World Bank, WHO, OECD, G7, G20, G77, African Union, UN, UNCTAD, CCSA, European Union, Amnesty International, and Google among others (for more information, see, for example [16–38]). These organizations have actively participated in shaping discussions, proposing frameworks, and calling for collaborative efforts to address the challenges and opportunities associated with data governance.

By including references from different years, the table reflects the ongoing nature of data governance discussions and the continuous efforts to adapt to evolving technological advancements and societal needs. The range of years also demonstrates the increasing global focus on data governance as a critical issue for the digital age.

This paper also highlights selected successful implementations of data governance for official statistics. These cases indicate the importance of effective data governance frameworks and practices in ensuring the accuracy, reliability, and integrity of statistical data. The examples include Statistics Canada's comprehensive data governance framework [39], Eurostat's quality assurance framework [40], the United States Census Bureau's data stewardship program [41], and the Australian Bureau of Statistics' Data Integration Partnership for Australia [42]. Additionally, the cases showcase Statistics New Zealand's Integrated Data Infrastructure [43], INSEE's data governance framework in France [44], the Central Statistics Office of Ireland's robust data governance practices [45], and Statistics Korea's data governance in enhancing official statistics, providing valuable insights for researchers, policymakers, and practitioners in the field.

Furthermore, organizations recognize that effective data governance is not solely a matter of compliance but also a strategic imperative. Properly governed data enables organizations to make informed decisions, enhance operational efficiency, identify new business opportunities, and improve customer experiences.

Accordingly, this paper encompasses a comprehensive evaluation of the existing framework for digital data governance in official statistics. It considers the entire data lifecycle, from collection to dissemination, and places a specific emphasis on the challenges and opportunities presented by the digital era. The paper also aims to contribute to the ongoing discussions and research surrounding data governance in official statistics, providing valuable insights and practical recommendations for statistical organizations navigating the digital landscape.

Official statistics, as state statistics, navigate a delicate path intertwined with various political considerations. One such critical issue is data sovereignty, which raises sensitivity regarding the utilization of unofficial data sources for international official statistics. The debates held at the United Nations Statistical Commission (UNSC) around 2016 and 2017 clearly demonstrate the level of caution required when addressing this matter.

This paper can be considered a comprehensive resource providing a holistic overview of data governance for official statistics. It encompasses past efforts, current challenges, and recommendations for the future, particularly in the context of handling digital data. Throughout the paper, highlighted boxes have been strategically included to provide key insights and help readers to grasp the main concepts. As a result, this paper has multiple aims, including providing a thorough understanding of data governance in official statistics and offering practical guidance for navigating the evolving landscape of digital data.

2. Official Statistics in the Digital Era

The digital era has brought about significant transformations in the field of official statistics. Official statistics refer to data collected, compiled, and analyzed by national statistical agencies or other authorized entities to provide accurate and reliable information about various aspects of society, such as population, economy, social indicators, and the environment. These statistics play a crucial role in informing evidence-based decision making, policy formulation, and monitoring progress towards sustainable development goals.

In the digital era, official statistics have witnessed a paradigm shift in terms of data sources, collection methods, processing techniques, and dissemination practices. The proliferation of digital technologies, the advent of big data, and the increasing availability of digital platforms have the potential to revolutionize the way official statistics are produced and utilized.

One of the key aspects of digital data is the utilization of diverse data sources. Traditional surveys and censuses are now supplemented with administrative data, data from social media platforms, sensor-generated data, and other digital sources. This expansion of data sources enables users to capture real-time information, enhance the timeliness and granularity of statistics, and provide more accurate and detailed insights into societal trends and phenomena.

Table 2 presents an overview of data generation by category, highlighting the volume and percentage of data generated in 2023 and projected for 2025. The data landscape is witnessing exponential growth across various sectors, fuelled by advancements in digital technologies. The volume of data is expected to soar from 120 zettabytes in 2023 to 218 zettabytes in 2025, representing a remarkable 82% growth rate. This surge in data production is driven by diverse sources, including social media, internet usage, the internet of things, business data, scientific data, government data, sensor data, personal data, healthcare data, and other categories.

Category	Volume (2023)	Percentage (2023)	Volume (2025)	Percentage (2025)	Growth Rate
Social Media	30 ZB	25%	54 ZB	25%	80%
Internet Usage	25 ZB	21%	38 ZB	17%	52%
Internet of Things	20 ZB	17%	34 ZB	16%	70%
Business Data	15 ZB	13%	25 ZB	11%	67%
Scientific Data	9 ZB	8%	16 ZB	7%	78%
Government Data	8 ZB	7%	13 ZB	6%	63%
Sensor Data	4 ZB	3%	7 ZB	3%	75%
Personal Data	3 ZB	3%	6 ZB	3%	100%
Healthcare Data	2 ZB	2%	4 ZB	2%	100%
Other Categories	10 ZB	8%	18 ZB	8%	80%
Total	120 ZB	100%	218 ZB	100%	82%

Table 2. Data creation by category in 2023 and 2025 (forecast).

Such extensive data generation necessitates robust data governance frameworks to ensure responsible and ethical handling of information. With the rapid expansion of data, organizations and governments face the challenge of implementing effective strategies for data management, privacy protection, and security. Moreover, the increasing availability and accessibility of data offer immense opportunities for data-driven decision making, innovation, and advancements in various sectors. Various stakeholders, including those in healthcare, business, science, and government have the opportunity to leverage the expanding data landscape to drive valuable insights, optimize operations, enhance services, and tackle intricate challenges. However, it also underscores the need for appropriate data governance practices, including data protection and privacy, and responsible data use to ensure the integrity, trustworthiness, and ethical handling of data in the digital era.

Furthermore, the digital era has necessitated the adoption of advanced data processing techniques and analytical tools. Statistical agencies are leveraging data analytics, machine learning, and artificial intelligence to handle large and complex datasets, derive meaningful insights, and detect patterns and trends. These technological advancements enable more efficient data processing, improved data quality, and enhanced statistical outputs.

In addition to data collection and processing, the digital era has the potential to revolutionize the dissemination of official statistics. Statistical agencies now employ interactive online platforms, data portals, and data visualization tools to make statistics more accessible, user-friendly, and engaging for policymakers, researchers, businesses, and the general public. This digital transformation in dissemination practices promotes data transparency, enables data exploration, and facilitates evidence-informed decision making.

The digital era has transformed official statistics, offering new opportunities and challenges. Diverse data sources, including social media, IoT, and business data, have the potential to revolutionize data collection, enhancing the accuracy and granularity of statistics. The exponential growth in data production calls for robust data governance frameworks to ensure responsible handling of information. Advanced data processing techniques and interactive dissemination platforms further enhance the value and accessibility of official statistics.

3. Data Quality

A critical aspect that warrants attention within the realm of data governance is data quality. Ensuring high-quality data is essential throughout the entire data lifecycle, as defined by the Generic Statistical Business Process Model (GSBPM). This model outlines the various stages involved in statistical production, from data collection and processing to analysis and dissemination.

To maintain data governance practices consistently, many countries have developed national quality assurance frameworks. These frameworks serve as comprehensive guidelines for assessing, monitoring, and improving data quality within a country's statistical system. They provide a structured approach to measure the accuracy, reliability, timeliness, and relevance of data, ensuring its fitness for specific purposes and promoting confidence in the resulting statistical outputs.

At the international level, the United Nations has established the Statistical Quality Assurance Framework (SQAF) to address data quality issues in statistical activities conducted by international organizations. The SQAF provides a common framework and methodology for assessing and enhancing data quality across different entities. However, given the evolving data landscape, there is a growing recognition that the SQAF may need to be updated to incorporate new dimensions arising from the use of big data, citizen science data, and geospatial data [47].

The emergence of big data, which refers to large and complex datasets, along with the increasing involvement of citizens in data collection through citizen science initiatives, and the widespread use of geospatial data, present unique challenges and opportunities for data governance. These new data sources and technologies require additional considerations and standards to ensure their appropriate integration into the existing data quality frameworks. This includes addressing issues related to data privacy, data accuracy, data interoperability, and ensuring ethical and responsible use of these data sources.

