On the Connotation, Emergence Background, and Major Issues of Government Data Governance (Postprint)

Authors: Xia Yikun

Date: 2023-08-26T00:00:00+00:00

Abstract

[Purpose/Significance] Research on government data governance issues is a cutting-edge hot topic in the field of government information management, and holds important theoretical guiding significance for effectively connecting and advancing key issues in government information management such as government information disclosure, open data, and public information reuse. [Method/Process] Through methods such as literature research, concept comparison, background analysis, and logical deduction, this paper focuses on exploring the connotation and constituent elements of the government data governance concept, and points out the inevitability of the emergence of the government data governance concept from perspectives such as data-driven administration, economy, society, and data risk prevention. [Result/Conclusion] From macro, meso, and micro perspectives, it provides a detailed analysis of the main content and procedures of government data governance, and then points out its basic characteristics in terms of the integrity of system structure, the ordering of management, the coupling of governance structure, and the urgency of risk response. Finally, it focuses on analyzing the main problems and challenges faced by government data governance in the processes of integration with IT, internal-external integration, and value integration.

Full Text

On the Connotation, Context, and Main Issues of Government Data Governance

Xia Yikun Center for Studies of Information Resources, Wuhan University, Wuhan 430072

Abstract

[**Purpose/Significance**] Research on government data governance represents a cutting-edge hotspot in the field of government information management, offering important theoretical guidance for effectively connecting and advancing key issues such as government information disclosure, open data, and public information reuse. [**Method/Process**] Through literature review, conceptual comparison, contextual analysis, and logical reasoning, this paper focuses on exploring the connotation and constituent elements of the government data governance concept, and demonstrates the inevitability of its emergence from the perspectives of data-driven administration, economic development, social governance, and data risk prevention. [**Result/Conclusion**] The paper analyzes in detail the main content and procedures of government data governance from macroscopic, mesoscopic, and microscopic perspectives, and identifies its fundamental characteristics including architectural integrity, managerial orderliness, structural coupling of governance, and the urgency of risk response. Finally, it focuses on analyzing the main problems and challenges faced by government data governance in the processes of integration with IT, internal-external fusion, and value convergence.

Keywords: government data governance; open data; development and utilization of government data; value of data assets; data risk

---

1. Conceptual Framework and Context of Government Data Governance

The evolution of government information management research has always been closely integrated with information management practices and the development of network information technology. Based on a review of relevant domestic and international literature, the concept of government data governance has benefited from the increasing maturity of enterprise data governance systems and the continuous enrichment and improvement of concepts such as e-government, digital government, and government governance. Currently, academic circles have achieved relatively clear conceptual definitions of data governance, though no consensus-based explicit definition of government data governance has yet emerged. However, a series of forward-looking explorations have been conducted regarding its content architecture, element composition, and organizational management systems.

1.1 Definition of Data Governance

Since the rise of IT governance in the 1980s, the concept of data governance was first proposed by enterprises such as IBM. From an enterprise perspective, data governance refers to the practices of policies, processes, and standards specifically organized and implemented for the effective use of structured or

unstructured information assets [4]. Others define it as the methods by which organizations manage data and information, proposing a series of policies and procedures covering the entire data lifecycle from collection to use and disposal [5]. From a data lifecycle perspective, data governance is a complete set of data processes to ensure that important data assets are formally managed within the enterprise [6]. From a conceptual comparison standpoint, IT governance focuses more on hardware development management of information infrastructure and information systems, while data governance emphasizes the management of data flow processes and the realization of expected benefits behind various information infrastructures and tools.

