Scientific Data

https://doi.org/10.1038/s41597-025-06525-0

Article in Press

Construction of a Theoretical Framework for Scientific Data Governance

Received: 15 July 2025

Accepted: 24 December 2025

Cite this article as: Qiu, Y., Hu, Z. Construction of a Theoretical Framework for Scientific Data Governance. *Sci Data* (2025). https://doi.org/10.1038/s41597-025-06525-0

Yanrui Qiu & Zhimin Hu

Construction of a Theoretical Framework for Scientific Data Governance

Yanrui Qiu[1], Zhimin Hu[1]

Affiliation:

[1] School of Health Policy and Management, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China.

Corresponding Author:

Dr. Zhimin Hu: huzhimin@pumc.edu.cn, School of Health Policy and Management, Chinese Academy of Medical Sciences & Peking Union Medical College, No. 9 Dongdan Santiao, Dongcheng District, Beijing 100730, China.

Abstract

The advancement of data-intensive sciences and artificial intelligence-driven sciences has introduced governance challenges for multi-source heterogeneous scientific data across diverse scenarios. Given the intricate entanglement of stakeholders, processes, and content in scientific data governance, this study intends to propose a theoretical framework to elucidate its complex dynamics and inform governance practices. The theoretical framework for scientific data governance consists of three core dimensions: data stakeholders, data lifecycle, and data governance elements. Non-systematic literature review was employed to identify the classification of data stakeholders and data lifecycle, and bibliometric analysis was used to extract the elements of scientific data governance. Meanwhile, based on the elements of data governance, five governance systems have been summarized, including organizational operation system, technical support system, risk prevention and control system, value realization system, and regulatory system.

Introduction

Contemporary society is undergoing a data revolution, witnessing an unprecedented expansion of data ecosystems within various organizations. Exponentially growing multi-source heterogeneous data streams have comprehensively permeated critical public sectors including healthcare systems, agricultural economies, intelligent manufacturing, judicial administration, smart transportation networks and scientific research. As human society enters the era of the data big bang[1], the scientific research paradigm has undergone profound changes. It has transitioned from the traditional empirical paradigm, theoretical paradigm, and simulation paradigm to the fourth paradigm driven by data[2], and is evolving towards the trend of the fifth paradigm transformation driven by artificial intelligence (AI).

Data is a discrete, limitless entity that has an unstructured and unprocessed shape，while big data is characterized by high volume, veracity, velocity, and variety [3]. The concept of "big data" first emerged in the late 1990s[4] and was defined in the early 2000s in terms of the 3Vs model, which refers to Volume, Velocity, and Variety[5]. Subsequently, the dimensions of Value and Veracity were introduced[6]. However, with the continuous advancements in artificial intelligence, the Internet of Things, omics technologies, etc., the traditional paradigm encapsulated by the term "big data" is no longer sufficient to accurately describe the complexities of the current data ecosystem. Currently, data sources are extensive and diverse in form, encompassing vast amounts of unstructured and semi-structured data, such as images, videos, and text. The core challenge of data has shifted from its initial "scale" to "heterogeneity" and "complexity." The transformation in the nature of data forms has created an urgent need for data governance theories capable of addressing multi-source heterogeneous data environments.

Scientific data refers to datasets generated through fundamental research, applied research, and experimental development in natural sciences and engineering disciplines, as well as raw observational records and derived datasets obtained via monitoring systems, field investigations, experimental testing, and detection processes that serve scientific inquiry. In essence, scientific data encompasses both data generated through scientific research and data utilized for scientific purposes, the latter of which includes data originally collected for non-scientific objectives but subsequently employed in research. The scientific data referred to in this study is synonymous with research data.

Throughout the evolution of data governance, several principles have emerged to provide direction and value orientation for data governance activities. The FAIR principles — encompassing Findable, Accessible, Interoperable, and Reusable—have established foundational criteria for data governance, thereby promoting open data sharing and enhancing the value of data. However, ethical issues such as data ownership after FAIRification, privacy protection of sensitive data, and informed consent for data reuse have not been adequately addressed. The introduction of the CARE principles has, to some extent, remedied the ethical shortcomings of the FAIR framework. The CARE principles emphasize Collective Benefit, Authority to Control, Responsibility, and Ethics. Its core philosophy shifts the focus from a "data-centric" approach to a "human-centric" one, imposing stringent requirements on the data governance ecosystem with ethics as a central thread.

Additionally, the TRUST principles — Transparency, Responsibility, User Focus, Sustainability, and Technology—propose core evaluation criteria for trusted digital repositories, thereby complementing the FAIR principles' lack of focus on the long-term preservation of scientific data. Meanwhile, the 4P model—Data as a Product, Data Platform, Data People, and Data Process—aims to further advance the assetization of data and enhance its commercial value. The proliferation of diverse data governance principles reflects the complex and evolving nature of data governance issues. Clarifying the core elements and specific dimensions of data governance is therefore of critical importance.

The advancement of data-intensive sciences and AI-driven sciences has introduced governance challenges for multi-source heterogeneous scientific data across diverse scenarios. Throughout their entire lifecycle from creation to destruction, scientific data are transferred among different organizations and stakeholders. While these entities employ various tools to standardize and normalize data to unlock and utilize its value, the data circulation process frequently triggers risks such as privacy breaches, data misuse, and data monopolization. How to achieve legally compliant and regulatory-aligned governance of the data lifecycle under the joint influence of multiple stakeholders, as well as what aspects and elements are included in effective governance, are pressing issues that need to be addressed. To address these gaps, this study aims to synthesize insights from existing literature on scientific data governance — focusing on stakeholder dynamics, lifecycle management, and governance elements — to construct a theoretical framework with governance systems. This framework seeks to provide theoretical foundation for scientific data governance and guide operational practices in this domain.

Definition of Data Governance

Data governance is highly scene-oriented. It varies significantly across domains — such as government and corporate data governance — based on the specific subject, objectives, and context[7]. The conception and description of data governance vary depending on the specific scenario. To date, a universally accepted definition of data governance has not yet been established.

Earlier on, some definitions of data governance were influenced by information and communications technology governance. In scientific literature, the terms "information" and "data" could be used interchangeably. As a result, academic sources mainly followed Weill and Ross's[8] definition of information governance, which is "specifying the decision rights and accountability to encourage desirable behavior in the use of information technology". Data governance was defined as a framework for specifying decision rights and accountability to encourage good behavior in the use of data[9]. Sunil Soares[10] argued that big data governance is part of a broader information governance program that formulates policy relating to the optimization, privacy, and monetization of big data by aligning the objectives of multiple functions. 2013 was generally regarded as the "Year of Big Data". Before that, scholars mostly viewed data governance as a derivative or subset of information governance. However, some practitioners believed that "data" and "information" were distinct concepts, where data constitutes simple facts, while information represents data contextualized within specific frameworks or processed data[11].