Updating the UN SQAF to incorporate these new dimensions would enable international organizations to effectively leverage the potential of these emerging data sources while maintaining the integrity and reliability of statistical outputs. It would also foster harmonization and consistency in data quality practices across different domains and promote a more inclusive and comprehensive approach to data governance in the era of rapid technological advancements.

Figure 4 represents the Google search volume index for the keyword "data quality" on a monthly basis from January 2004 to May 2023. Analyzing the data, we can observe fluctuations in the search volume index over time. In the early years (2004–2006), the index shows varying levels of interest in the topic, with occasional spikes and drops. In recent years, from 2020 to 2023, the search volume index remains relatively stable, indicating a sustained level of interest in the topic of data quality. This suggests that data quality continues to be a relevant and sought-after subject among data users.



Figure 4. Search volume index for data quality (2004–2023). (Data source: Google Trends).

Data quality is a critical aspect of data governance, and comprehensive frameworks are in place to ensure its maintenance throughout the data lifecycle. National quality assurance frameworks provide structured guidelines to assess and improve data quality within statistical systems. Internationally, the UN's Statistical Quality Assurance Framework (SQAF) plays a vital role in addressing data quality issues. However, the evolving data landscape calls for updates to incorporate new dimensions from big data, citizen science data, and geospatial data. This would enable effective utilization of these emerging data sources while ensuring integrity and reliability in statistical outputs.

4. Digital Data Governance in Official Statistics

Digital data governance refers to the frameworks, principles, and practices employed to ensure the effective and responsible management of digital data in the context of official statistical activities. It encompasses a range of processes, policies, and mechanisms that govern the collection, processing, storage, sharing, and dissemination of digital data to uphold data quality, integrity, and privacy.

It is important to highlight the significance of the UN's Fundamental Principles of Official Statistics. These principles serve as the foundation and guiding framework for official statisticians, holding great importance within the official statistics community. They provide essential guidelines and standards for the collection, production, and dissemination of official statistics.

In addition to the UN's fundamental principles, there are also the principles governing international statistical activities that apply to international organizations. These principles serve as a framework for coordination and collaboration among statistical agencies and organizations at the international level. They ensure consistency, comparability, and quality of statistical data produced by different entities, facilitating effective data sharing and analysis for global decision-making processes.

Together, these frameworks, including the UN's fundamental principles of official statistics and the principles governing international statistical activities, contribute to the establishment of robust data governance practices and promote the ethical and reliable use of data at both national and international levels.

However, in 2022, the mandate of the United Nations Statistical Commission (UNSC) underwent a significant change from focusing solely on statistics to encompassing both data and statistics. This expansion of the mandate provides a political rationale to discuss a broader set of principles related to data governance and management. This shift acknowledges the increasing recognition of the vital role that data plays in informing policy decisions and shaping global development agendas.

4.1. Understanding Digital Data Governance in the Context of Official Statistics

Digital data governance in official statistics is a critical aspect of managing the complexities associated with handling digital data. Unlike traditional data sources, digital data possesses distinct characteristics that require specialized governance approaches. One key characteristic is the vast volume of digital data generated and collected on a daily basis. The exponential growth in digital data presents both opportunities and challenges for official statistical agencies. On one hand, it enables the availability of massive amounts of data that can potentially enhance the accuracy and granularity of statistical analyses. On the other hand, managing and processing such large volumes of data necessitates robust governance practices to ensure efficient and effective data handling.

The velocity of digital data adds another layer of complexity. Digital data are often generated, updated in real-time or near real-time, requiring statistical agencies to develop agile data governance processes to keep up with the pace of data production. This necessitates the adoption of automated data collection, processing, and analysis techniques to enable timely and up-to-date statistical outputs. Additionally, the variety of digital data sources and formats poses a challenge for data integration and harmonization. The users must develop mechanisms to aggregate, standardize, and reconcile data from diverse sources, including social media platforms, web scraping, administrative records, and sensor networks. This requires careful consideration of data integrability, metadata management, and data integration frameworks [48].

Furthermore, the veracity of digital data refers to its accuracy, reliability, and trustworthiness. As digital data can be generated by a multitude of sources, including individuals, organizations, and automated systems, ensuring data quality becomes paramount. Data governance practices need to address issues such as data validation, verification, and data provenance to maintain the credibility and trust associated with official statistical outputs. It is crucial to establish data quality frameworks and implement rigorous data validation processes to mitigate the risks of errors, biases, and misinformation that can arise from digital data sources.

While harnessing the power of digital technologies, official statistical agencies must also prioritize safeguarding the trust, credibility, and confidentiality associated with their outputs. Data governance practices need to adhere to strict privacy and confidentiality principles to protect individual privacy rights and sensitive information. This involves implementing robust data anonymization techniques, secure data storage systems, and stringent access controls to prevent unauthorized use or disclosure of data. Additionally, ensuring compliance with data protection regulations and ethical guidelines is essential for maintaining public trust and confidence in official statistical agencies.

Digital data governance in official statistics recognizes the unique characteristics of digital data, such as its volume, velocity, variety, and veracity. It involves the application of governance principles and practices to effectively navigate these complexities while harnessing the benefits of digital technologies. By implementing robust data governance frameworks, official statistical agencies can ensure the reliability, accuracy, and integrity of statistical outputs, while upholding the trust, credibility, and confidentiality associated with their work.

4.2. Key Principles and Components of Digital Data Governance

(a) Data Quality: Ensuring the accuracy, reliability, and relevance of digital data is a fundamental aspect of digital data governance in official statistics. Robust quality assurance processes are implemented to verify the integrity and validity of the data. These include conducting thorough data validation, performing checks for outliers and inconsistencies, and adhering to internationally recognized statistical standards and methodologies. By maintaining high standards of data quality, official statistical agencies can enhance the reliability and credibility of their statistical outputs, enabling informed decision making and policy formulation.

(b) Data Privacy and Confidentiality: Protecting individual privacy and maintaining the confidentiality of sensitive data is of the utmost importance in digital data governance. It is necessary for official statistical agencies to implement rigorous data anonymization techniques to remove personally identifiable information and ensure data are rendered anonymous. Secure storage systems and strict access controls are employed to prevent unauthorized access or disclosure of data. By prioritizing data privacy and confidentiality, official statistical agencies can foster public trust and confidence, encouraging data providers to share their information while safeguarding individuals' rights and sensitive information.

(c) Data Security: Data security plays a vital role in digital data governance. It is required for official statistical agencies to adopt robust security measures to protect digital data from unauthorized access, breaches, and cyber threats. This involves implementing encryption protocols to secure data transmission and storage, regularly updating security systems, conducting vulnerability assessments, and establishing disaster recovery mechanisms to mitigate risks. By prioritizing data security, official statistical agencies can ensure the integrity, availability, and confidentiality of digital data, safeguarding it from potential threats.

(d) Data Ethics: Ethical considerations are crucial in the governance of digital data and analysis. It is necessary for official statistical agencies to adhere to ethical guidelines and principles when handling digital data. This includes obtaining informed consent from data providers, ensuring transparency in data collection and usage, maintaining fairness in data representation and analysis, and being accountable for the responsible use of data. By upholding data ethics, official statistical agencies can promote trust, fairness, and integrity in their data governance practices.

(e) Data Integration and Interoperability: Digital data governance involves addressing the challenges of data integration and interoperability. It is crucial for official statistical agencies to develop mechanisms to seamlessly integrate and harmonize digital data from multiple sources, platforms, and formats. This requires the establishment of data integration frameworks, the development of data standardization processes, and the implementation of metadata management practices. By enabling data integration and interoperability, official statistical agencies can enhance the comprehensiveness and accuracy of their statistical analyses, enabling a more holistic understanding of complex phenomena [49].

(f) Stakeholder Engagement: Engaging relevant stakeholders is a crucial component of digital data governance in official statistics. It is vital for official statistical agencies to foster collaboration, dialogue, and partnerships with data providers, data users, and the public. By involving stakeholders in the governance processes, agencies can gain valuable insights, address emerging challenges, and ensure that data governance practices align with the needs and expectations of the wider community. Stakeholder engagement promotes transparency, inclusiveness, and accountability in the management and governance of digital data.