Conceptually, data governance is closely related to and mutually supportive of information governance and data management. On the one hand, data governance constitutes an important component and key subset of information governance. Although both concepts include basic elements such as risk and value management, their objectives, objects, tools, and methods differ in implementation. Information governance emphasizes maximizing information benefits while minimizing costs and risks, whereas data governance focuses on data asset development and management at the data level, seeking to ensure data quality and reduce data risks through appropriate systems and processes. On the other hand, governance refers to decisions made to ensure effective management and use of resources, while management involves executing those decisions [7]. "Data management is the development, execution, and supervision of data plans, policies, programs, and practices to control, protect, deliver, and enhance the value of data and information assets" [6]. Data governance expresses governance over data management and data utilization [8], encompassing all institutional arrangements, technical applications, methodological innovations, and process optimizations for effectively protecting and controlling data. In short, "data governance is the advanced planning and control of data management" [6], reflecting the decision-making authority and responsibility exercise in each link of the data lifecycle to ensure effective management and resource allocation. Data management is both the concrete implementation of data governance concepts and an important component of the data governance structure, reflecting the execution of data processes and tasks in each link to meet business or regulatory requirements.

1.2 Contextual Analysis of Government Data Governance

From the perspective of evolution in information content, technology, and methodology, the emergence of the government data governance concept does not simply replace government information management but represents the in-depth development and refined manifestation of government information management in the digital age, with profound contextual influences (see [Figure 1: see original paper]).

(1) Data-driven administration requires efficient government data governance as its guarantee. The effective operation of government re-

lies on data as its foundation. Diverse data from various sources with different standards constitute the lifeline and fundamental elements of government governance. With the development of information technologies such as big data and cloud computing, data quality, data supply, and data security have more direct impacts on government administration. Low-quality data supply significantly reduces administrative efficiency. A telephone survey of U.S. state government officials showed that 70% of respondents believed data problems had interfered with and affected the effective conduct of business operations, manifested as inaccurate data, data loss, incomplete data, difficult access, and lack of comparability [9]. How to transcend existing constraints in systems, technology, and markets, and shift from passive control to active management of government data, and from single-node management of data processes to holistic data solutions to achieve value-added government data assets has become an urgent issue in current government data management. This requires strengthening overall planning of government data resources from a global perspective and establishing a cross-regional, cross-departmental open data sharing system to form a new pattern of government data governance with multi-party interaction between government and society.

(2) Data-driven economy requires clear government data governance as its foundation. The asset value of government data has been widely recognized by society. In Australia, accessing government data could potentially generate value worth $25 million annually [10]. The large-scale digitalization and servitization of government data products not only activate data demand and information exchange between government, enterprises, and society but also expand data resource development forms from single content processing to data intermediation, APP development, enterprise intelligence, platform applications, and other value realization methods, releasing the enormous industrial boundaries contained in government information services. However, for enterprises, on the one hand, the macro-level property rights relationships regarding government data ownership and usage rights for open reuse of data resources are not clear, and data circulation and transaction systems remain imperfect. On the other hand, existing government data acquisition licensing mechanisms, pricing systems, business models, data market regulatory guidance methods, and even personal data protection systems directly affect the supply status of government data, urgently requiring the establishment of a new data governance ecosystem based on value development that radiates government data value to all levels of the digital economy.

(3) Data-driven society requires comprehensive government data governance as its guide. The biggest change brought to the social domain by the data surge is the profound transformation of the relationship between government and society. On the one hand, the extensive penetration and flow of data in social life objectively endows the public with individual rights including data rights, enhancing individual data awareness and data acquisition and utilization capabilities. This requires high-quality public data quality assurance mechanisms and precise, rapid data response mechanisms to actively guide pub-

lic data consumption and effective participation in government data activities. On the other hand, the opening and sharing of government data has greatly changed and shaped social forms, essentially removing the centralization issues of traditional government data control and fostering the diversification of governance subjects. This requires systematic and open government data governance platforms to provide society with opportunities for participation in value creation and corresponding data utilization and control norms to maintain the healthy and orderly use of government data by society, thereby constructing a new data-driven social ecology and achieving flattened social governance.