The use of data governance in the research context has been implemented within academic and government agency programs for many decades. For instance, the Human Genome Project (1990s) established mandatory data sharing policies, internationally unified data formats, and dedicated data repositories (e.g., GenBank), which enabled effective management and global sharing of large-scale scientific data. Similarly, Planetary Data System (established in 1989) of National Aeronautics and Space Administration (NASA) introduced rigorous data archiving standards, metadata specifications, and long-term preservation mechanisms to ensure continued accessibility and reusability of valuable space science data. Thia reflects its core concerns of data governance in data lifecycle and quality management. Dimensions such as data collection protocols, metadata management, data quality control, and data standardization have long been integral to scientific domains, even if not always explicitly labeled as "governance". Moreover, data governance practices are often codified through data policies and overseen by roles such as the chief data officer. A notable recent contribution in the field of research data

management is the *NIST Research Data Framework (RDaF)*[12] published by the National Institute of Standards and Technology (NIST). The RDaF employs a lifecycle approach structured around six high-level stages to organize essential information related to research data management and dissemination. It also incorporates fourteen overarching themes, collectively aiming to help shape the future of research data management and open data access.

In recent years, most definitions of data governance have primarily focused on decision rights (who has the authority to make decisions related to data) and accountability (who is responsible for data-related decisions) in a business environment, rather than in a research context[13]. For instance, Sergi Nadal[14] argued that data governance refers to what decisions must be made to ensure effective data management and usage, as well as who makes these decisions. Al-Badi, Ali[15] defined big data governance as the management of huge volumes of an organization's data, exploiting it in the organization's decision-making using different analytical tools. Zhang[16] considered data governance as the standardized management of legitimate data through the intervention of certain procedures, rules, and even values, transforming it into a strategic asset for the enterprise. A more representative view is that put forward by the Data Governance Institute (DGI)[17], which states that Data Governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods. Such perspectives are more applicable to the governance of business data generated in corporate operations and place greater emphasis on the value attributes of data as an asset.

Furthermore, different research perspectives emphasize varying priorities in data governance. Some studies focus on the supervision and control of data resource processing. For example, Fothergill[13] et al. regarded data governance as a strategy for the overall management of the availability, usability, integrity, quality, and security of data in order to ensure that the potential of the data is maximized whilst regulatory and ethical compliance is achieved within a specific organizational context. The *DAMA Guide to the Data Management Body of Knowledge* defines data governance as "planning, oversight, and control over management of data and data-related resources" [18]. Some other research emphasizes governance across the entire data lifecycle. For instance, Jang, Kyoung-ae[19] argued that data governance involves functions that enable organizations to ensure high data quality throughout the entire data lifecycle. Damian O. Eke[20] defined data governance as the principles,

procedures, frameworks, and policies that ensure acceptable and responsible processing of data in each stage of the data lifecycle; from collection, storage, processing, curation, sharing and use, to deletion. There are also studies emphasize quality control specifications within data governance. For instance, the International Business Machines Corporation (IBM) Data Governance Council[21] defines data governance as a data management discipline focused on the quality, security, and availability of organizational data, ensuring data integrity and security through the definition and implementation of policies, standards, and procedures.

Building upon the aforementioned perspectives, this paper posits that the definition of data governance should encompass essential elements including governance objects, governance processes, and governance objectives. This study primarily adopts the definition established in the *Data Governance Standardization White Paper* released by the China Communications Standards Association (CCSA)[22], which conceptualizes data governance as the implementation of effective management and control throughout the entire lifecycle of diverse data categories-including personal data, enterprise data, government data, and public data-through a systematic framework comprising laws and regulations, management systems, standard specifications, and technological tools. This governance framework aims to fulfill data application requirements across multiple scenarios such as enterprise management, industry regulation, national governance, and international collaboration. Scientific data governance can accordingly be defined as the effective management and control of the entire lifecycle of scientific data through a systematic framework comprising laws and regulations, management systems, standard specifications, and technological tools, in order to meet the requirements of data application in scientific research scenarios.

Methods

This study employs a non-systematic literature search to identify a series of theoretical frameworks for data governance (table 1). By reviewing existing frameworks, this study provides insights to inform the development of a scientific data governance framework. These theoretical frameworks vary in their emphases. For instance, the framework of the IBM Data Governance Council[21] and the framework by Zhang[16] and Kyoung-ae[19] both focus on enterprise data governance, while the framework of Data Management Association International (DAMA) centers on specific data management practices[18]. The framework of NIST is designed to help research data management and open sharing[12]. Kieran's framework is mainly applied to the governance of genomic data[23].

Table 1. Existing Data Governance Theoretical Frameworks

Source Publication	Components of Data Governance
The DAMA Guide to the Data Management Body of Knowledge[18]	data architecture; data modeling and design; data storage and operations; data security; data integration and interoperability; documents and control; reference and master data; data warehousing and business intelligence; meta-data; and data quality.
What is Data Governance?[21]	program goals, roles and duties; data standards, policies and processes; auditing procedures; data governance tools
DGI Data Governance Framework[17]	mission and value; beneficiaries of data governance; data products; controls; accountabilities; decision rights; policy and rules; data governance processes, tools, and communication; DG work program; participants
NIST Research Data Framework[12]	community engagement; cost implications and sustainability; culture; curation and stewardship; data quality; data standards; diversity, equity, inclusion, and accessibility; ethics, trust, and the care principles; legal considerations; metadata and provenance; reproducibility and the fair data principles; security and privacy; software tools; training, education, and workforce development
Data Governance: A Conceptual Framework, Structured review, and Research Agenda[24]	① governance mechanisms: structural, procedural, and relational mechanisms ② organizational scope: within-organization and inter-organization scope ③ data scope: traditional data and big data ④ domain scope: data quality, data security, data architecture, data lifecycle, metadata, and data storage and infrastructure
Data Matters: A Strategic Action Framework for Data Governance[16]	design of governance mechanisms; investment in data technologies; data collaboration; perceptions of data potential; developing data-related capabilities; establishment of data legitimacy
Towards Better Governance of Human Genomic Data[23]	enabling data access; compliance with applicable national laws and international agreements; supporting appropriate data use and mitigating potential harms; promoting equity in the access to, and use and analysis of genomic data; using genomic data for public benefit
Development of Data Governance Components Using DEMATEL and Content Analysis[19]	① data compliance: strategy, policy, methodology, metrics, system architecture ② data management: data lifecycle, data monitoring, quality management, quality value, data security ③ data organization: role and responsibility, organizational structure
Designing Data Governance[25]	data principles; data quality; metadata; data access; data lifecycle