Data governance in official statistics involves ensuring accuracy, privacy, security, ethics, integration, and stakeholder engagement. It includes robust quality assurance, data anonymization, security measures, ethical considerations, data integration, and partnerships. These practices enhance data reliability, protect privacy, ensure security, promote fairness, and enable comprehensive analysis.

5. Framework for Digital Data Governance in Official Statistics

5.1. Designing a Comprehensive Framework for Digital Data Governance

In the digital era, the effective governance of data is paramount for official statistical agencies. To address the challenges and complexities associated with digital data, a comprehensive framework for digital data governance needs to be designed. This framework serves as a roadmap for guiding official statistical agencies in managing, protecting, and utilizing digital data in a responsible and effective manner.

The design of the framework involves identifying key principles, standards, and best practices that govern the collection, storage, processing, and dissemination of digital data. It encompasses various aspects, such as data quality, privacy, security, ethics, interoperability, and stakeholder engagement. The framework should be flexible and adaptable to accommodate the evolving nature of digital data and the changing needs of official statistical practices. Here is an example for digital data governance.

- 1. Governance Structure and Accountability:
 - Establish a dedicated governance body responsible for overseeing digital data governance initiatives.
 - Define clear roles, responsibilities, and accountability for data governance at various levels within the organization.
 - Ensure senior management commitment and support for data governance efforts.
- 2. Data Strategy and Policies:
 - Develop a data strategy that aligns with the organization's goals and objectives.
 - Define data governance policies that outline principles, standards, and guidelines for managing digital data.
 - Address legal, ethical, and regulatory considerations in data governance policies.
- 3. Data Quality Management:
 - Implement robust quality assurance processes to ensure the accuracy, reliability, and relevance of digital data.
 - Establish data quality metrics and monitoring mechanisms to measure and improve data quality over time.
 - Integrate data validation, cleansing, and enrichment techniques into data processing workflows.

4.

- Data Privacy and Confidentiality:
- Develop and enforce policies and procedures to protect individual privacy and maintain the confidentiality of sensitive data.
- Implement rigorous data anonymization techniques to de-identify personally identifiable information.
- Establish secure data storage and access control mechanisms to prevent unauthorized disclosure.
- 5. Data Security and Cybersecurity:
 - Implement robust data security measures to protect digital data from unauthorized access, breaches, and cyber threats.
 - Apply encryption and authentication protocols to ensure secure data transmission and storage.
 - Conduct regular security assessments, vulnerability testing, and incident response planning.
- 6. Data Ethics and Compliance:
 - Embed ethical considerations into data governance practices, including informed consent, fairness, transparency, and accountability.
 - Ensure compliance with relevant data protection regulations, industry standards, and ethical guidelines.
 - Promote responsible data use and address emerging ethical challenges associated with digital data.
- 7. Data Integration and Interoperability:
 - Establish data integration frameworks and standards to enable seamless integration of digital data from multiple sources and formats.
 - Implement metadata management practices to facilitate data discovery, understanding, and interoperability.
 - Foster collaboration with external data providers to enhance data integration efforts.
- 8. Stakeholder Engagement and Communication:
 - Foster collaboration, dialogue, and partnerships with data providers, data users, and the public.
 - Engage stakeholders in the governance processes, solicit feedback, and address their concerns.
 - Communicate transparently about data governance initiatives, policies, and practices to build trust and confidence.
- 9. Monitoring, Evaluation, and Continuous Improvement:
 - Develop mechanisms for monitoring and evaluating the effectiveness of the data governance framework.
 - Regularly review and update data governance policies and practices to adapt to evolving needs and emerging technologies.
 - Foster a culture of continuous improvement and learning within the organization.

5.2. Integration of Data Governance Principles into Digital Statistical Processes

The integration of data governance principles into digital statistical processes is essential for ensuring the effectiveness and efficiency of data governance efforts. By incorporating data governance principles into every stage of the statistical process, official statistical agencies can establish a holistic and consistent approach to managing digital data.

One key benefit of integrating data governance principles is the standardization of data collection. By defining clear protocols, methodologies, and standards for data collection, agencies can ensure that data are collected in a uniform and reliable manner. This reduces the risk of errors, inconsistencies, and biases in the data, enhancing the accuracy and reliability of statistical outputs.

In addition, the integration of data governance principles enables the implementation of robust quality assurance processes. By incorporating data validation, cleansing, and enrichment techniques into digital statistical processes, agencies can identify and address data anomalies, outliers, and discrepancies. This helps to maintain data quality and integrity throughout the statistical process, ensuring that the final outputs are of high quality and fit for purpose.

Moreover, the integration of data governance principles promotes interoperability and data sharing across different agencies and organizations. By adopting common data standards, definitions, and formats, agencies can facilitate the exchange and integration of data from various sources. This enables a more comprehensive and holistic analysis of data, providing a richer understanding of complex phenomena and supporting evidence-based decision making.

The framework for digital data governance in official statistics involves the design of a comprehensive governance framework and the integration of data governance principles into digital statistical processes. This holistic approach ensures that official statistical agencies can effectively manage and utilize digital data while upholding data quality, privacy, security, ethics, and stakeholder engagement. By implementing such a framework, agencies can enhance the reliability, integrity, and trustworthiness of their statistical outputs in the digital era.

6. Metadata and Data Governance

Metadata plays a crucial role in data governance, as it provides valuable information about the structure, content, and context of digital data. It acts as a foundation for effective data management, discovery, understanding, and interoperability. In this section, we will explore the relationship between metadata and data governance and the importance of incorporating metadata management practices within the framework of data governance.

Figure 5 illustrates an informative data governance framework for national e-government. The diagram depicts data governance as a comprehensive and multifaceted approach that encompasses the establishment of policies and regulations, the development of leadership for institutional coordination and national strategy, the cultivation of an enabling data ecosystem, and the streamlining of data management processes.

As depicted by the four pillars in the figure, data governance relies on the dynamic interplay between policies, institutions, people, processes, and enabling technologies. Notably, metadata plays a crucial role within this framework, as it serves as a fundamental element without which the entire system would lack a critical component.

The visual representation emphasizes the holistic nature of data governance, highlighting the interconnectedness of its various dimensions. It reinforces the significance of adopting a comprehensive approach to effectively manage and govern data in the context of national e-government initiatives [50].



Figure 5. Metadata and data governance framework for e-government.

6.1. The Importance of Metadata in Comprehensive Data Governance

Metadata describes the characteristics, attributes, and properties of data, enabling users to understand its meaning, purpose, and relevance. Metadata facilitates data integration, quality assessment, data lineage, and data usage tracking. It supports data governance processes, such as data classification, data access control, and data lifecycle management. By incorporating metadata into data governance practices, organizations can achieve the following:

- 1. Data Understanding: Metadata helps users comprehend the meaning and context of data, allowing for effective data interpretation and analysis. It enables data consumers to understand the origin, accuracy, and limitations of data, leading to more informed decision making.
- Data Integration: Metadata provides insights into the structure, format, and relationships between different data sources, facilitating data integration and interoperability. It enables organizations to harmonize disparate datasets, align data definitions, and establish data mappings.
- 3. Data Quality Management: Metadata plays a vital role in data quality management by providing information about data lineage, data validation rules, and data transformations. It supports data profiling, data cleansing, and data enrichment processes, ensuring the accuracy, consistency, and completeness of data.
- 4. Data Security and Privacy: Metadata aids in data governance efforts related to data security and privacy. It helps identify sensitive data elements, data access controls, and data retention policies. Metadata also assists in tracking data usage and monitoring compliance with data protection regulations.

6.2. Metadata Management Practices within Data Governance

To effectively incorporate metadata into data governance, it is essential for organizations to implement robust metadata management practices. These practices involve:

- 1. Metadata Documentation: Organizations should document metadata attributes, definitions, and relationships in a structured and consistent manner. This documentation should cover data source information, data transformations, data usage policies, and data stewardship responsibilities.
- 2. Metadata Repository: Establishing a centralized metadata repository allows for the storage, organization, and retrieval of metadata. The repository should support metadata search, browsing, and version control. It should also facilitate metadata sharing and collaboration among stakeholders.
- 3. Metadata Standards and Governance: Organizations should define metadata standards and guidelines that align with industry best practices. These standards should address metadata formats, naming conventions, metadata modeling techniques, and metadata exchange protocols. Metadata governance frameworks can be established to ensure compliance with these standards.
- 4. Data Lineage and Impact Analysis: Metadata should capture and track the lineage of data, documenting its origin, transformation processes, and usage. Data impact analysis should be conducted to understand the implications of changes in data structures or data sources on downstream processes and stakeholders.
- 5. Data Cataloging and Discovery: Metadata should enable data cataloging and discovery, allowing users to search for and identify relevant data assets. This promotes data transparency and facilitates data sharing among authorized users.

