Comparative Analysis of Data Management and Data Governance and Its Implications for Scientific Data Open Sharing Policy Development (Postprint)

Authors: Sheng Xiaoping, Song Dacheng

Date: 2023-04-01T00:00:00+00:00

Abstract

[Purpose/Significance] This study conducts a comparative analysis of the differences and connections between data management and data governance to provide references for formulating scientific data open sharing policies.

[Method/Process] Employing comparative analysis, this paper examines the similarities and differences between data management and data governance across five aspects: definition and connotation, function, objectives, principles, and focus areas, thereby deriving implications for the formulation of China's scientific data open sharing policies.

[Results/Conclusion] Data management and data governance exhibit significant differences in definition and connotation, function, objectives, principles, and focus areas, while also possessing intrinsic connections. Data governance is the key to successful implementation of data management. Clarifying this relationship helps identify deficiencies in China's current scientific data management policies, provides references for improving scientific data management approaches in the future, and facilitates the planning and formulation of practical detailed rules for scientific data open sharing.

Full Text

Preamble

Volume 64, Issue 22, November 2020
ChinaXiv Cooperative Journal

A Comparative Analysis of Data Management and Data Governance and Its Implications for Formulating Scientific Data Open Sharing Policies

Sheng Xiaoping, Song Dacheng
Department of Library, Information and Archives, Shanghai University, Shanghai 200444

Abstract:
[**Purpose/Significance**] This paper compares and analyzes the differences and connections between data management and data governance to provide references for formulating scientific data open sharing policies. [**Method/Process**] Using comparative analysis, this study examines the similarities and differences between data management and data governance across five dimensions: definition and connotation, functions, objectives, principles, and focus areas, and subsequently analyzes their implications for developing China's scientific data open sharing policies. [**Result/Conclusion**] Data management and data governance exhibit significant differences in definition and connotation, functions, objectives, principles, and focus areas, yet they also share intrinsic connections. Data governance is the key to successful data management implementation. Clarifying their relationship helps identify deficiencies in current scientific data management policies in China, provides references for improving scientific data management methods, and enables the planning and formulation of practical scientific data open sharing rules.

Keywords: data management; data governance; comparative analysis; scientific data; open sharing policy
Classification Number: G203
DOI: 10.13266/j.issn.0252-3116.2020.22.001

Data is critically important to the success of almost all organizations, regardless of size, and many have attempted to implement data governance initiatives [1]. Data governance is also a key domain and important lever for China to implement its national big data strategy and accelerate the construction of a digital China [2]. However, data governance and data management are easily confused concepts, with four distinct viewpoints currently existing: (1) They represent different levels of activity—data governance focuses on macro-level evaluation, guidance, and supervision of data management activities, while data management concentrates on micro-level actions to implement data governance decisions [3]; data governance sits above data management, representing a redistribution of data content and rights [4]. (2) Data management encompasses data governance—for instance, data governance is considered part of data management [5], the central activity of data management [6], or an important component of corporate data management strategy [7]. (3) Data governance encompasses data management—for example, data governance includes three layers: data services, data management, and data coordination [8]. (4) Data management equals data governance—for instance, data governance is a data manage-

ment concept involving organizational capabilities to ensure high data quality throughout the data lifecycle [9]. These perspectives reveal considerable confusion and ambiguity in understanding data management and data governance. Moreover, while some domestic policy documents, such as the *Big Data Industry Development Plan (2016-2020)*, emphasize data governance importance without mentioning data management, others like the *Scientific Data Management Measures* stress data management while neglecting data governance. This raises critical questions: What exactly are the differences between data governance and data management? How can we effectively utilize both to implement scientific data open sharing in China? Few studies have thoroughly examined these issues. Therefore, it is essential to clarify the relationship between data governance and data management, analyze their similarities and differences, and thereby formulate effective scientific data open sharing policies and promote the implementation of scientific data open sharing activities.

2. Comparative Analysis of Data Management and Data Governance

The similarities and differences between data management and data governance can be compared across five dimensions.