(4) Data risk prevention requires a sound government data governance system as its support. The tremendous changes in the government data ecological environment have given rise to and intensified risks and hidden dangers in the development and utilization of government data, such as sensitive data protection and data leakage. "Since information workers within government agencies may freely interact with confidential data in large data warehouses, there is potential to forget the value and sensitive nature of information" [2], which prompts major adjustments and changes to traditional government data security management strategies that emphasize technology. It is necessary to elevate data security to a core issue of government data management at the national level for global planning, and to upgrade from purely technical management to management with dual attributes of technology and management, shifting from separate data classification and closed risk control levels to coordinated and unified data classification and technical standards based on management security.

2. Main Content and Structural Characteristics of Government Data Governance

Understanding the connotation of government data governance can be interpreted from macroscopic, mesoscopic, and microscopic perspectives. Macro-level government data governance represents the government's macro-management of the data industry, data economy, and even the entire social datafication process, such as national data strategies [3]. The mesoscopic level involves government data governance principles, data governance systems, data quality, and data lifecycle management. The microscopic level generally focuses on the definition of data elements, the structure, storage, and movement of data elements, mainly including data architecture management, data development, database operation management, data security management, master data management, data warehousing, document and content management, metadata management, and data quality management.

In 2017, a U.S. survey of chief data officers and chief analytics officers at federal and state levels found that current government data governance work concentrates on machine learning, data management initiatives, big data, deep data mining and analysis, partnerships, the Internet of Things, data analysis innovation, data center optimization, open data principles, open data frameworks,

artificial intelligence, recruiting and retaining data talent, and establishing data analysis teams [11]. Thus, government data governance in the usual sense basically concentrates on mesoscopic and microscopic levels. Compared with enterprise data governance, although both share basically the same constituent elements and main content (see), they differ significantly in terms of objects, scope, methods, and standard requirements. Government data governance not only breaks through single organizational boundaries, being broader and more macroscopic, but also has complex internal and external structures, variable governance environments, and higher standard requirements and target expectations. It extends from simple data asset value preservation and appreciation to comprehensive data sharing across society and full economic and social benefits, with richer connotations and more challenging implementation.

Specifically, the particularity of government data governance mainly manifests in the following aspects:

(1) Emphasizing architectural integrity. In the digital environment, isolated government information systems cannot effectively connect with other systems and establish data associations. The inherent maladies of government information management—such as inconsistent standards, varied formats, self-contained systems, and fragmentation—have blocked government data correlation and deep mining. Therefore, an integrated data governance approach is needed to replace the previous fragmentation in information management and segmented information system construction with unified institutional arrangements and standard systems (see [Figure 2: see original paper]), forming a multi-level, multi-agent comprehensive government data management and institutional arrangement, technical support from top-down unified guidance and bottom-up standardized advancement. This should be carried out with a more open, standardized, scientific, and inclusive attitude throughout the complete government data lifecycle, thereby effectively avoiding fragmented data management.

(2) Greater emphasis on managerial orderliness. On the one hand, government data activities are extensive and complex, requiring both national-level overall planning and top-level design of government data governance and standardized and orderly operational standards in specific data quality control and data dissemination micro-activities to ensure data activity uniformity and lay the foundation for subsequent data correlation and sharing. On the other hand, orderliness also implies inclusiveness in data governance, achieving close connection and division of labor between internal and external government data activities. This requires breaking through departmental interests and industry monopolies within the administrative system, strengthening quality assurance and risk prevention in data collection, storage, processing, aggregation, openness, and dissemination, continuously expanding the scope and scale of data openness, and ensuring the authority and accuracy of data resources. Simultaneously, it is necessary to play the leading role of government data governance, continuously activate upstream and downstream markets for government data,

and maintain standardized order, open sharing, value capture, and value realization in the dynamic circulation of internal and external data systems (see [Figure 3: see original paper]).