Metadata and data governance are closely intertwined, with metadata serving as a critical enabler of effective data management and governance. Incorporating metadata management practices within the framework of data governance ensures that data assets are well-documented, understood, integrated, and secured. By prioritizing metadata within data governance initiatives, organizations can unlock the full potential of their digital data, enabling informed decision making, data interoperability, and organizational agility.

7. Implementing Digital Data Governance: Framework, Challenges and Best Practices

This section delves into the practical aspects of establishing and executing a digital data governance initiative. It outlines a comprehensive framework that encompasses the key elements necessary for successful implementation, including defining objectives, developing policies and frameworks, assigning roles, assessing data quality, ensuring privacy and security, promoting collaboration, monitoring compliance, and embracing emerging technologies. This section also addresses the challenges that organizations may encounter during the implementation process, such as resistance to change, resource constraints, and technical complexities. Moreover, it highlights best practices derived from successful implementations, providing valuable insights for organizations seeking to enhance their digital data governance practices. By following this section's guidance, organizations can navigate the complexities of implementing digital data governance and maximize the benefits of robust data management and governance practices.

7.1. Framework

- Establish Clear Objectives and Scope: Define the objectives and goals of digital data governance within your organization. Clearly outline the scope of data governance initiatives, including the types of data
- and systems covered.
 Develop Data Governance Policies and Frameworks: Create policies and frameworks that outline data governance principles, standards, and procedures. Define data classification, data ownership, data lifecycle management, and data access controls.
- Identify Data Stewards and Roles: Assign data stewards responsible for overseeing data governance processes. Clearly define the roles and responsibilities of data owners, data custodians, and data users.
- Assess Data Quality and Metadata Management: Implement processes for assessing and improving data quality. Establish metadata management practices to document data attributes, lineage, and definitions.
- Ensure Data Privacy and Security: Develop policies and procedures to protect sensitive data and ensure compliance with privacy regulations. Implement security measures such as encryption, access controls, and data masking.
- 6. Implement Data Integration and Interoperability: Establish data integration standards and practices to ensure seamless data flow across systems. Promote interoperability between different data sources and systems.
- Enable Stakeholder Engagement and Collaboration: Involve stakeholders in decision-making processes and encourage collaboration. Foster a data-driven culture and provide training and awareness programs for employees.
- Monitor and Audit Data Governance: Establish monitoring mechanisms to track compliance with data governance policies and standards. Conduct regular audits to assess the effectiveness of data governance processes and identify areas for improvement.
- 9. Embrace Emerging Technologies and Trends:

Stay updated with emerging technologies relevant to digital data governance such as AI, blockchain, and cloud computing. Adapt governance practices to incorporate new trends, such as responsible AI and privacy-enhancing technologies.

 Continuously Improve Data Governance: Regularly review and refine data governance policies and frameworks to adapt to evolving business needs. Seek feedback from stakeholders and incorporate lessons learned from successful implementations and challenges faced.

7.2. Overcoming Challenges in Implementing Digital Data Governance

Implementing digital data governance can be a complex and challenging endeavor. Organizations may encounter various obstacles during the implementation process. Some common challenges include:

- Lack of awareness and understanding: One of the primary challenges is the lack of awareness and understanding about the importance and benefits of digital data governance. It is crucial to educate stakeholders about the significance of data governance and its potential impact on the organization's overall performance.
- 2. Resistance to change: Resistance to change can hinder the implementation of data governance initiatives. People may be resistant to new processes, policies, or technologies, making it difficult to gain buy-in and cooperation. Effective change management strategies, including clear communication, training, and involving stakeholders in the decision-making process, can help overcome resistance.
- 3. Data complexity and variety: Dealing with the vast volume, variety, and velocity of digital data can be overwhelming. Organizations need to address the challenges associated with integrating and managing diverse data sources, formats, and structures. Implementing data integration technologies, metadata management systems, and data standardization practices can help overcome these complexities.
- 4. Data quality issues: Ensuring data accuracy, reliability, and consistency can be a significant challenge. Poor data quality can lead to erroneous insights and decisions. Establishing robust data quality management processes, including data validation, cleansing, and enrichment, can help address these issues.
- 5. Regulatory compliance: Organizations often face challenges in ensuring compliance with data protection and privacy regulations. It is important to navigate the complex landscape of data regulations and implement appropriate measures to protect sensitive data, such as implementing data anonymization techniques, obtaining necessary consents, and establishing secure data storage and transmission protocols.
- 6. Stakeholder alignment: Achieving alignment among various stakeholders can be a significant challenge in data governance implementation. Different departments or individuals may have conflicting priorities or different interpretations of data governance requirements. Engaging stakeholders early on, facilitating open communication, and involving them in the decision-making process can foster collaboration and alignment.
- 7. Resource constraints: Implementing effective data governance requires dedicated resources, including skilled personnel, technology infrastructure, and budgetary allocations. Limited resources can impede progress and hinder the successful implementation of data governance initiatives. Organizations should prioritize resource allocation and consider investing in training programs, hiring specialized talent, and adopting scalable technologies.
- 8. Change sustainment: Sustaining the changes and ensuring the longevity of data governance practices can be challenging over time. Without ongoing support and reinforcement, organizations may revert to old habits or face difficulties in adapting to evolving data governance requirements. It is crucial to establish governance frameworks, regular assessments, and continuous improvement processes to embed data governance into the organization's culture and ensure its long-term success.

7.3. Best Practices and Lessons Learned from Successful Implementations

To navigate the challenges and ensure a successful implementation of digital data governance, organizations can follow best practices and learn from successful implementations. Some key practices include:

- 1. Develop a clear vision and strategy: Start by defining a clear vision for data governance and align it with the organization's goals and objectives. Develop a comprehensive data strategy that outlines the desired outcomes, priorities, and roadmap for implementation.
- 2. Establish a dedicated governance structure: Create a dedicated governance body or team responsible for overseeing data governance initiatives. Define clear roles, responsibilities, and reporting lines to ensure accountability and effective decision making.
- 3. Engage stakeholders: Involve stakeholders from different levels and functions within the organization, including senior management, IT, data owners, and data users. Foster collaboration, communication, and shared ownership of data governance initiatives.
- 4. Adopt a phased approach: Implementing data governance in phases allows for incremental progress and minimizes disruption. Start with a pilot project or a specific data domain, demonstrate success, and then expand to other areas.
- 5. Ensure data privacy and security: Prioritize data privacy and security considerations throughout the implementation process. Implement appropriate measures to protect data, including encryption, access controls, and compliance with data protection regulations.
- Foster a data-driven culture: Promote a data-driven culture within the organization by encouraging data literacy, training, and awareness programs. Foster a mindset of data ownership, accountability, and continuous improvement.
- 7. Monitor and measure progress: Establish metrics and performance indicators to track the effectiveness of data governance initiatives. Regularly evaluate and assess the impact of data governance on data quality, decision making, and organizational performance.
- 8. Learn from others: Look to successful implementations and case studies from other organizations to learn from their experiences and best practices. Engage with industry experts, attend conferences, and participate in communities of practice to stay updated on the latest trends and insights in digital data governance.

7.4. Selected Successful Implementation of Data Governance for Official Statistics

The following examples highlight the successful implementation of data governance in official statistics organizations, demonstrating how robust frameworks and practices contribute to the production of reliable and trusted statistical data.