2.1 Comparison of Definitions and Connotations

Simply put, data management involves managing data and information by representing it in a formalized manner suitable for communication, interpretation, or processing [10]. This represents a narrow definition. Broadly speaking, data management encompasses the development, execution, and supervision of data plans, policies, procedures, and practices to deliver, control, protect, and enhance the value of data and information assets throughout their lifecycle [6]17. Broad data management includes data governance. However, data management typically focuses on defining data elements, how to store, structure, and move them to ensure reliability and usability [11], representing lifecycle management of data. This paper compares narrow data management with data governance.

Data governance involves exercising authority and control over data asset management (including planning, monitoring, and implementation) [6]67. In other words, data governance is the process of establishing standards, rules, policies, and oversight to ensure compliance with data management best practices [12]. It requires comprehensive management of data availability, relevance, usability, integrity, and security within an organization [13]. Data governance entails [14]: (1) establishing and managing public policies and affairs regarding data, including data-related laws and policies and data quality control and management strategies; (2) exercising data decision-making authority, including determining data meanings, usage locations, accuracy, and applicable rules; (3) controlling data velocity or volume, including data sharing according to classification rules (confidential, sensitive, public); (4) controlling data actions, including establish-

ing and monitoring appropriate processes for managing data definition, production, and use at all organizational levels; (5) ensuring data control, including hierarchical management, protection, and auditing of data access; and (6) influencing data decisions, ensuring the right people make correct data decisions at the right time.

2.2 Comparison of Functions

Data management and data governance serve different functions. Generally, data management functions primarily include [15]: (1) data architecture management—defining data architecture and data asset management blueprints; (2) data development—analyzing, designing, implementing, testing, deploying, and maintaining data, identifying structured and unstructured data as valuable organizational resources; (3) data operations management—providing operational support from data acquisition to disposal; (4) data security management—protecting data privacy, confidentiality, and security with appropriate access controls; (5) data quality management—defining, monitoring, and improving data quality; (6) reference and master data management—managing various master and reference data; (7) data warehouse and business intelligence management—providing data storage and business intelligence through data warehouses or repositories; (8) document and content management—managing data files and content; and (9) metadata management—integrating, controlling, and providing metadata.

Effective data governance offers numerous important functions [16-17]: (1) supporting organizational strategic initiatives by clarifying goals for cost reduction, improved decision-making processes, and enhanced efficiency; (2) helping develop and implement data quality management protocols, guidelines, and roadmaps to provide unique data and information management capabilities; (3) providing consistent definitions, methods, and procedures for better data acquisition, improved data quality, security, consistency, and organizational trust in data, while reducing redundant data storage and risks from inaccurate data; (4) facilitating standardized data quality processes with clearly defined metrics to measure and track data quality across the organization, improving coordination between business units and enabling more effective data management; (5) enabling convenient and timely access to data across organizational departments and external business data, preparing business units for better information sharing; (6) promoting internal network development for effective data transfer, exchange, and allocation, gaining competitive advantages and enhancing reciprocity, connectivity, and cohesion between business units; (7) fostering a data quality culture, promoting broader understanding of data across products and business units and broader thinking about quality issues; (8) improving compliance with organizational rules, enhancing data accountability, and preventing negative business impacts from accidental data disclosure; and (9) improving relationships and communication with customers (data providers and users), elevating organizational standing in industry and society.

Despite functional differences, data management and data governance share intrinsic connections. Both treat data as the management object. Data management functions focus on lifecycle management, including planning, documentation, organization, analysis improvement, sensitive data protection, storage and backup, post-project care, effective sharing, and reuse discovery—practices that can lay the foundation for data governance [18]. Data governance focuses on data strategy, quality, security, compliance, and accountability, representing both a redistribution of data content and rights [4] and high-level planning and control of data management [11]841, providing guidance [15]37, decision-making, and oversight for data management execution.