Demystifying Data Governance for Process Mining: Insights from a Delphi Study[26]	data architecture; data integration and interoperability; data modelling and design; data quality; data security; data storage and operations; data warehousing and business intelligence; documents and content; meta-data; reference and master data; supporting organizational policies and programs
Coordinating Decision-Making in Data Management Activities: A Systematic Review of Data Governance Principles[9]	① organization: decision rights, balanced roles, stewardship, ownership, separation of duties and concern, improved coordination of decision making ② alignment: meeting business needs, aligning business and IT, developing data strategy, defining data quality requirements, reducing error of use, effective policies and procedures ③ compliance: accountability, policy enforcement, due diligence, privacy, openness, security, data quality measurement ④ common understanding: shared data commons, use of standards, metadata management, standardized data models, standardized operations, facilitates communication
A Contingency Approach to Data Governance[11]	data quality management roles; decision areas; main activities; responsibilities
Exploring Big Data Governance Frameworks[15]	identify organizations structure; stakeholders selection; big data scope determination; policies and standards setting; optimize and compute; measure and monitor quality; data storage; communication and data management

Despite differences in the construction across theoretical frameworks, some commonalities can be extracted. The frameworks of the IBM Data Governance Council[21], the DGI[17], Kyoung-ae[19], Paul[9], Kristin[11], and Ali[15] all address roles or stakeholders in data governance. Similarly, frameworks by the IBM Data Governance Council[21], NIST[12], Rene[24], Kyoung-ae[19], Vijay[25], and Paul[9] incorporate discussions of processes or data lifecycles. Meanwhile, data quality management, data standard setting, metadata management, organizational structure design, data security maintenance, and data value release can be regarded as the content elements included in good data governance. Thus, a generalized data governance framework can be conceptualized as an interconnected and interactive system composed of stakeholders, the data lifecycle, and data governance elements.

In the dimensions of stakeholders and lifecycle, the classification of scientific data governance aligns closely with that of general data governance and demonstrates relatively fixed categorization. Therefore, this study

primarily draws upon existing scholarly literature regarding classifications within these two dimensions. However, the categorization of data governance elements shows high relevance to data sources and application scenarios. To address this characteristic, this research employs bibliometric methods to systematically extract scientific data governance elements, thereby constructing a more comprehensive and objective theoretical framework.

The Web of Science, recognized as the most comprehensive multidisciplinary academic resource globally, covers over 12,000 core scholarly journals and is widely regarded as a reliable and authoritative source for academic research[27, 28]. Consequently, it was selected as the primary database for this bibliometric analysis.

This study utilized the Web of Science Core Collection as the literature retrieval database. To ensure the quality of the literature, the indexed sources were restricted to the Science Citation Index Expanded (SCI-EXPANDED) and the Social Sciences Citation Index (SSCI). The retrieval formula was TS = (“scientific data” OR “research data”) AND TS = (“governance” OR “sharing”) AND TS = (“framework”), with the publication year limited to 2000–2024. The initial search yielded 7,624 publications. The literature type was limited to "Articles," and the language was restricted to "English," resulting in the final inclusion of 6,525 publications. Detailed inclusion and exclusion criteria are illustrated in Figure 1 (Figure 1 goes here). The list of included articles and other detailed information can be found in Supplementary Table S1.

The selected literature was exported as plain text files and imported into Thomson Data Analyzer (TDA) version 3.0 for keywords extraction. Keywords were derived from “combined keywords and phrases”, including authors’ keywords, Keywords Plus, and phrases from titles and abstracts. Keywords records ≥50 times were retained for analysis. Synonyms were manually merged, and nonsensical terms were removed. The keywords were categorized into several governance elements based on practical relevance. Meanwhile, the frequency and percentage of each governance element was calculated.

Results

Stakeholders of Scientific Data Governance

Stakeholder theory was defined as "all individuals or groups who can affect the achievement of an organization's objectives or are affected by the pursuit of those objectives"[29]. Stakeholder theory constitutes a framework for

understanding how diverse stakeholders interact to co-create and exchange value[30]. Stakeholders in scientific data governance and open sharing refer to individuals or organizations engaged in activities related to scientific data governance and open sharing.

Smith et al. [31] classified key roles within data-sharing systems into three categories: 1) data-sharing objects, which encompass data of varying granularity and types; 2) data-sharing actors, including providers (those contributing data) and consumers (those utilizing data); and 3) data-sharing facilitators, such as distributors offering value-added services and entities responsible for establishing data standards or developing technical tools. Wang[32] identified stakeholders in research data governance as data producers, storers, users, disseminators, and policymakers. Similarly, Gao et al.[33]categorized stakeholders in medical data sharing into data providers, medical-data-sharing platforms, and data demanders, emphasizing the role of intermediaries in specialized contexts. By synthesizing these perspectives and considering the unique characteristics of scientific data, this study categorizes the principal stakeholders in scientific data governance into four distinct groups: data providers, data users, data sharing facilitators, and policymakers.

Several international organizations and alliances have compiled a list of individuals and organizations associated with data governance and open sharing activities. The *OECD Principles and Guidelines for Access to Research Data from Public Funding*[34] systematically identifies critical actors, encompassing researchers, research institutions, funding agencies, government agencies, data archives, academic associations, private sector, and user groups. The European Commission's *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*[35] outlines the range of stakeholders, incorporating researchers, publishers, funding agencies, research institutions, repository managers, enterprises, citizens, and policymakers. Similarly, the League of European Research Universities (LERU) *Roadmap for Research Data*[36] articulates a structure involving researchers/data owners, data scientists/data stewards, libraries, and management/faculty/administration, along with external service providers. Furthermore, Science International's *Open Data in a Big Data World*[37] conceptualizes stakeholder roles across scientific ecosystems, including scientists, research institutions/universities, publishers, funding agencies, professional associations/scholarly societies/academies, and libraries/archives/repositories.

From these documents, it is evident that data governance practitioners primarily include researchers and research

institutions (encompassing universities, institutes, hospitals, and academies), government bodies, enterprises, the public, research funding agencies, publishers, data centers, information centers, libraries, archives, repositories, external service providers, and professional associations. These stakeholders collectively contribute to and are impacted by data governance frameworks, underscoring the multifaceted nature of effective data management and governance in the research ecosystem.

Table 2 presents the stakeholder classifications of scientific data governance and open sharing along with their representative individuals and organizations. Note that some individuals and organizations may have multiple stakeholder identities. For example, researchers and research institutions are both data providers and data users.