- 1. Statistics Canada's Data Governance Framework [39]: Statistics Canada, the national statistical agency of Canada, has implemented a comprehensive data governance framework. The framework includes clear policies, procedures, and guidelines for data management and governance. It defines roles and responsibilities, establishes data standards and quality controls, and ensures compliance with privacy regulations. The framework has helped Statistics Canada enhance data accuracy, reliability, and accessibility, thereby improving the quality of official statistics produced.
- 2. Eurostat's Quality Assurance Framework [40]: Eurostat, the statistical office of the European Union, has developed a quality assurance framework for official statistics. This framework encompasses data governance principles and practices aimed at ensuring the highest standards of quality for statistical data. It involves comprehensive quality assessments, adherence to data validation and verification processes, and continuous improvement measures. The framework has contributed to the production of reliable, comparable, and timely official statistics across the European Union.
- 3. United States Census Bureau's Data Stewardship Program [41]: The United States Census Bureau has implemented a robust data stewardship program to govern its official statistical data. The program focuses on data quality, security, privacy, and accessibility. It involves the establishment of data governance boards, data stewardship training programs, and the implementation of data management policies and

procedures. The Data Stewardship Program has played a critical role in maintaining the integrity and trustworthiness of the Census Bureau's official statistical data.

- 4. Australian Bureau of Statistics' Data Integration Partnership for Australia [42]: The Australian Bureau of Statistics (ABS) has established the Data Integration Partnership for Australia (DIPA) to govern the integration of multiple data sources for producing official statistics. DIPA operates under a robust data governance framework that ensures privacy protection, data security, and the ethical use of data. It facilitates data sharing and collaboration with partner organizations while maintaining strict controls and safeguards. The ABS's implementation of DIPA has resulted in more accurate and comprehensive official statistics.
- 5. Statistics New Zealand's Integrated Data Infrastructure (IDI) [43]: Statistics New Zealand has implemented the IDI, a world-leading data integration platform. The IDI securely combines and analyzes data from various government agencies to produce valuable insights for official statistics. The data governance framework ensures strict privacy controls, data quality checks, and compliance with legal and ethical standards. The IDI has enabled Statistics New Zealand to generate more accurate and timely official statistics while protecting individual privacy.
- 6. National Institute of Statistics and Economic Studies (INSEE) of France [44]: INSEE has developed a comprehensive data governance framework to ensure the quality and integrity of official statistics. The framework includes standardized procedures for data collection, validation, and dissemination. It emphasizes transparency and involves collaboration with data providers, statistical experts, and stakeholders. IN-SEE's data governance practices have contributed to the production of reliable and trusted official statistics in France.
- 7. Central Statistics Office (CSO) of Ireland [45]: The CSO has established a robust data governance framework to manage and govern official statistics. The framework incorporates data quality controls, privacy protection measures, and adherence to international statistical standards. It also promotes data sharing and collaboration with government agencies, researchers, and the public. The CSO's data governance practices have strengthened the credibility and usefulness of official statistics in Ireland.
- 8. Statistics Korea's Data Governance Program [46]: Statistics Korea has implemented a comprehensive data governance program to ensure the accuracy and reliability of official statistics. The program includes data quality assessments, data lineage tracking, and standardized data management practices. It also focuses on data integration and interoperability to streamline statistical processes. Statistics Korea's data governance program has played a crucial role in producing high-quality and consistent official statistics.

Implementing digital data governance poses challenges, including lack of awareness, resistance to change, data complexity, and data quality issues. Best practices include developing a clear vision, establishing a dedicated governance structure, engaging stakeholders, adopting a phased approach, prioritizing data privacy and security, fostering a data-driven culture, monitoring progress, and learning from others. By following these practices and learning from successful implementations, organizations can overcome challenges and ensure effective digital data governance.

8. UN Newly Approved Data Governance

The recent approval of data governance by the United Nations (UN) can be considered a game changer [11]. In May 2023, the United Nations System Chief Executives Board for Coordination (CEB) endorsed the position paper titled "International Data Governance —Pathways to Progress", which had received approval from the High-Level Committee on Programmes (HLCP) in March of the same year.

This endorsement by the UN, a globally recognized and influential organization, signifies a significant milestone in acknowledging the importance of data governance at an international level. It demonstrates that data governance has become a key concern on the

UN's agenda and reinforces its commitment to addressing the challenges and opportunities associated with data management.

The position paper itself represents the culmination of extensive efforts spanning over two years. It builds upon the World Bank's World Development Report: Data for Better Lives, released in 2021, which called for global data governance. Additionally, it takes into account various blogs, papers, and an interim report submitted to the HLCP in March 2022.

The UN's adoption of data governance reflects the growing recognition of the critical role that data plays in driving progress and development across various sectors. By endorsing this position paper, the UN acknowledges the need for coordinated and comprehensive approaches to data governance that can facilitate better data management, enhance data quality and privacy, promote data sharing, and ensure responsible and ethical use of data.

The UN's involvement in data governance brings considerable credibility and influence to the field. It has the potential to encourage governments, organizations, and stakeholders around the world to prioritize and invest in effective data governance frameworks. The UN's global platform can serve as a catalyst for collaboration and knowledge sharing, enabling countries to learn from one another's experiences and best practices in data governance implementation.

Furthermore, the UN's endorsement of data governance demonstrates its commitment to leveraging data as a valuable resource for sustainable development, humanitarian efforts, and evidence-based decision making. This move highlights the recognition that harnessing the power of data in a responsible and inclusive manner is essential for achieving the UN's sustainable development goals (SDGs).

The UN's recent adoption of data governance is a game changer because it signifies the growing importance of data governance at the international level and brings increased attention and momentum to this critical area. It sets the stage for global collaboration, knowledge sharing, and the development of comprehensive frameworks that can address the challenges and maximize the benefits of the data-driven era we are living in.

9. Future Trends and Directions in Digital Data Governance

9.1. Emerging Technologies and Their Impact on Digital Data Governance

The landscape of digital data governance is continuously evolving due to the emergence of new technologies. These technologies have a significant impact on how data are collected, managed, analyzed, and governed. Some key emerging technologies and their implications for digital data governance include:

- 1. Artificial Intelligence (AI) and Machine Learning (ML): AI and ML technologies are revolutionizing data governance by enabling advanced analytics, data processing, and decision making. These technologies can automate data quality checks, identify patterns and anomalies, and enhance data governance processes.
- Blockchain: Blockchain technology offers decentralized and immutable data storage and verification, providing enhanced security and transparency. It has the potential to transform data governance by enabling secure data sharing, provenance tracking, and consent management.
- 3. Internet of Things (IoT): IoT devices generate vast amounts of data, requiring robust data governance practices. Data governance needs to adapt to handle the unique challenges posed by IoT, such as data volume, real-time processing, and privacy considerations.
- 4. Cloud Computing: Cloud-based solutions offer scalable storage, processing power, and collaboration capabilities. Data governance in the cloud requires organizations to address security, privacy, and compliance concerns while leveraging the benefits of cloud infrastructure.

9.2. Future Directions and Opportunities for Advancing Digital Data Governance

As digital data governance continues to evolve, there are several future directions and opportunities for organizations to advance their data governance practices:

- 1. Privacy-enhancing technologies: With growing concerns around data privacy, organizations can explore and adopt privacy-enhancing technologies, such as differential privacy, secure multi-party computation, and homomorphic encryption. These technologies enable data analysis while preserving privacy and confidentiality.
- 2. Data ethics and responsible AI: As AI and ML become more prevalent, organizations should prioritize data ethics and responsible AI practices. These include considering the ethical implications in data collection, algorithm design, bias mitigation, and ensuring fairness and accountability.
- 3. Data governance for big data and unstructured data: As the volume and variety of data continue to increase, organizations need to develop data governance approaches that can effectively manage big data and unstructured data sources, such as social media data, text documents, and multimedia content.
- 4. Federated data governance: With the rise in data sharing collaborations and partnerships, federated data governance models are gaining importance. Organizations need to establish frameworks for data governance that facilitate secure and trusted data sharing while maintaining control and compliance.
- 5. Enhanced data transparency and access: Organizations should strive to enhance data transparency and provide greater access to data for internal and external stakeholders. This includes implementing data catalogs, data portals, and open data initiatives to promote data sharing and reuse.
- 6. Data governance in the context of emerging domains: As new domains and industries embrace digital transformation, such as healthcare, smart cities, and autonomous vehicles, data governance needs to adapt to address domain-specific challenges, regulatory requirements, and ethical considerations.
- 7. Collaboration and knowledge sharing: Given the complexity of digital data governance, collaboration and knowledge sharing among organizations, academia, and government entities are crucial. Establishing communities of practice, sharing best practices, and collaborating on data governance standards can drive advancements in the field.