2.3 Comparison of Objectives

Data management and data governance do not share identical objectives. Data management objectives include [6]18: (1) understanding organizational and stakeholder data and information needs; (2) acquiring, storing, protecting, and ensuring data asset integrity; (3) ensuring data and information quality, including accuracy, completeness, integration, timeliness, relevance, and usefulness, as well as clear and consensus-based data definitions; (4) protecting stakeholder data privacy and confidentiality; (5) preventing unauthorized and inappropriate data use; and (6) ensuring effective data use to add organizational value. In summary, data management objectives should be specific, measurable, achievable (or executable), realistic, and timely for designated timeframes [15]19.

Unlike data management objectives, data governance's overarching goal is ensuring organizations can effectively manage data as assets [6]71. This can be subdivided into [20]: (1) enabling better decision-making; (2) reducing business friction; (3) protecting data stakeholder needs; (4) training managers and staff to adopt common approaches to data issues; (5) establishing standardized, repeatable processes; (6) ensuring process transparency; and (7) reducing costs and improving efficiency through coordinated efforts.

Thus, data management and governance share common objectives in ensuring stakeholder data needs and using data effectively to improve efficiency or create value. However, data management emphasizes understanding institutional or user data needs, organizing and providing relevant data, and maximizing data value utilization. Data governance focuses on more effective data asset management, transforming data into actionable knowledge, and reducing data risks [21].

2.4 Comparison of Principles

Owning data does not equate to knowing how to handle it. Data is an asset with unique properties—it is not consumed like financial or physical assets during use and requires an organizational perspective. Data management principles primarily include [6]21-23: (1) Data value quantification—data value should be expressed in economic terms as it is a valuable organizational asset, requiring con-

sistent methods to measure qualitative and quantitative value, including costs of low-quality data and benefits of high-quality data for better decision-making; (2) Data quality—organizations must understand stakeholder quality requirements and measure quality accordingly; (3) Metadata management—managing any asset requires data about that asset, with metadata coming from processes including architecture, modeling, management, governance, system development, IT, business operations, and analytics; (4) Planning management—managing data requires planning even in small organizations with complex technical and business processes, as data created in different places must be planned and managed from architectural and workflow perspectives to ensure consistency; (5) Team collaboration—data management requires diverse skills and expertise beyond single teams, necessitating cross-functional collaboration; (6) Dynamic management—data is fluid and changing, requiring dynamic perspectives; (7) Lifecycle management—different data types have different lifecycle characteristics and management needs; (8) Risk management—managing data includes managing data-related risks, as data represents organizational risk through potential loss, theft, or misuse; (9) Technology relevance—data management is closely related to IT, requiring approaches that ensure technology serves strategic data needs; and (10) Leadership support—effective data management requires leadership commitment.

Data governance principles differ [20]: (1) Consistency—ensuring alignment with business needs and common data understanding; (2) Multi-level—occurring across the entire organization and between departments; (3) Integrity—participants are honest and forthright in interactions and decisions; (4) Transparency—governance and management workflows are transparent; (5) Auditability—decisions, processes, and controls are auditable; (6) Accountability and checks and balances—clearly defining responsibilities with checks and balances; (7) Standardization—introducing and supporting data standardization; and (8) Change management—supporting proactive and reactive change management.

2.5 Comparison of Focus Areas and Related Activities

Data management and data governance have distinct focus areas with different primary activities. Data management focus areas include data architecture management, data quality management, data warehouse and business intelligence management, data storage and operations management, data integration and interoperability management, document and content management, reference and master data management, data modeling and design management, metadata management, and data security management [6]45-46.

Data governance focus areas generally include seven aspects [15]16-17, [22-23]: (1) data governance focused on data management plans; (2) data governance focused on data structures; (3) data governance focused on data quality; (4) data governance focused on data warehouses and business intelligence; (5) data governance focused on data access; (6) data governance focused on data usage;

(7) data governance focused on privacy, regulations, and security; and (8) data governance focused on data management controls.