Table 2. Classification of Stakeholders in Scientific Data Governance

Stakeholders	Representative Individuals and Organizations
data providers	Researchers and research institutions, enterprises, the public, etc.
data users	Researchers and research institutions, government bodies, enterprises, the public, etc.
data sharing facilitators	Research funding agencies, publishers, data centers, information centers, libraries, archives, repositories, external service providers, etc.
policymakers	Government bodies, professional associations, etc.

Lifecycle of Scientific Data Governance

The inherent characteristics of data, which include ease of sharing, replication, and recombination, endow them with potential reusability. However, the precondition for unlocking the value of data is the proper collection, processing, and preservation of data. Data loss or damage may lead to substantial economic costs and missed opportunities. Therefore, a data management plan based on the entire lifecycle of data is of vital importance[38]. A comprehensive data lifecycle management framework can be used to plan, organize, and manage the entire lifecycle of data from acquisition to disposal, as well as the relationships between these sequential phases[3].

The life cycle refers to the biological process from birth through growth, senescence, to demise[39]. The concept of "information life cycle" was proposed by Marchand and Horton in 1986, which was divided into six stages: creating, collecting, organizing, developing, utilizing, and purging information[40]. Given the strong

relationship between data and information, the advancement of data science has subsequently led scholars to propose the concept and models of the data lifecycle.

The data lifecycle is the sequence of stages that data follow from the moment they enter a system to the moment they are erased from the system[41]. The lifecycle of scientific data is conceptualized as a theoretical framework encompassing the entire process from data generation through development and maturation until its eventual obsolescence or deletion[42]. During this journey, data passes through different stages, which vary among scholars' models depending on data type, field, and objectives, and even the terminology for the same phase may differ[3]. Currently, no all-inclusive data lifecycle model has been found yet.

Multiple academic and institutional perspectives have contributed to data lifecycle conceptualization. Table 3 presents several existing classifications of data lifecycle. IBM's framework emphasizes five core phases: creation, storage, sharing/usage, archival, and deletion[43]. Damian[20] defined data governance with a lifecycle of collection, storage, processing, management, sharing, use, and destruction. Fothergill[13] divided the data lifecycle into collection, processing, management, sharing, application, and destruction. Kumar et al.[44] considered data lifecycle management phases as production, storage, use, sharing, archiving, and destruction or reuse. Shameli-Sendi[45] noted the data lifecycle includes generation, modification, processing, transmission, presentation, and final storage. Blazquez et al. [38] present a granular nine-stage framework encompassing study/planning, collection, documentation/quality assurance, integration, preparation, analysis, publishing/sharing, storage/maintenance, and reuse.

Specialized research data lifecycle models have also emerged. NIST[12] identified six lifecycles of research data management: envision, plan, generate/acquire, process/analyze, share/use/reuse, preserve/discard. Jetten et al.[46] considered the research data lifecycle as involving planning, creating, processing/analyzing, using, preserving, and accessing research data. Ball's Research 360 model[47] focuses on scientific research data through design, collect/capture, interpret/analyze, manage/preserve, release/publish, discover and reuse phases.

Table 3. Existing Classification of Data Lifecycle

Source Publication	Classification of Data Lifecycle
What is data lifecycle management	creation, storage, sharing/usage, archival, deletion

(DLM)? [43]		
International Data Governance for Neuroscience[20]	collection, storage, processing, management, sharing, use, destruction	Building upon these theoretical
Responsible Data Governance of Neuroscience Big Data[13]	collection, processing, management, sharing, application, destruction	
Data Life Cycle Management in Big Data Analytics[44]	production, storage, use, sharing, archiving, destruction/reuse	
An Efficient Security Data-Driven Approach for Implementing Risk Assessment[45]	generation, modification, processing, transmission, presentation, final storage	
Big Data Sources and Methods for Social and Economic Analyses[38]	study/planning, collection, documentation/quality assurance, integration, preparation, analysis, publishing/sharing, storage/maintenance, reuse	
NIST Research Data Framework (RDaF)[12]	envision, plan, generate/acquire, process/analyze, share/use/reuse, preserve/discard	
The Role of CRIS's in the Research Life Cycle[46]	planning, creating, processing/analyzing, using, preserving, accessing research data	
Review of Data Management Lifecycle Models[47]	design, collect/capture, interpret/analyze, manage/preserve, release/publish, discover, reuse	

foundations and considering the unique attributes of scientific data, this study proposes a refined lifecycle framework comprising seven critical phases: data collection, data storage, data processing, data management, data sharing, data application, and data deletion. Scientific data originates from various research entities. The approach of "centralized management and distributed storage" will facilitate effective governance of such data. The head scientific data center can conduct unified processing and management of scientific data stored in different subcenters and nodes, which can achieve hierarchical classification, standardization, and quality improvement of multi-source heterogeneous data. The processed data can then be shared securely, applied in various scenarios to increase its value, and eventually deleted.

Elements of Scientific Data Governance

To clarify the key elements of scientific data governance, this study employs bibliometric analysis to extract keywords from relevant literature on scientific data governance frameworks and subsequently categorizes and

counts the frequency of these keywords. The classification results are presented in Table 4.

Table 4. Classification of Keywords for the Elements of Scientific Data Governance

Elements	Frequency / Percentage	Combined Keywords and Phrases
technological infrastructure	3898 / 11.43%	internet; tools; network; technologies; machine learning; infrastructure; deep learning; artificial intelligence; federated learning; web; internet of things; blockchain technology; software; federated learning; cloud computing; algorithm; devices; sensors; information technology; sites
data resources	3208 / 9.41%	information; knowledge; resources; factors; value; database; elements; products; data sources
public attitudes	2552 / 7.48%	understanding; trust; perspective; perception; attention; response; will; public; attitudes; view; awareness; expectations; opinions
operation mechanism	2333 / 6.84%	system; collaboration; coordination; culture; mechanism; cooperation; accountability; operation; function; maintenance; autonomy
organizational structure	2113 / 6.20%	community; environment; organizations; platform; institutions; structure; architecture; setting; building; establishment
data quality control	1962 / 5.75%	quality; accuracy; assessment; review; reliability; validation; validity; robustness; data quality; usability; surveillance; usefulness; evaluation; consistency
data standards and specifications	1951 / 5.72%	variety; wide range; classification; diversity; indicators; interoperability; standards; reproducibility; degree; criteria; applicability; heterogeneity; different levels; different types; taxonomy
talent team building	1943 / 5.70%	role; training; characteristics; practitioners; members; leadership; managers; staff; motivation; expertise; professionals; incentives; recruitment
data ontology construction	1906 / 5.59%	relationship; concept; scope; relevance; association; domain; definition; ontology; field; species; constraints; correlation
data services	1689 / 4.95%	application; services; thematic analysis; visualization; serve; consumers; data mining; content analysis
access control	1612 / 4.73%	approach; barriers; access; control; accessibility
data security	1296 / 3.80%	security; risk; safety; prevention; vulnerability; data security; safe
funding sources	1241 / 3.64%	support; cost; contribution; foundation; account; provision; commitment; investment; amount; funding