10. Data, Information, and Knowledge Governance

Data, information, and knowledge are interconnected concepts that represent different levels of understanding and utilization of information within an organization. While data governance, information governance, and knowledge governance are related, they address different aspects and levels of governance within the data-to-knowledge continuum.

- 1. Data Governance: Data governance primarily focuses on the management and control of data assets within an organization. It involves establishing policies, processes, and controls to ensure the quality, integrity, privacy, and security of data. Data governance aims to ensure that data are accurate, consistent, and available for use across the organization. It involves activities such as data quality management, data security, data integration, data lifecycle management, and data stewardship. Data governance is concerned with the technical and operational aspects of managing data as a valuable organizational asset.
- 2. Information Governance: Information governance takes a broader perspective and encompasses the management and control of all forms of information within an organization, including data. It involves establishing policies, procedures, and practices to ensure that information is effectively managed throughout its lifecycle. Information governance focuses on aligning the organization's information management practices with legal, regulatory, and business requirements. It involves activities such as information classification, records management, information privacy, information security, knowledge management, and compliance. Information governance considers the strategic and operational aspects of managing information as a valuable organizational asset.
- 3. Knowledge Governance: Knowledge governance focuses on the management and control of knowledge assets within an organization. It is concerned with capturing,

organizing, sharing, and utilizing knowledge to enhance decision making and drive innovation. Knowledge governance involves establishing policies, processes, and systems to facilitate knowledge creation, capture, storage, retrieval, and dissemination. It includes activities such as knowledge sharing, knowledge retention, expertise management, communities of practice, and learning initiatives. Knowledge governance aims to harness the collective knowledge of individuals within the organization and create a knowledge-sharing culture.

While data governance primarily deals with the management of data assets, information governance extends to include other forms of information, and knowledge governance focuses specifically on managing knowledge assets. These three levels of governance are interconnected and build upon each other. Effective data governance is a foundation for information governance, and effective information governance provides the groundwork for knowledge governance. It is essential for organizations to consider all three levels of governance to ensure that data, information, and knowledge are effectively managed, utilized, and leveraged for organizational success.

Data governance, information governance, and knowledge governance are interrelated but distinct levels of governance within the data-to-knowledge continuum. They address different aspects of managing and utilizing data, information, and knowledge within an organization, each with its specific focus, objectives, and activities. A comprehensive governance framework that encompasses all three levels is essential for organizations to optimize the value and impact of their data, information, and knowledge assets.

Figure 6 represents the trends in data governance (blue), information governance (orange), and knowledge governance (yellow) worldwide, on a monthly basis, starting from January 2004 and ending in May 2023. The figure clearly illustrates the varying levels of attention given to data governance, information governance, and knowledge governance over time. From the graph, it is evident that knowledge governance consistently received the lowest level of focus throughout the entire period covered.



Figure 6. Worldwide search volume index for data, information, and knowledge governance from Jan 2004 to May 2023. (Data source: Google Trends).

In the early years, up until around 2017, both data governance and information governance attracted similar levels of attention. The lines representing data governance and information governance appear to be relatively close in value, indicating a comparable level of emphasis on managing and controlling data assets and information within organizations.

However, a notable shift occurred around 2017. From that point onwards, data governance experienced a significant increase in attention, surpassing both information governance and knowledge governance. The blue line representing data governance steadily rises and remains consistently higher than the other two lines.

This shift suggests that organizations began to prioritize data governance more prominently, recognizing its critical role in managing and governing data assets effectively. The dominance of data governance over information governance and the continued negligible attention given to knowledge governance highlight the increasing recognition of the importance of data in decision making and organizational strategies.

While the figure indicates a dominant focus on data governance in recent years, it is possible that increased awareness and emphasis on data governance could lead to a subsequent rise in the attention paid to information governance and knowledge governance.

The figure provides a snapshot of the observed trends up until the present time, and it is important to consider that governance practices and priorities can evolve over time. As organizations become more proficient in managing their data assets and establish robust data governance frameworks, they may shift their attention towards leveraging that data to derive valuable insights and drive informed decision making. This shift could potentially result in increased attention to information governance, which involves ensuring the accuracy, reliability, and usability of information within an organization.

Furthermore, as organizations recognize the strategic value of knowledge management and the need to effectively capture, share, and utilize knowledge assets, there could be a subsequent increase in the attention paid to knowledge governance. This would involve implementing frameworks and processes to manage knowledge assets, facilitate knowledge sharing, and foster a culture of continuous learning and innovation.

It is important to note that this analysis is based on the provided data points, and other contextual factors may influence the observed trends. Nevertheless, the figure clearly demonstrates that knowledge governance consistently received the lowest attention, while data governance emerged as the dominant governance category, surpassing information governance from around 2017 onwards.

It is also worth noting that the interpretation of the figure and the potential future trends are speculative and dependent on various factors such as organizational priorities, industry dynamics, and emerging technologies. Continued monitoring and analysis of data governance, information governance, and knowledge governance trends will provide a more comprehensive understanding of their interplay and evolving importance in the future.

11. Ensuring Ethics and Trustworthiness in Data Governance

Ensuring ethics and trustworthiness in data governance requires considering several important factors. Here are some key considerations:

- 1. Data Privacy: Protecting individuals' privacy is crucial in data governance. Implement strong privacy measures to ensure that personal information is handled securely and only used for the intended purposes. Comply with relevant data protection laws and regulations, such as the General Data Protection Regulation (GDPR) or any applicable local laws.
- 2. Data Security: Maintain robust security measures to safeguard data against unauthorized access, breaches, or malicious activities. Employ encryption, access controls, and regular security audits to ensure the integrity and confidentiality of the data.
- 3. Transparency: Foster transparency in data governance by clearly communicating how data are collected, used, and shared. Provide individuals with clear information about the purpose and nature of data processing, as well as their rights regarding their personal information.
- 4. Consent and Opt-Out Mechanisms: Obtain informed consent from individuals before collecting and processing their data. Offer clear mechanisms for individuals to opt out of data collection and processing activities if they choose to do so. Respect users' preferences and ensure their choices are honored.
- 5. Data Quality and Accuracy: Ensure that data collected and processed is accurate, reliable, and up to date. Implement procedures for data validation, data cleansing, and regular data quality checks to maintain the integrity and trustworthiness of the information.

- 6. Ethical Use of Data: Establish guidelines and policies to ensure the ethical use of data. Avoid biases, discrimination, or unfair practices when analyzing and applying data insights. Consider the potential societal impact of data usage and strive to minimize harm while maximizing benefits.
- Accountability and Governance Frameworks: Implement clear accountability structures and governance frameworks to oversee data management practices. Assign responsibilities, define roles and authorities, and regularly assess and monitor compliance with data governance policies.
- Data Retention and Deletion: Establish policies for data retention and deletion to ensure that data are only stored for as long as necessary. Define appropriate data retention periods based on legal requirements, business needs, and data usage purposes.
- 9. Regular Audits and Compliance Monitoring: Conduct regular audits to assess compliance with data governance policies and applicable regulations. Monitor and address any potential breaches, issues, or gaps in data ethics or trustworthiness promptly.
- 10. Data Literacy and Training: Promote data literacy within the organization and provide appropriate training to employees involved in data governance. Ensure that they understand the importance of ethics and trustworthiness in data handling and processing.

By focusing on these factors, organizations can establish a strong foundation for ethical and trustworthy data governance practices, fostering confidence among stakeholders and building a positive reputation in the use of data.

Ensuring ethics and trustworthiness in data governance requires that attention is paid to several crucial factors. These include prioritizing data privacy and implementing robust security measures to protect against unauthorized access. Transparency is key, with clear communication about data collection, usage, and individual rights. Obtaining informed consent and providing opt-out mechanisms are essential for respecting individuals' choices. Maintaining data quality and accuracy through validation and cleansing processes is vital. The ethical use of data involves avoiding biases and discrimination, while accountability and governance frameworks ensure responsible practices. Data retention and deletion policies should align with legal requirements and business needs. Regular audits and compliance monitoring help to identify and address issues promptly. Lastly, promoting data literacy and training employees foster a culture of ethical and trustworthy data governance. By addressing these factors, organizations can cultivate an environment that values ethics and trust in data handling and processing.