Based on these focus area divisions, further comparative analysis of related activities reveals differences and similarities.

(1) Comparison of Different Focus Areas and Activities. Most focus areas and related activities differ between data management and data governance (see Table 1).

(2) Comparison of Same Focus Areas and Activities. Although data management and data governance share three common focus areas—data quality management, data architecture management, and data warehouse and business intelligence management—they emphasize different activities (see Table 2). In data quality management, data management focuses on defining, executing, improving, and monitoring data quality, while data governance centers on quality, integrity, and availability. In data architecture management, data management builds architecture frameworks and practices, whereas data governance focuses on policies, standards, metadata schemes, and authority-responsibility management. In data warehouse and business intelligence management, data management involves understanding, defining, developing, and maintaining systems, while data governance concentrates on usage rules, stakeholder authority-responsibility clarification, and data asset value determination.

3. Distinctions and Connections Between Data Management and Data Governance

In summary, distinctions between data management and data governance primarily manifest in: (1) Data governance involves planning, decision-making, supervision, and control of data management, while data management manages data lifecycle activities (planning, collection, description, synthesis, analysis, preservation, evaluation, reuse). (2) Data management's main objective is providing high-quality relevant data to maximize value utilization, while data governance's primary goal is ensuring effective data asset management and operations to maximize value and minimize risks. (3) Data management follows principles of data value quantification, data quality, metadata management, data planning, dynamic management, lifecycle management, technology relevance, and leadership support, focusing on data attributes, processes, and skill requirements. Data governance emphasizes consistency, transparency, compliance, auditability, standardization, and accountability, focusing on data responsibility, quality, standards, and compliance. (4) Both share three focus areas—data architecture, data quality, and data warehouse/business intelligence management—but with different activities. Other data management focus areas include data storage and operations, data integration and interoperability, document and content management, reference and master data management, data modeling and design, metadata, and data security. Other data governance focus areas include data management planning, monitoring, data access, data

usage, and data privacy, regulations, and security issues. (5) Data governance does not involve every detail of data management but focuses on data availability, integrity, consistency, reliability, security, architecture, quality, privacy, and compliance [25].

Conversely, data governance and data management possess intrinsic close connections: (1) Data governance essentially belongs to the broad category of data management [26] and is key to successful narrow data management implementation. (2) Both are indispensable for any organization. Without effective data management, organizations face numerous issues [10]58-59: information systems cannot interconnect due to incompatible definitions and development; data cannot be shared between systems despite technical connectivity; communication breakdowns and information loss occur; unnecessary data transcription and reentry increase R&D time and costs; competitive advantages weaken; user and employee frustration arises. Without data governance, organizations cannot effectively handle data, track and supervise utilization, identify trustworthy, secure, compatible, confidential, and accessible data, or maximize governance investment while minimizing data breach risks. (3) Both share objectives in ensuring data quality and security and maximizing data asset value. (4) Both emphasize leadership support, risk management, collaborative management, and dynamic management principles.

4. Implications for Formulating China's Scientific Data Open Sharing Policies

Understanding the distinctions and connections between data governance and data management helps examine scientific data management policies domestically and internationally and guides scientific data open sharing practices, particularly for formulating China's policies in two key aspects.

First, clarifying deficiencies in scientific data management policies provides references for future improvements. The *Scientific Data Management Measures* issued by the State Council on March 17, 2018, first defined responsibilities and practices for the State Council's science and technology administrative department, relevant departments, provincial science and technology departments, research institutes, universities, enterprises, and scientific data centers during scientific data management processes (collection, submission, preservation, sharing, utilization) and "confidentiality and security" aspects [27], significantly promoting scientific data management in China. While establishing the general principle of "openness as the norm, non-openness as the exception," the Measures do not specify implementation details for scientific data open sharing. Notably, they omit scientific data governance provisions, failing to address data ownership, open access timeframes, data publication and citation mechanisms [28], or to regulate critical issues like personal data protection, privacy governance, data quality, data rights, and data services in scientific data sharing. Although emphasizing scientific data management, the Measures neglect scientific data governance measures. In fact, efficient and successful