policies and regulations	1179 / 3.46%	policy; guidelines; health policy; regulation; policy makers
privacy protection	1170 / 3.43%	privacy; identification; conservation; data privacy; confidentiality; protection; privacy protection
metadata management	955 / 2.80%	datasets; form; sample; group; metadata; categories; feature extraction
data ownership	928 / 2.72%	rights reserved; interest; ownership; benefits; knowledge management; empowerment
data circulation	904 / 2.65%	dissemination; information sharing; delivery; knowledge sharing; transition; exchange; transformation; transmission; mobility
ethics framework	619 / 1.82%	ethics; principles; protocol; integrity; disciplines
informed consent	323 / 0.95%	consensus; consent; agreement; informed consent
data fairness	318 / 0.93%	transparency; equity; average

Scientific data governance constitutes not a collection of discrete components, but rather an integrated organism comprising processes, entities, contents, and their dynamic interrelationships. Only when these interactions evolve into institutionalized operational systems can they provide stable and sustainable guidance for scientific data governance practices. Accordingly, based on the functional differences of scientific data governance elements, this paper reorganizes the governance elements, forming five scientific data governance systems — organizational operation system, technical support system, risk prevention and control system, value realization system, and regulatory system.

Organizational Operation System

The governance and open sharing of scientific data necessitate reliance on physical organizations, such as scientific data centers, information hubs, data sharing portals, and biobanks. These organizations, on the one hand, can aggregate scientific data to make it findable, accessible, interoperable, and reusable (FAIR). On the other hand, they can offer human resources, financial support, and regulatory accountability mechanisms for effective data governance.

In response to exponential data growth and the global consensus on open science, the international community has established numerous scientific data centers. Notable examples include the three global biological data

centers established during the 1980-1990s: the National Center for Biotechnology Information (NCBI) in the United States, the European Bioinformatics Institute (EBI), and the DNA Data Bank of Japan (DDBJ). These institutions have significantly advanced biomedical research through standardized governance and sharing of third-party data resources. Furthermore, initiatives like The Cancer Genome Atlas (TCGA) the UK Biobank demonstrate hierarchical data sharing models leveraging large-scale research project outputs.

Contemporary scientific data governance has entered a phase characterized by systematic and sustainable competition. Future development trajectories for scientific data centers will manifest three distinct trends: 1) Evolution of governance architectures toward a “head center—subcenter—node” model, 2) Transition of technical infrastructure to AI-driven platforms, and 3) Expansion of value realization into knowledge services and data elements circulation. To sustain progress, it is imperative to strengthen institutional capacity building through stable talent recruitment and diversified funding mechanisms. Concurrently, the implementation of robust operational maintenance protocols, comprehensive supervision frameworks, and transparent accountability systems remains crucial to ensure legal and regulatory compliance throughout data sharing processes.

A robust accountability system is crucial for data governance institutions. A sound accountability framework begins with clearly defined roles and responsibilities, relies on transparent operations and continuous oversight, is safeguarded by performance evaluations and periodic audits, and achieves a closed management loop through effective enforcement and remediation. At the practical level, to fulfill ethical review and data protection obligations, the United States commonly establishes Institutional Review Boards (IRBs), China has implemented an ethical review committee system, and the European Union has developed a multi-tiered regulatory architecture through the European Data Protection Board (EDPB) and Data Protection Officers (DPOs) embedded in various institutions.

Technical Support System

Scientific data governance requires an effective technical support system, encompassing both software and hardware infrastructure, as well as technical tools for data management. Furthermore, this study underscores that modern data governance must transcend traditional models reliant on manual audits and evolve into a technology-empowered "Governance-by-Design" paradigm embedded within entire workflow.

Hardware infrastructure comprises physical facilities such as high-capacity storage devices, high-performance computing systems, virtual desktop infrastructure, multi-tiered power disaster recovery architectures, local backup facilities, and remote disaster recovery centers.

Software infrastructure encompasses digital intelligence technologies and practical tools that facilitate efficient governance throughout the entire data lifecycle. For instance, during the data collection phase, structured consent frameworks and term-based agreements are employed to define the scope and permissions of data gathering. In the data storage phase, distributed storage and cloud computing technologies provide highly available and scalable storage capabilities. Additionally, a data asset view and data lineage chain centered on metadata management are established, integrated with blockchain technology to record key operations, forming an end-to-end, tamper-proof traceability system. During data processing, federated learning, secure multi-party computation, and homomorphic encryption enable analytical models under the "usable yet invisible" principle, while privacy-enhancing technologies such as differential privacy, data desensitization, and anonymization are employed to mitigate the risk of sensitive data leakage. Throughout data management, behavioral standards, training programs, and accountability mechanisms are implemented to enforce data subject responsibility. Meanwhile, a regulatory center is deployed to convert institutional provisions into machine-executable code, enabling automatic early warning and disposal of behaviors that violate preset policies. In the data sharing phase, hierarchical and categorical access clauses, data sharing agreements, and data access committees serve as critical safeguards for secure and compliant data sharing.

Data governance technical tools are specialized methodologies for standardized data management to improve reusability and operability. These tools primarily include data ontology construction, metadata management, data quality control, and data standards and specifications. Data ontology construction aims to create a structured, formalized knowledge model that describes conceptual relationships and rules in specific domains. Metadata, defined as "data about data," describes characteristics such as data attributes, structure, provenance, usage, and quality. Metadata management involves the creation, storage, maintenance, sharing, and utilization of metadata. Data quality control employs systematic methods, tools, and workflows to ensure data accuracy, completeness, consistency and reliability. Data standards and specifications refer to a set of rules and standards formulated for the definition, structure, format, encoding, storage and use of data.

Providing robust software and hardware environments for scientific data governance, coupled with continuous improvements in standardized data processing, serves as a critical pathway to enhance the intrinsic value of scientific data and ensure the sustainable operation of scientific data platforms.

Risk Prevention and Control System

Scientific data faces diversified and complex risks during open sharing processes, necessitating a robust risk prevention and control framework to enhance governance. This framework primarily encompasses the following dimensions: public attitudes, access control, data security, privacy protection, informed consent, and data fairness. Establishing a comprehensive risk mitigation system is critical to safeguarding the rights of data contributors and promoting legally compliant scientific data sharing.