12. Consent Management and Provenance

Consent management and provenance are two critical components that are closely interconnected in ensuring effective data governance. Let us explore how they influence each other:

12.1. Consent Management

Consent management refers to the process of obtaining and managing individuals' consent for the collection, processing, and use of their data. It involves providing clear information to individuals about the purpose and nature of the data processing, their rights regarding their personal information, and obtaining their informed consent.

12.2. Provenance

Provenance in data governance refers to the ability to trace and understand the origin, ownership, and transformation history of data throughout its lifecycle. It involves capturing metadata that documents the source, context, and transformations applied to the data, ensuring transparency and accountability.

12.3. The Relationship

Consent management and provenance work together to enhance data governance practices in the following ways:

12.3.1. Trust and Transparency

Effective consent management ensures that individuals are fully aware of how their data are collected and used. By obtaining informed consent, organizations establish a foundation of trust with data subjects. Provenance, on the other hand, enhances transparency by providing a clear trail of data lineage, showing how data has been collected, processed, and transformed. The combination of consent management and provenance promotes transparency, fostering trust in data governance practices.

12.3.2. Compliance and Accountability

Consent management is essential for organizations to comply with data protection regulations and ensure accountability. By obtaining explicit consent, organizations demonstrate their commitment to respecting individuals' rights and complying with legal requirements. Provenance complements this by allowing organizations to track and demonstrate compliance through an auditable record of data lineage. It enables organizations to prove that data has been handled in accordance with consent and regulatory obligations, enhancing accountability.

12.3.3. Data Governance Controls

Consent management and provenance contribute to establishing effective data governance controls. Consent management ensures that data are collected and processed within the boundaries defined by individuals' consent, supporting compliance with privacy regulations. Provenance provides visibility into how data has been used, enabling organizations to enforce data governance policies and ensure that data are used appropriately and in alignment with consent.

12.3.4. Data Subject Rights

Consent management plays a crucial role in respecting data subject rights, such as the right to access, rectify, or erase personal data. Organizations need to maintain accurate records of consent to demonstrate compliance with these rights. Provenance assists in honoring these rights by enabling organizations to identify the exact location, use, and history of an individual's data, facilitating the exercise of data subject rights.

By integrating consent management and provenance, organizations can establish a comprehensive and accountable data governance framework. This integration ensures that data are collected and processed with individuals' consent, while also providing visibility and traceability throughout the data lifecycle, thereby promoting transparency, compliance, and trust.

13. Conclusions and Recommendation

Throughout the paper, the significance of digital data governance in official statistics and its critical role in ensuring the accuracy, reliability, privacy, and security of digital data has been explained. The paper extensively discusses the key principles and components of digital data governance, including data quality, privacy, security, ethics, integration, and stakeholder engagement, providing in-depth insights into each aspect.

The paper sheds light on the challenges encountered in implementing digital data governance and presents best practices and insights gained from successful implementations. It delves into emerging technologies that are shaping the future of digital data governance, such as AI, blockchain, IoT, and cloud computing. Furthermore, it explores upcoming trends and directions in digital data governance, encompassing privacy-enhancing technologies, responsible AI, big data governance, and federated data governance.

Based on the findings and insights presented in this paper, the following recommendations are offered for policymakers and practitioners to enhance digital data governance in official statistics:

- 1. Establish a comprehensive digital data governance framework: Policymakers should develop and implement a dedicated framework that addresses the unique challenges and requirements of digital data governance in official statistics. This framework should encompass data quality management, privacy and confidentiality, data security, ethics, data integration, stakeholder engagement, and continuous improvement.
- 2. Invest in capacity building and training: Policymakers and organizations should prioritize capacity building and training programs to enhance the skills and knowledge of professionals involved in digital data governance. This includes training on data governance principles, emerging technologies, data privacy, security protocols, and ethical considerations.
- 3. Foster collaboration and knowledge sharing: Policymakers should encourage collaboration and knowledge sharing among official statistical agencies, academia, industry experts, and other stakeholders. This can be achieved through the establishment of communities of practice, conferences, workshops, and platforms for sharing best practices, experiences, and challenges.
- 4. Ensure legal and regulatory compliance: Policymakers should review and update existing legal and regulatory frameworks to address the unique challenges of digital data governance. This includes data protection laws, privacy regulations, security standards, and ethical guidelines. Policymakers should also foster international cooperation to harmonize data governance practices across borders.
- 5. Embrace emerging technologies strategically: Policymakers and practitioners should monitor and evaluate emerging technologies and strategically adopt those that can enhance data governance in official statistics. This includes exploring the use of AI, blockchain, IoT, and cloud computing in a manner that aligns with data privacy, security, and ethical considerations.
- 6. Promote transparency and accountability: Policymakers should promote transparency in data governance practices by making information about data collection, processing, and usage readily available to the public. Organizations should establish mechanisms for accountability and feedback to ensure the responsible and ethical use of digital data.

By adhering to these guidelines, policymakers and professionals can bolster the effectiveness of digital data governance in official statistics, resulting in enhanced data quality, strengthened privacy safeguards, and heightened public confidence. Consequently, this will facilitate evidence-based decision making, policy development, and sustainable progress.

Author Contributions: Conceptualization, H.H. and S.M.; methodology, H.H. and S.M.; investigation, H.H. and S.M.; writing—review, H.H. and S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