scientific data management requires scientific data governance support. A prerequisite for successfully implementing nationwide scientific data open sharing is establishing a sound scientific data open sharing governance system, including a "national-provincial-institutional-individual" four-level governance structure, clarifying rights, obligations, and responsibilities of different stakeholders, and establishing mechanisms for scientific data quality tracking and evaluation, security and privacy protection, and open sharing based on availability, integrity, consistency, reliability, security, standardization, and compliance. Only by effectively integrating scientific data management and governance and adding scientific data governance measures in future *Scientific Data Management Measures* revisions can scientific data open sharing be more effectively implemented.

Second, planning and formulating practical scientific data open sharing rules. Under the guidance of the *Scientific Data Management Measures*, the Chinese Academy of Sciences (CAS) pioneered industry-specific scientific data management rules on February 11, 2019, with the *CAS Scientific Data Management and Open Sharing Measures (Trial)* [29]. These trial measures effectively combine data management concepts with data governance thinking to regulate scientific data management and open sharing behaviors, including: (1) integrating data management philosophy and data governance ideology to formulate general principles of "coordinated planning, standardized management, security and control, and sustainable development for scientific data work, implementing main responsibilities, innovating management mechanisms, strengthening capacity building, and promoting open sharing" (Chapter 1, Article 4); (2) constructing a preliminary CAS scientific data governance system by clarifying responsibilities of the CAS Cybersecurity and Informatization Leading Group, its office, CAS departments, affiliated legal entities, and CAS scientific data centers (Chapter 2, "Division of Responsibilities"), assigning scientific data management to specific units or departments; (3) applying data management thinking to regulate submission and management of scientific project data and paper-associated data (Chapters 3 and 4), clarifying practices for affiliated institutions to openly share scientific data (Chapter 5); and (4) establishing guarantee mechanisms and security confidentiality mechanisms for scientific data management and open sharing (Chapters 6 and 7), strengthening both management thinking and flexible governance measures. Overall, these trial measures represent a pioneering initiative for implementing China's national big data strategy and promoting scientific data open sharing. However, they inadequately define individual (researcher, user) rights, responsibilities, and obligations in scientific data management and open sharing, resulting in an incomplete scientific data governance system. Moreover, few domestic institutions have issued scientific data open sharing rules. This indicates China must strengthen policy research on scientific data management and governance and practical research on open sharing, integrating data management and governance concepts to establish and improve a scientific data open sharing policy system ensuring sustainable development of scientific data open sharing practices.