(3) More prominent coupling of governance structures. Reasonable data governance structures need to expand from focusing on completing specific data tasks to adopting a long-term data perspective throughout public administration as a whole, closely integrating the advancement of government data governance processes with relevant data activities and management elements. For example, the close integration of government data governance with government information disclosure activities, aided by multi-source heterogeneous data collection systems and fine-grained data extraction and viewpoint mining, will greatly improve the scale and quality of information disclosure and accelerate government information response. The close integration with open data activities not only enhances cross-regional, cross-industry, and cross-departmental self-organization and coordination of government data management but also proposes an integrated government data processing, storage, and dissemination utilization model, forming a systematic data security risk review and prevention and control system. The close integration with public information reuse activities, through the batch opening of high-quality government datasets, can not only promote the integration of government and social data resources and enhance enterprise data consumption and reproduction capabilities but also create a new situation of co-governance through the integration of public data resources.

(4) More urgent risk response. "While information technology brings various advantages to state power, it also brings enormous risks and challenges to national security" [12]. In addition to technical failures, system vulnerabilities, and human fraud that may cause centralized government data storage or top-down data control to suffer illegal hacker attacks, the upgrading of associated information technology and algorithmic languages makes cross-fusion and analytical mining of multi-data sources highly likely to cause data leakage, data fraud, and personal data rights infringement, threatening data asset development security, data sharing and utilization security, and even national security. This has generated an inherent need to strengthen holistic government data risk prevention (see [Figure 4: see original paper]), involving the accurate division and reasonable use of data management rights, data control rights, data personality rights, and data property rights. It requires comprehensive supervision of data quality, application software, platform systems, infrastructure, and data operations.

3. Main Problems and Challenges in Current Government Data Governance

The transition from document management to data governance presents government data management departments with numerous complex contradictions that require gradual resolution. These include contradictions between the cross-

boundary complexity of multi-dimensional data flow associations and the rigidity of government data management systems and mechanisms; between the infinity of data scale and types and limited data processing capacity; between diverse user data needs and limited government data supply levels; and between unlimited business opportunities in government data asset development and limited reuse capacity and value realization levels in data markets. Behind these contradictions lie deep-seated issues urgently requiring resolution in government data governance, such as the evolution and reorganization of the government data governance ecosystem, the expansion and reconstruction of the government data industry value chain, cross-boundary collaboration of government data resources (such as integrating various government data, enterprise data, and social data), intelligent management of government data processes, improvement of government data institutions, rapid development of government data technology, and the expansion and competition of government data governance subjects. Government data governance faces a series of unpredictable pressures.

3.1 Problems and Challenges from IT Integration

U.S. survey results of state government officials show that 17% of data problems are related to technology application [9]. The integration of data governance and IT requires addressing both public information infrastructure construction and core content such as order, efficiency, and effectiveness in data flow processes. How can digital technologies such as big data, machine learning, and predictive analytics be quickly and effectively integrated into key links of government data collection, storage, processing, and dissemination to ensure people's ability to control data processing, dissemination, and utilization? How can the superposition of data openness and related technology applications properly balance data reuse with preventing sensitive information disclosure and anonymous identity recognition? How do the scalability of data from different sources, the accessibility of different subjects, and the output of different data products/services impact and reshape existing departmentalized government data organizational systems, posing higher functional requirements for public data platforms? How can technology-driven intelligent management and human-computer interaction be achieved? These issues require the establishment of comprehensive privacy and security protection mechanisms and a scientific and unified data governance institutional framework at all stages of government data development and utilization.

3.2 Problems and Challenges from Internal-External Fusion

Government data governance breaks the invisible data barriers between government departments, enterprises, and society, achieving not only cross-government department data exchange and sharing but also orderly promoting the fusion of government data with enterprise and social data. This transformation from closed systems to open systems has reconstructed government data management

methods and raised new issues requiring attention:

First is the division of responsibilities among data governance subjects. Traditional data management methods confined to single institutions no longer meet actual needs. It is necessary to establish a collaborative data governance organization model across government departments and even across government, enterprises, and society to negotiate and handle data development and utilization issues. Under the existing government data management system, the authority of this model's effectiveness and its decision-making and supervision execution have certain looseness and unpredictability, posing unprecedented challenges to government data governance capabilities.