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

References

- United Nations Statistical Commission. Fundamental Principles of Official Statistics. 2013. Available online: https://unstats.un. org/unsd/dnss/gp/FP-New-E.pdf (accessed on 30 May 2023).
- United Nations. Principles and Recommendations for Population and Housing Censuses: Revision 3. 2017. Available online: https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/ Population-and-Housing-Censuses/Series_M67rev3-E.pdf (accessed on 30 May 2023).
- World Bank. Data for Better Lives: A Decade of World Development Indicators. 2020. Available online: https://openknowledge. worldbank.org/bitstream/handle/10986/34158/9781464815678.pdf (accessed on 30 May 2023).
- Eurostat. European Statistics Code of Practice. 2020. Available online: https://ec.europa.eu/eurostat/documents/3859598/1121 5121/KS-GQ-20-001-EN-N.pdf/7d38e7b8-48f3-314f-177d-4c5c7c92e6e3 (accessed on 30 May 2023).
- International Monetary Fund (IMF). External Sector Report: Statistical Appendix. 2014. Available online: https://www.imf.org/ external/pubs/ft/scr/2014/cr1404.pdf (accessed on 30 May 2023).
- World Health Organization (WHO). Handbook on Health Inequality Monitoring: With a Special Focus on Low- and Middle-Income Countries. 2015. Available online: https://apps.who.int/iris/bitstream/handle/10665/132232/9789241548632_eng.pdf (accessed on 30 May 2023).
- Organisation for Economic Co-operation and Development (OECD). OECD Guidelines on Measuring Trust. 2020. Available online: https://www.oecd.org/gov/trust-in-government-guidelines.pdf (accessed on 30 May 2023).
- Statistics Canada. The Value of Official Statistics. 2017. Available online: https://www150.statcan.gc.ca/n1/pub/11-634-x/2017 001/tbl/tbl1-eng.htm (accessed on 30 May 2023).
- 9. Australian Bureau of Statistics. The Value of Official Statistics. 2013. Available online: https://www.abs.gov.au/websitedbs/D3 310114.nsf/home/The+Value+of+Official+Statistics (accessed on 30 May 2023).
- 10. Statistical Office of the European Union (Eurostat). The Economic Importance of Statistics. 2020. Available online: https://ec.europa.eu/eurostat/documents/64157/4376083/Economic-importance-of-statistics.pdf (accessed on 30 May 2023).
- United Nations. International Data Governance: Pathways to Progress. United Nations System Chief Executives Board for Coordination—Advance Unedited Version of 4 May 2023. Available online: https://unsceb.org/international-data-governancepathways-progress (accessed on 12 May 2023).
- 12. MacFeely, S.; Me, A.; Fu, H.; Veerappan, M.; Hereward, M.; Passarelli, D.; Schüür, F. Towards an International Data Governance Framework. *Stat. J. IAOS* **2022**, *39*, 703–710. [CrossRef]
- MacFeely, S.; Me, A.; Fu, H.; Schweinfest, S. We Urgently Need a Global Data Convention. World Economic Forum—Pioneers of Change Blog, 13 November 2020. Available online: https://www.weforum.org/agenda/2020/11/global-data-convention (accessed on 12 May 2023).
- Me, A.; Fu, H.; MacFeely, S. A Global Data Convention to Safeguard Sustainable Development. World Data Forum Data Blog, 5 October 2021. Available online: https://unstats.un.org/unsd/undataforum/blog/a-global-data-convention-to-safeguardsustainable-development/ (accessed on 12 May 2023).
- 15. MacFeely, S. In Search of the Data Revolution: Has the Official Statistics Paradigm Shifted? *Stat. J. IAOS* **2020**, *36*, 1075–1094. [CrossRef]
- 16. G7. Carbis Bay G7 Summit Communique. Our Shared Agenda for Global Action to Build Back Better. 2021. Available online: https://www.consilium.europa.eu/media/50361/carbis-bay-g7-summit-communique.pdf (accessed on 15 May 2023).
- 17. G7. G7 Roadmap for Cooperation on Data Free Flow with Trust. G7 Digital and Technology Track—Annex 2. 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/986160/Annex_2_ __Roadmap_for_cooperation_on_Data_Free_Flow_with_Trust.pdf (accessed on 30 May 2023).
- G20. G20 Osaka Leaders Declaration. 2019. Available online: https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/ en/documents/final_g20_osaka_leaders_declaration.html (accessed on 15 May 2023).
- 19. G77. Statement on behalf of the Group of 77 and China by the Delegation of the Republic of Cuba at the Informal Consultations on the Global Digital Compact on the Third Thematic Deep Dive: 'Data protection' (New York, 24 April 2023). Available online: http://www.g77.org/statement/getstatement.php?id=230424b (accessed on 30 May 2023).
- European Union. Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European Data Governance and Amending Regulation (EU) 2018/1724 (Data Governance Act). 2022. Available online: https://eur-lex.europa. eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R0868 (accessed on 15 May 2023).
- World Bank. World Development Report 2021: Data for Better Lives. 2021. Available online: https://www.worldbank.org/en/publication/wdr2021 (accessed on 10 May 2022).
- UNCTAD. Digital Economy Report 2021—Cross-Border Data Flows and Development: For Whom the Data Flow. 2021. Available online: https://unctad.org/system/files/official-document/der2021_en.pdf (accessed on 10 May 2022).
- 23. United Nations. Our Common Agenda—Report of the Secretary General. 2021. Available online: https://www.un.org/en/content/common-agenda-report/assets/pdf/Common_Agenda_Report_English.pdf (accessed on 15 May 2023).
- UNU. A Breakthrough for People and Planet: Effective and Inclusive Global Governance for Today and the Future. Report of the High-Level Advisory Board on Effective Multilateralism (HLAB). 2023. Available online: https://highleveladvisoryboard.org/breakthrough/pdf/56892_UNU_HLAB_report_Final_LOWRES.pdf (accessed on 15 May 2023).

- 25. Google. Global Digital Compact. Google Submission. 2023. Available online: https://www.un.org/techenvoy/sites/www.un. org.techenvoy/files/GDC-submission_Google.pdf (accessed on 16 May 2023).
- OECD. Moving Forward on Data Free Flow with Trust: New Evidence and Analysis of Business Experiences; OECD Digital Economy Papers; OECD: Paris, France, 2023.
- 27. Transform Health. Health Data Governance Principles. 2021. Available online: http://healthdataprinciples.org/principles (accessed on 24 May 2023).
- G7. G7 Global Plan for UHC Action Agenda. Communique from G7 2023 Hiroshima Summit. 2023. Available online: https://www.mhlw.go.jp/content/10500000/001096405.pdf (accessed on 23 May 2023).
- UHC2030. From Commitment to Action—Action Agenda on Universal Health Coverage from the UHC Movement. 2023 UN High-Level Meeting on Universal Health Coverage. 2023. Available online: https://www.uhc2030.org/un-hlm-2023/actionagenda-from-the-uhc-movement/ (accessed on 23 May 2023).
- Mathieson, C. Why Do We Need to Responsibly Unlock Data to Transform Health? Datasphere Initiative Blog. 6 April 2023. Available online: https://www.thedatasphere.org/news/why-do-we-need-to-responsibly-unlock-data-to-transform-health/ (accessed on 24 May 2023).
- United Nations. A Global Digital Compact—An Open, Free and Secure Digital Future for All. Our Common Agenda Policy Brief. Available online: https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-gobal-digi-compact-en.pdf (accessed on 25 May 2023).
- 32. Davis, T. "Data Governance and the Datasphere", Literature Review. Available online: https://www.thedatasphere.org/wp-content/uploads/2022/11/Datagovernance-and-the-Datasphere-Literature-Review-2022.-Tim-Davies.pdf (accessed on 30 May 2023).
- Coyle, D. Common Governance of Data: Appropriate Models for Collective and Individual Rights. Ada Lovelace Institute Blog. 30 October 2020. Available online: https://www.adalovelaceinstitute.org/blog/common-governance-of-data/ (accessed on 29 May 2023).
- World Health Organization. Global Strategy on Digital Health 2020–2025. 2021. Available online: https://www.who.int/docs/ default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf (accessed on 31 May 2023).
- 35. The Lancet and Financial Times Commission. Governing health futures 2030: Growing up in a digital world. *Lancet* 2021, 398, 1727–1776. [CrossRef] [PubMed]
- 36. WHO. Health Data Governance Summit: Governing Data for Better Health. Meeting Report. 30 June 2021. Available online: https://cdn.who.int/media/docs/default-source/world-health-data-platform/events/health-data-governance-summit/ summit-report_final.pdf?sfvrsn=1c6ff885_8&download=true (accessed on 31 May 2023).
- European Commission. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. Strasbourg, 3.5.2022 COM(2022) 197 Final 2022/0140(COD). 2022. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 (accessed on 31 May 2023).
- Lynch, C. Exclusive: Global South Thwarts UN Future Summit Plans. Devex. 26 May 2023. Available online: https://www.devex. com/news/exclusive-global-south-thwarts-un-future-summit-plans-105573 (accessed on 31 May 2023).
- 39. Available online: https://www.statcan.gc.ca/en/about/datastrategy (accessed on 30 May 2023).
- 40. Available online: https://ec.europa.eu/eurostat/web/quality/european-quality-standards/quality-assurance-framework (accessed on 30 May 2023).
- Available online: https://www.census.gov/about/policies/privacy/data_stewardship/dsep_committee.html (accessed on 30 May 2023).
- 42. Available online: https://www.abs.gov.au/about/data-services/data-integration (accessed on 30 May 2023).
- 43. Available online: https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/ (accessed on 30 May 2023).
- 44. Available online: https://www.insee.fr/en/accueil (accessed on 30 May 2023).
- 45. Available online: https://www.cso.ie/en/aboutus/lgdp/csodatapolicies/ (accessed on 30 May 2023).
- 46. Available online: https://kostat.go.kr/anse/ (accessed on 30 May 2023).
- 47. Reister, M.; MacFeely, S.; Me, A.; Hereward, M.; Schweinfest, S. A Statistical Quality Assurance Framework for the United Nations. *Stat. J. IAOS* **2022**, *38*, 725–732. [CrossRef]
- MacFeely, S. Big Data and Official Statistics. In *Big Data Governance and Perspectives in Knowledge Management*; Kruger, S., Kruger, M., Eds.; IGI Global: Hershey, PA, USA, 2018; pp. 25–54.
- MacFeely, S.; Barnat, N. Statistical Capacity Building for Sustainable Development: Developing the fundamental pillars necessary for modern national statistical systems. *Stat. J. Int. Assoc. Off. Stat.* 2017, 33, 895–909. [CrossRef]
- UN DESA 2020, UN E-Government Survey. 2020; p. 166. Available online: https://publicadministration.un.org/egovkb/ Portals/egovkb/Documents/un/2020-Survey/2020%20UN%20E-Government%20Survey%20(Full%20Report).pdf (accessed on 30 May 2023)

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