Public attitudes directly influence the collection and sharing of scientific data, particularly for data related to health and genomics, which often contain personally identifiable information. Indiscriminate collection or sharing may lead to severe ethical issues, including privacy breaches and individual discrimination. Consequently, obtaining informed consent from data subjects during collection is imperative. Equally essential are robust measures to ensure data security and privacy protection throughout data processing, sharing, and application. To prevent data misuse and privacy violations, scientific data sharing should adopt a graded classification system with access control mechanisms implemented on sharing platforms. Open sharing models may be applied to anonymized datasets devoid of identifiable information, while restricted access or application-based review protocols should govern datasets with re-identification risks. Finally, scientific data governance and open sharing aim to foster a more equitable and well-regulated data ecosystem. Vigilance is required to address inequities such as the data divide and data monopolies, ensuring fairness in data accessibility and utilization.

Value Realization System

The ultimate objective of scientific data governance and open sharing lies in unleashing data value, which facilitates the application of science, technology, and data to advance human well-being and drive societal development. The data value realization system originates from data resources, materializes through data circulation and service, and is rooted in the delineation of data ownership.

Data resources are the source of data value. In this study, data resources encompass data products, data assets, and data elements, which represent different evolutionary forms of data. Once raw data are processed into value-laden resources, they enter circulation and generate diverse data services. However, the circulation and service phases involve tricky benefit distribution challenges, where conflicts among stakeholders may lead to chaotic practices. To mitigate risks and inequities, clear data ownership delineation is imperative. Rational ownership demarcation clarifies the rights and obligations of data stakeholders, ensuring their activities remain subject to legal oversight and accountability. Available reference approaches for data ownership demarcation include the tripartite rights separation (encompassing data resource ownership rights, data processing and usage rights, and data product operation rights) and the data value dichotomy (distinguishing between data transformer rights and data integrator rights), among others.

Regulatory System

The regulatory system serves as the top-level design for scientific data governance. The legality and compliance of data collection, processing, sharing, and application activities constitute the prerequisite for stakeholders' engagement in data governance processes, while the development and refinement of regulatory system adapt to the practical needs of data governance. The regulatory system for scientific data governance primarily encompasses policy regulations and ethical frameworks.

Globally, several representative policies and regulations have been promulgated in response to the issue of data governance. These include the European Union's *General Data Protection Regulation* and *Data Governance Act*, the United States' *National Security and Personal Data Protection Act*, *Genomic Data Sharing Policy*, and *Data Management and Sharing Policy*, the United Kingdom's *Data Protection Act*, as well as Germany's *Federal Data Protection Act* and *Recommendations on Data and Algorithms*.

Concurrently, substantial advancements in data ethics governance have emerged worldwide. Notable frameworks include the United Kingdom's *Data Ethics Framework*, the United States' *Federal Data Strategy: Data Ethics Framework*, Germany's *Opinion of the Data Ethics Commission*, and Switzerland's *Ethical Framework for Responsible Data Processing in Personalized Health Research*. These policies, regulations and ethical frameworks provide operational paradigms and normative references for global data governance practices, reflecting evolving societal expectations regarding responsible data stewardship.

Discussion

Through the identification of data stakeholders and data lifecycle, along with the extraction and integration of data governance elements, this study proposes a final theoretical framework for scientific data governance, as shown in Figure 2 (Figure 2 goes here). The framework consists of three dimensions: the data lifecycle, data stakeholders, and data governance components, with the latter divided into five data governance systems. To achieve legally compliant governance of the entire data lifecycle under the joint influence of multiple stakeholders, one may refer to the scientific data governance framework proposed in this study. This framework emphasizes synergistic governance involving diverse stakeholders, full lifecycle management, and complex governance systems. This is elaborated as follows.

The synergy mechanism refers to the process, mode of action, and procedures through which elements or entities interact to achieve established or agreed upon goals[48]. Synergistic governance theory, a governance strategy that offers coordination to balance interests, emphasizes the self-organized activities of participating subsystems, the diversity of participating entities, and the contingency of governance methods[32]. The stakeholders, processes, and content involved in scientific data governance exhibit inherent diversity and complexity. Therefore, establishing collaborative governance characterized by the active participation of multiple stakeholders and the coordination of complex governance systems throughout the entire lifecycle of scientific data becomes critically essential.

Multi-stakeholder synergistic governance focuses on the common participation and equal consultation of stakeholders. Data governance practitioners should clarify their rights and obligations, enhance their awareness of synergistic governance, and safeguard privacy and security during data circulation. Policy makers should establish a clear data property rights system to reduce interest conflicts among stakeholders and create a harmonious data trading environment. When data governance is based on a particular platform, priority should be given to ensuring the fairness and inclusiveness of decision making and accountability mechanisms. Secondly, it is necessary to strengthen the integration and interoperability of data resources on scientific data platforms. Technologies like AI and blockchain can be used to enable the omni-domain circulation and sharing of scientific data, and unify data standards to reduce barriers to data sharing and utilization.

Scientific data governance necessitates the coordinated operation of sophisticated governance systems anchored in institutionalized organizations. On the one hand, the organizations require the foundational supports of sufficient human, financial, and material resources, as well as scientific management systems. On the other hand, advanced technological infrastructures capable of enhancing the quality and standardization of multi-source heterogeneous data are also needed. The ultimate objective resides in maximizing data value realization while rigorously ensuring legal compliance and risk control. Within this paradigm, organizational operation system and technical support system serve as the cornerstone, risk prevention and control system and regulatory system function as safeguards, while value realization system represents the targeted outcome. The synergistic integration of these systems collectively ensures comprehensive governance efficacy through structural coherence and functional complementarity.

The complex governance system for scientific data spans the entire lifecycle of data governance. However, the key focus of the governance system varies across each lifecycle stage, and the level of involvement of each governance system also differs throughout these stages. During the stages of data storage, processing, and management, the governance focus should be placed on the development of the organizational operation system and the technical support system, establishing standardized data operation procedures and a technology-enabled, embedded "Governance-by-Design" mechanism within operational workflows. In the data sharing phase, it is essential to strengthen the risk prevention and control system, ensuring security, compliance, and privacy protection during data circulation and external provisioning. The data utilization stage should leverage the value realization system to promote the value transformation and efficiency release of data elements, thereby supporting data-driven decision-making and data-empowered social governance. The regulatory system permeates all phases from data collection to destruction, providing institutional guarantees and ethical constraints for data governance, and must be consistently upheld as the baseline for all data governance activities.