Second is the unification of data standards. The fusion and aggregation of datasets from different sources require unifying or re-converting previous respective data formats, metadata, and other standards, such as data standard system construction, data quality terminology and evaluation dimensions, data file exchange norms and principles, master data distribution principles, and the design and application management of technical metadata, business metadata, and operational metadata, to promote cross-boundary fusion of multi-source datasets.

Third is coordinating the relationship between standardized management and maintaining flexibility. With the complex and changing data environment, while standardizing government information systems, it is also necessary to resist one-size-fits-all management approaches. For example, developing diversified and precise data push solutions based on diverse user group needs and capabilities, setting different levels of user access permissions according to data sensitivity, and establishing different data disclosure and acquisition licensing systems according to data types and industry market characteristics.

Fourth is the improvement and perfection of government data management regulations and institutions. Cross-department, cross-industry, and cross-domain data fusion often causes conflicts between internal data management systems of various agencies, requiring the breaking of conventions to establish integrated information management systems that conform to the government data governance ecosystem and a culture of data development and utilization based on trust, as well as data security accountability mechanisms.

In addition, there are issues such as constructing data quality assurance systems, data overload and fragmentation, and balancing data openness with confidentiality and privacy protection. Resolving these issues requires forming comprehensive privacy and security protection mechanisms and a scientific and unified data governance institutional framework throughout all stages of government data development and utilization.

3.3 Problems and Challenges from Value Fusion

For a long time, government information management practice has overemphasized the regulation of information flow order and management efficiency while neglecting information value development, cost compensation, and benefits. The introduction of the data governance concept, while strengthening the asset attributes of government data, also emphasizes the reconstruction and mining of the government data value system and seeks solutions from theoretical, institutional, and methodological perspectives:

First is the innovation of government data value creation. This involves both the transformation of data resource allocation methods from public supply to the integrated symbiosis of multiple systems combining market allocation and public allocation, and the balance between social benefits, economic benefits, and the trade-offs between immediate data interests and long-term benefits in government data asset management. Meanwhile, government data transmission and information sharing beyond time and space objectively accelerate the value exchange, transfer, and appreciation of data products. This requires exploring how to meet the needs of different user types from the perspective of value realization paths and methods, and optimizing government data product processing and production flows and enriching business models from the dimensions of value driving methods, value realization forms, and value chain structures, thereby forming a long-term mechanism for data value realization.

Second is the reconstruction of the relationship between data value-added development and government departments' human, financial, material, and management resources. The core issue is the ownership and benefit distribution of government data assets. While continuously improving the management of ownership, usage rights, and revenue rights of government data assets, it is also necessary to consider how to strengthen cost control of government data governance, maintain the sustainability of data governance, and establish a fair and reasonable data cost-sharing mechanism to allocate the costs and benefits of value creation. At the micro level, it is necessary to further consider how to restrain the unequal power of data development and utilization among different stakeholders such as government, enterprises, society, and individual hackers, clarifying the functional boundaries, activity trajectories, and task lists of different subjects to achieve a balance between maximizing the economic and social value of government data and individual data value.

Third is the integration of user value development and government data value realization. Data governance emphasizes the combination of data value creation and value sharing, focusing on user needs interpretation and user experience enhancement. However, the public product attributes of government data, which advocate digital services, public participation, democratic expansion, and the differential and individual nature of user needs, do not completely align with asset management objectives. It is necessary to deepen user value perception from data openness and sharing, timely grasp the disruptive change of data dom-

inance shifting from government to a balance between government and users, and make corresponding adjustments in management focus and methods.