References

[1] PLOTKIN D. Data stewardship: an actionable guide to effective data management and data governance[M]. Waltham: Morgan Kaufmann, 2014: 1.
[2] Li Haiyun. Xi Jinping: Implement national big data strategy and accelerate digital China construction[EB/OL]. [2020-07-20]. http://www.xinhuanet.com/2017-12/09/c_{1122084706}.htm.
[3] Bao Dongmei, Fan Yingjie, Li Ming. Data governance and its framework in university libraries[J]. Library and Information Service, 2015, 59(18): 134-141.
[4] Ren Yazhong. From data management to data governance: transformation of library functions in the big data environment[J]. Journal of the Sichuan Society for Library Science, 2017(4): 6-9.
[5] HARPER J. Distinguishing data management from data governance[EB/OL]. [2020-07-20]. https://www.dataversity.net/distinguishing-data-management-from-data-governance/.
[6] DAMA International. Data management body of knowledge[M]. 2nd ed. Basking Ridge: Technics Publications, 2017: 35.
[7] Data Republic. Data governance vs data management[EB/OL]. [2020-07-20]. https://www.datarepublic.com/blog/data-governance-vs-data-management/.
[8] Lin Ning. From data management to data governance[N]. China Ocean News, 2015-07-08 (003).
[9] Data governance is a data management concept—Wikipedia[EB/OL]. [2020-07-20]. http://www.esensoft.com/industry-news/data-governance-899.html.
[10] GORDON K. Principles of data management: facilitating information sharing[M]. 2nd edition. Swindon: BCS Learning and Development Ltd, 2013: 57.
[11] AL-RUITHE M, BENKHELIFA E, HAMEED K. A systematic literature review of data governance and cloud data governance[J]. Personal and ubiquitous computing, 2019, 23(5/6): 839-859.
[12] Enterprise Data Management Council. Data management capability assessment model (DCAM)[EB/OL]. [2020-07-20]. https://dgpo.org/wp-content/uploads/2016/06/EDMC_{DCAM}-*{{{WORKING}}*{DRAFT}}}*{VERSION}0.7.pdf.
[13] IBM. What is data governance?[EB/OL]. [2020-07-20]. https://www.ibm.com/analytics/data-governance.
[14] SEINER R S. Non-invasive data governance: the path of least resistance and greatest success[M]. Basking Ridge: Technics Publications, 2014: 21-24.
[15] MOSLEY M, BRACKETT M, EARLEY S, et al. The DAMA guide to the data management body of knowledge (DAMA-DMBOK)[M]. Bradley Beach: Technics Publications, 2009: 19-37.
[16] BHANSALI N. The role of data governance in an organization[M]//BHANSALI N. Data governance: creating value from information assets. Boca Raton: CRC Press, 2014: 1-18.
[17] ZHANG J. Operationalizing data quality through data governance[M]//BHANSALI N. Data governance: creating value from information assets. Boca Raton: CRC Press, 2014: 65-92.
[18] BRINEY K. Data management for researchers: organize, maintain and

share your data for research success[M]. Exeter: Pelagic Publishing, 2015: 6.
[19] SPACEY J. Data governance vs data management[EB/OL]. [2020-07-20]. https://simplicable.com/new/data-governance-vs-data-management/.
[20] The Data Governance Institute. Goals and principles for data governance[EB/OL]. [2020-07-20]. http://www.datagovernance.com/adg{{{data}}}{{governance}}}{goals}/.
[21] COHN B L. Data governance: a quality imperative in the era of big data, open data and beyond[J]. I/S: a journal of law and policy for the information society, 2015, 10(3): 811-826.
[22] THOMAS G. The DGI data governance framework[EB/OL]. [2020-07-20]. http://www.datagovernance.com/wp-content/uploads/2014/11/dgi_{framework}.pdf.
[23] The Expert Advisory Group on Data Access (EAGDA). Governance of data access[R/OL]. [2020-07-20]. https://wellcome.ac.uk/sites/default/files/governance-of-data-access-eagda-jun15.pdf.
[24] KHATRI V, BROWN C V. Designing data governance[J]. Communications of the ACM, 2010, 53(1): 148-152.
[25] MANDULA K. What is the difference between data governance and data management[EB/OL]. [2020-07-20]. http://www.beckerassociates.ca/the-difference-between-governance-and-management/.
[26] Fan Zhenjia. Enterprise scientific data governance and stewardship subjects: practice and analysis of NSR[J]. Library and Information Service, 2017, 61(1): 56-63.
[27] General Office of the State Council. Notice on issuing the Scientific Data Management Measures[EB/OL]. [2020-07-20]. http://www.most.gov.cn/mostinfo/xinxifenlei/fgzc/gfxwj/gfxwj
[28] Xing Wenming, Hong Cheng. Openness as the norm, non-openness as the exception—interpreting scientific data sharing and utilization in the Scientific Data Management Measures[J]. Library Tribune, 2019, 39(1): 117-124.
[29] Chinese Academy of Sciences. Notice on issuing the CAS Scientific Data Management and Open Sharing Measures (Trial)[EB/OL]. [2020-07-20]. http://www.cas.cn/tz/201902/P020190220358041915907.pdf.

Author Contributions:
Sheng Xiaoping: paper writing and revision;
Song Dacheng: data collection and content analysis.

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