The theoretical framework for scientific data governance constructed in this study is designed to encompass all entities and levels involved in scientific data processing activities, without being limited to any specific disciplinary field. However, no theoretical framework can claim universal applicability. While this study employs quantitative methods to enhance the objectivity and logical consistency of the framework as much as possible, how to promote its implementation across different levels and disciplinary domains remains an issue worthy of further discussion.

In terms of data stakeholders and the data lifecycle dimension, various applying entities can adjust specific elements according to their contexts, with the core principle being adherence to the concepts of collaborative governance by multiple stakeholders and holistic governance across the entire data lifecycle. In the dimension of data governance elements, this study primarily follows the principle of comprehensiveness during the extraction process. Therefore, practitioners at different levels can select corresponding modules for prioritized implementation and emphasis based on their specific needs during application. For example, at the laboratory level, emphasis can be placed on elements within the technical support system, focusing on standardized protocols for data collection to enhance data quality. At the regulatory level, such as government departments, the establishment and refinement of regulatory systems should be the primary concern. At the institutional level, such as scientific data centers, comprehensive configuration should be pursued, using this framework as a reference to address deficiencies.

Scientific data inherently possess interdisciplinary characteristics, and scientific data frameworks face challenges related to adaptability across different disciplinary fields and interdisciplinary sharing. When referring to this framework, different disciplines should make adjustments based on their specific data characteristics. For instance, elements such as privacy protection, informed consent, and public attitudes are primarily applicable to life sciences and medical fields. The key to reducing barriers in cross-disciplinary data sharing, understanding, and usage lies in unifying data standards and specifications, including metadata management protocols. Promoting the establishment of national-level scientific data standards and mandating their application during data collection and processing stages in publicly funded research projects can significantly address these challenges.

Conclusion

The principal contribution of this study lies in establishing an integrated theoretical framework for scientific data governance, aiming to provide systematic referential guidance for governing scientific data. This research initially identifies three core dimensions of scientific data governance through non-systematic literature review: data stakeholders, data lifecycle, and data governance elements. For the classification of stakeholder and lifecycle dimensions, the analysis synthesizes perspectives from existing scholarship while incorporating distinctive characteristics of scientific data. Regarding the data governance elements, bibliometric analysis was

employed to extract critical elements through keyword frequency analysis in thematic literature. However, given the methodological constraints of non-systematic literature selection, incomplete coverage of academic resources, and inherent interpretive bias in categorical delineation, the proposed theoretical framework may contain certain limitations. Practitioners are advised to refine and optimize this scientific data governance framework according to specific application scenarios and requirements.

Declarations

Ethics Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Data Availability

The data that support the findings of this study are available from Web of Science. A full list of consulted articles and their detailed information is provided in Supplementary Table S1.

Code Availability

The software used for bibliometric analysis is Thomson Data Analyzer (TDA) version 3.0.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

QYR collected and analyzed data and materials, constructed the scientific data governance theoretical framework, and prepared the original draft. HZM modified the scientific data governance theoretical framework, revised and reviewed the manuscript, and obtained the funding support for the article. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to acknowledge the Scientific Data Governance and Open Sharing Team for their

contributions to the adjustment and refinement of the theoretical framework of scientific data governance. The contributing members include Xiaofeng Jia, Youbing Ran, Meng Yu, and Borui Zhang. This work was supported by the Noncommunicable Chronic Diseases–National Science and Technology Major Project (Grant No. 2023ZD0509701) and the Medical and Health Technology Innovation Project of the Chinese Academy of Medical Sciences (Grant No. 2021-I2M-1-057).

Figure Legends

Figure 1. Flowchart of Literature Retrieval for the Scientific Data Governance Framework
Figure 2. Theoretical Framework of Scientific Data Governance

References

[1]Pesenson MZ, Pesenson IZ, McCollum B. The Data Big Bang and the Expanding Digital Universe: High-Dimensional, Complex and Massive Data Sets in an Inflationary Epoch[J]. Advances in Astronomy. 2010;2010.

[2]Hey T, Tansley S, Tolle K. The Fourth Paradigm: Data-Intensive Scientific Discovery[M]: Microsoft Press; 2009.

[3]Shah SIH, Peristeras V, Magnisalis I. DaLiF: a data lifecycle framework for data-driven governments[J]. Journal of Big Data. 2021;8(1).

[4]Cox M, Ellsworth D. Managing big data for scientific visualization[J]. 1997.

[5]Laney D. 3D Data Management: Controlling Data Volume, Velocity, and Variety[J]. 2001.

[6]Bello-Orgaz G, Jung JJ, Camacho D. Social big data: Recent achievements and new challenges[J]. Information Fusion. 2016;28:45-59.

[7]Li S, Yueliang Z. Research on the Data Governance Framework of Institutional Research Data Repository Alliance[J]. Library Tribune. 2018;38(08):61-7.

[8]Weill P, Ross J. IT Governance: How Top Performers Manage IT Decision Rights for Superior Results[M]: Harvard Business School Press; 2004.

[9]Brous P, Janssen M, Vilminko-Heikkinen R, editors. Coordinating Decision-Making in Data Management Activities: A Systematic Review of Data Governance Principles2016; Cham: Springer International Publishing.

[10]Soares S. IBM InfoSphere: A Platform for Big Data Governance and Process Data Governance[M]. US: MC Press; 2013.

[11]Weber K, Otto B. A Contingency Approach to Data Governance[M]2007.

[12]Hanisch R, Kaiser D, Yuan A, Medina-Smith A, Carroll B, Campo E. NIST Research Data Framework (RDaF) Version 2.0.[EB/OL]. (2023) [2025-08-31]. https://doi.org/10.6028/NIST.SP.1500-18r2.

[13]Fothergill BT, Knight W, Stahl BC, Ulnicane I. Responsible Data Governance of Neuroscience Big Data[J]. Front Neuroinform. 2019;13:28. Epub 20190424.

[14]Nadal S, Jovanovic P, Bilalli B, Romero O. Operationalizing and automating Data Governance[J]. J Big Data. 2022;9(1):117. Epub 20221210.

[15]Al-Badi A, Tarhini A, Khan AI. Exploring Big Data Governance Frameworks[J]. Procedia Computer Science. 2018;141:271-7.

[16]Zhang QQ, Sun XB, Zhang MC. Data Matters: A Strategic Action Framework for Data Governance[J]. Information & Management. 2022;59(4).

[17]DGI. DGI Data Governance Framework[EB/OL]. [2025-03-04]. https://datagovernance.com/the-dgi-data-governance-framework/#:~:text=The%20DGI%20Data%20Governance%20Framework%20is%20a%20logical,decisions%20about%20and%20taking%20action%20on%20enterprise%20data.