In summary, the key to government data governance in the digital environment lies in efficiently integrating internal and external government data resources to achieve innovation in data development and utilization models. The focus is on improving and perfecting government data governance strategies and institutional systems, establishing the participation level of multiple data subjects under government leadership, and building data development and utilization partnerships. This involves the expansion and reconstruction of the government data value chain, and the construction of a cross-boundary, collaborative, and holistic management system for data security protection and risk prevention accompanying the expansion of data scale and time-space dimensions during the processes of technology fusion, internal-external fusion, and value fusion.

To achieve these requirements, efforts are needed in the following aspects: Enhancing understanding to deeply comprehend the significance of government data governance for digital government construction and data value realization, and establishing integrated data governance thinking and an open and collaborative data culture within the government administrative system; Strengthening the institutional supply capacity of government data governance, further improving relevant government information legal systems concerning confidentiality, personal data protection, data reuse, and data quality control, and paying attention to internal coordination and unity among information regulations; Improving the government data governance system and refining corresponding management details at the data level in current government information systems to continuously enhance the ability of various subjects to mine, analyze, and deeply utilize government data; Promoting the transformation of government data management forms from closed business models based on administrative objectives and task control to cross-boundary open business models based on value networks, achieving a shift in government decision-making models from experience-driven delayed decisions to data-driven real-time decisions; Paying greater attention to the value attributes of data and the convenience and security of user utilization, advocating integrated and precise data management, and committing to enriching government data management theories and innovating practical application methods.

References

[1] DAWES S S, HELBIG N. Information strategies for open government: challenges and prospects for deriving public value from government transparency [C] // WIMMER M A, CHAPPELET J L, JANSSEN M, et al. Electronic government EGOV 2010. Berlin: Springer, 2010: 50-60.

[2] THOMPSON N, RAVINDRAN R, NICOSIA S. Government data does not mean data governance: lessons learned from a public sector application audit

[J]. Government information quarterly, 2015, 32(3): 316-322.

[3] 黄璜. 美国联邦政府数据治理: 政策与结构 [J]. 中国行政管理, 2017(8): 47-56.

[4] PHANSE K. Data governance using SAP MDM [EB/OL]. [2016-12-05]. http://www.sdu.sap.com/irj/sdn/go/portal/prtroot/docs/library/uuid/600022998-5dd17-2b10-dbaa-8e3ab357fa55.

[5] KORHONEN J J, MELLERI I, HIEKKANEN K, et al. Designing data governance structure: an organizational perspective [J]. GSTF journal on computing, 2013, 2(4): 11-19.

[6] DAMA International. The DAMA guide to the data management body of knowledge (DAMA-DMBOK) [M]. USA: Technics Publications, LLC, 2009.

[7] FU X, WOJAK A, NEAGU D, et al. Data governance in predictive toxicology: a review [J]. Journal of cheminformatics, 2011, 3(1): 1-16.

[8] British Academy and the Royal Society. Data management and use: governance in the 21st century [R/OL]. [2017-12-13]. https://royalsociety.org/topics-policy/projects/data-governance/archive/insights/ON%20Report_{{{061114}}_{{FNL}}}.pdf.

[9] BARRETT K, GREENE R. The causes, costs and consequences of bad government data [EB/OL]. [2017-10-09]. http://www.governing.com/topics/mgmt/gov-bad-data.html.

[10] GRUEN N, HOUGHTON J, TOOTH R. Open for business: how open data can help achieve the G20 growth target [EB/OL]. [2017-07-05]. http://www.omidyar.com/sites/default/files/file_{AnalysisonGovernmentDataGovernance}.pdf.

[11] State of the Union: data and analytics in government [EB/OL]. [2017-08-20]. http://coriniumintelligence.com/chiefdataofficergovernment.

[12] 摩根索. 国家间政治 [M]. 孙芳, 译. 北京: 北京大学出版社, 2005: 156.

Acknowledgments: This paper is one of the research results of the National Natural Science Foundation of China project "Research on the Reconstruction of Public Information Resource Reuse System under Open Data" (Project No. 71373195).

Note: Figure translations are in progress. See original paper for figures.