[18]Mark Mosley MB, Susan Earley, Deborah Henderson. The DAMA Guide to the Data Management Body of Knowledge - DAMA-DMBOK[M]: Technics Publications, LLC DAMA International; 2009.

[19]Jang KA, Kim WJ. Development of data governance components using DEMATEL and content analysis[J]. Journal of Supercomputing. 2021;77(4):3695-709.

[20]Eke DO, Bernard A, Bjaalie JG, Chavarriaga R, Hanakawa T, Hannan AJ, et al. International data governance for neuroscience[J]. Neuron. 2022;110(4):600-12. Epub 20211215.

[21]IBM. What is data governance?[EB/OL]. [2025-03-04]. https://www.ibm.com/think/topics/data-governance.
[22]Association CCS. Data Governance Standardization White Paper[EB/OL]. (2021) [2025-03-05]. https://13115299.s21i.faiusr.com/61/1/ABUIABA9GAAg7YGHjgYo4PjDtgQ.pdf.
[23]O'Doherty KC, Shabani M, Dove ES, Bentzen HB, Borry P, Burgess MM, et al. Toward better governance of human genomic data[J]. Nature Genetics. 2021;53(1):2-8.
[24]Abraham R, Schneider J, vom Brocke J. Data governance: A conceptual framework, structured review, and research agenda[J]. International Journal of Information Management. 2019;49:424-38.
[25]Khatri V, Brown CV. Designing Data Governance[J]. Communications of the Acm. 2010;53(1):148-52.
[26]Goel K, Martin N, ter Hofstede A. Demystifying data governance for process mining: Insights from a Delphi study[J]. Information & Management. 2024;61(5).
[27]Jiang ST, Liu YG, Zheng H, Zhang L, Zhao HT, Sang XT, et al. Evolutionary patterns and research frontiers in neoadjuvant immunotherapy: a bibliometric analysis[J]. International Journal of Surgery. 2023;109(9):2774-83.
[28]Zhang LL, Ling J, Lin MW. Carbon neutrality: a comprehensive bibliometric analysis[J]. Environmental Science and Pollution Research. 2023;30(16):45498-514.
[29]Yan P. Analysis of the Participation of Archival Departments in Scientific Data Management from the Perspective of Stakeholders[J]. Archives World. 2019(03):47-9.
[30]Sheng XP, Wu H. Analysis of the Motivations of Different Stakeholders in the Activity of Open Sharing of Scientific Data[J]. library and Information Services. 2019;63(17):40-50.
[31]Smith K, Seligman L, Swarup V. Everybody share: The challenge of data-sharing systems[J]. Computer. 2008;41(9):54-+.
[32]wang D. Research on Scientific Data Governance Model from the Perspective of Open Data Maturity: Heilongjiang University; 2022.
[33]Gao Y, Zhu ZL, Yang J. An Evolutionary Game Analysis of Stakeholders' Decision-Making Behavior in Medical Data Sharing[J]. Mathematics. 2023;11(13).
[34]OECD. OECD principles and guidelines for access to research datafrom public funding[EB/OL]. (2007) [2025-02-27]. https://www.oecd.org/en/publications/oecd-principles-and-guidelines-for-access-to-research-data-from-public-funding_9789264034020-en-fr.html.
[35]COMMISSION E. Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 Version 1.0 [EB/OL]. (2013) [2025-02-27]. https://ai.tecnico.ulisboa.pt/files/sites/52/guidelines-scientific-publicationsresearch-data-in-h2020.pdf.
[36]LERU. LERU roadmap for research data[EB/OL]. (2013) [2025-02-27]. https://www.leru.org/files/LERU-Roadmap-for-Research-Data-Full-paper.pdf.
[37]International S. Open data in a big data world[EB/OL]. (2015) [2025-02-27]. https://council.science/wp-content/uploads/2017/04/open-data-in-big-data-world_short.pdf.
[38]Blazquez D, Domenech J. Big Data sources and methods for social and economic analyses[J]. Technological Forecasting and Social Change. 2018;130:99-113.
[39]Wen FF. Research on the Construction of the Policy System for Government Data Opening in China 2019.
[40]Pei L, Wang JC. Review of Research Progress in Information Lifecycle Management[J]. Journal of Information. 2010;29(09):7-10+20.

[41]Simonet A, Fedak G, Ripeanu M. Active Data: A programming model to manage data life cycle across heterogeneous systems and infrastructures[J]. Future Generation Computer Systems-the International Journal of Escience. 2015;53:25-42.

[42]Chen SX. Research on the Construction of FAIR Evaluation Index System for Scientific Data Platform 2024.

[43]IBM. What is data lifecycle management (DLM)?[EB/OL]. https://www.ibm.com/think/topics/data-lifecycle-management.

[44]Rahul K, Banyal RK. Data Life Cycle Management in Big Data Analytics[J]. Procedia Computer Science. 2020;173:364-71.

[45]Shameli-Sendi A. An efficient security data-driven approach for implementing risk assessment[J]. Journal of Information Security and Applications. 2020;54.

[46]Jetten M, Simons E, Rijnders J. The role of CRIS's in the research life cycle. A case study on implementing a FAIR RDM policy at Radboud University, the Netherlands[J]. Procedia Computer Science. 2019;146:156-65.

[47]Ball A. Review of Data Management Lifecycle Models[J]. 2012.

[48]Cao TZ. Synergetics Approach to the Study of Collaborative Mechanisms in Public Administration Implementation: Theoretical Rationality and Interdisciplinary Foundations[J]. Journal of Zhejiang Provincial Party School. 2009;25(01):37-42.

Data source: Web of Science Core Collection

Publication year: 2000-2024 | **Index sources: SCI-EXPANDED, SSCI**

TS=(“scientific data” OR “research data”) AND TS=(“governance” OR “sharing”) AND TS=(“framework”) N=7624

Document types=Articles: N=6555

Languages=English: N=6525

A total of 6525 articles included

Elements

Collection

Storage

Processing

Management

Sharing

Application

Deletion

Organizational Operation System

Organizational Structure	Talent Team Building
Operation Mechanism	Funding Sources

Technical Support System

Technological Infrastructure	Data Ontology Construction
Data Quality Control	Metadata Management
Data Standards and Specifications	

Risk Prevention and Control System

Public Attitudes	Informed Consent
Access Control	Privacy Protection
Data Security	Data Fairness

Value Realization System

Data Resources	Data Circulation
Data Ownership	Data Services

Regulatory System

Ethics Framework	Policies and Regulations

Stakeholders

Data providers

Data users

Data sharing Facilitators

Policymakers

Lifecycle