ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202307.00309

A Two-Dimensional Perspective on Scientific Data Management Systems: Postprint Interpretation of the "Measures for the Management of Scientific Data"

Authors: Xing Wenming, HONG Fanglin, Li Xiaoyan

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] To interpret the "Management Measures for Scientific Data", aiming to provide references for better understanding and implementation of these Measures and for promoting scientific data management and sharing practices in China. [Method/Process] The interpretation is conducted from a two-dimensional perspective encompassing the data lifecycle and stakeholders. [Results/Conclusion] The "Management Measures for Scientific Data" establishes a scientific data management system with the data lifecycle as the warp and stakeholders as the weft. These two elements are integrated and mutually complementary, jointly forming China's scientific data management system characterized by "national coordination with division of responsibilities among departments and regions."

Full Text

Preamble

Volume 63, Issue 23, December 2019

A Two-Dimensional Perspective on Scientific Data Management Systems: Interpreting the *Scientific Data Management Measures*

Xing Wenming[1], Hong Fanglin[1], Li Xiaoyan[2],[3]
[1]School of Public Administration, Xiangtan University, Xiangtan 411105
[2]National Science Library, Chinese Academy of Sciences, Beijing 100190
[3]Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190

Abstract

[**Purpose/Significance**] This paper interprets the *Scientific Data Management Measures* to provide a reference for better understanding and implementing these measures, thereby promoting scientific data management and sharing practices in China. [**Method/Process**] The interpretation is conducted from a two-dimensional perspective encompassing the data lifecycle and stakeholder responsibilities. [**Result/Conclusion**] The *Measures* establish a scientific data management system that uses the data lifecycle as the longitudinal framework and stakeholder responsibilities as the latitudinal framework. These two dimensions integrate and complement each other, collectively forming China's management system characterized by "national coordination with division of responsibilities among departments and regions."

Keywords: *Scientific Data Management Measures*; data lifecycle; stakeholders
Classification Number: G250
DOI: 10.13266/j.issn.0252-3116.2019.23.004

Scientific data has become a crucial strategic resource, scientific asset, and public information resource in today's era of digital research. It represents a key component for achieving national scientific and technological innovation, economic and social development, and national security. The formulation and development of scientific data management policies have attracted widespread attention and discussion in academic circles, with consensus emerging that managing scientific data and related elements throughout the entire data lifecycle is a highly policy-driven, specialized, and collaborative endeavor requiring national-level policies and regulations to ensure data sharing, management, security, and confidentiality [1-2]. On March 17, 2018, the General Office of the State Council issued the *Scientific Data Management Measures* (hereinafter referred to as the *Measures*), which provides national-level guidance and a legal basis for the collection, submission, preservation, sharing, utilization, confidentiality, and security of scientific data.

Following the promulgation of the *Measures*, various sectors have offered interpretations from different perspectives, focusing on three main aspects: (1) introducing and publicizing the main content, principles, and basic requirements of the *Measures* [3-5]; (2) interpreting its core ideas, such as the characteristics highlighted by Ye Yujiang, Director of the Basic Research Department of the Ministry of Science and Technology, at the press conference: first, drawing on advanced domestic and international experiences to strengthen full lifecycle data management; second, ensuring data security as the top priority; third, vigorously promoting open sharing of scientific data resources under the principle of "openness as the norm, non-openness as the exception"; and fourth, addressing weak links in China's scientific data work by proposing targeted measures regarding principal responsibilities, intellectual property rights, and submission mechanisms [6]; and (3) specific scholarly analyses, such as examining the principles, approaches, and safeguards for scientific data sharing and utilization [8], as

well as highlighting innovations in submission, utilization, and security backup [9].

While these interpretations provide valuable insights, most focus on specific provisions or particular domains, leaving room for a more systematic and holistic understanding. As China's first national-level policy on scientific data management and sharing, the *Measures* are guided by two main threads: first, clarifying key tasks for scientific data management and sharing from a data lifecycle perspective; second, defining the responsibilities and division of labor among various stakeholders from a management perspective, enabling relevant parties to fulfill their duties and collaborate effectively. This paper interprets the *Measures* from these two dimensions to offer a new perspective for better understanding and implementing the policy and for advancing China's scientific data sharing and management practices.

2. Interpreting the *Measures* from a Data Lifecycle Perspective

The lifecycle concept, originally a biological term describing the birth, growth, maturity, and death of organisms, has been extended and developed into an important research methodology applicable to various phenomena [10]. Researchers have applied this theory to scientific data studies, such as Wei Yue and Liu Guifeng's framework for research data management policies covering acquisition, organization, preservation, sharing, and security [11], and Huang Ruhua and Lai Tong's government data openness lifecycle model encompassing creation, collection, organization, processing, storage, publication, discovery, acquisition, and evaluation [12]. The vitality and broad applicability of lifecycle theory are evident in the *Measures*, which are organized around the scientific data lifecycle: collection and production, submission and preservation, and sharing and utilization.

2.1 Scientific Data Collection and Production

Data collection and production constitute the initial stage of data management, where standardized and orderly collection is key to ensuring data quality. Regarding collection and production, the *Measures* emphasize both the importance of standards and the need to establish a "scientific data quality control system." Article 9(2) states that entities must "conduct scientific data collection, production, processing, and long-term preservation in accordance with relevant standards and specifications to ensure data quality." Article 11 further specifies that "legal entities and scientific data producers shall organize data collection, production, and processing according to relevant standards" and emphasizes that "legal entities shall establish a scientific data quality control system to ensure data accuracy and usability."

This demonstrates that ensuring data quality through standardized specifications is a clear policy direction. China has long prioritized standardization in

scientific data, developing numerous collection standards. A search of the National Standard Information Public Service Platform reveals relevant standards as shown in .

2.2 Scientific Data Submission and Preservation

Data submission and preservation are prerequisites for sharing [13]. To realize the value of scientific data, China has actively promoted submission and preservation work. Since 2004, the Ministry of Science and Technology and the Ministry of Finance have established national science and technology resource sharing service platforms in eight fields: basic science, agriculture, forestry, oceanography, meteorology, seismology, earth system science, and population and health [14]. These platforms have accumulated experience in data submission and management, laying a foundation for further improving submission systems and informing the *Measures*.

Articles 12-15 of the *Measures* regulate scientific data submission based on classification management principles, summarized in . The *Measures* establish different requirements for data from various sources and contexts. For data from government-funded projects, Article 13 stipulates that "scientific data from national science and technology programs shall be submitted by the project lead unit to relevant scientific data centers, with a submission certificate obtained," effectively establishing a "submit before acceptance" mechanism.

Following submission, ensuring data integrity and security is essential for sharing. The *Measures* address this from multiple angles: (1) establishing preservation systems and mechanisms, with Article 16 requiring legal entities to "establish scientific data preservation systems, equip necessary facilities for storage, management, services, and security, and ensure data integrity and safety"; (2) improving cybersecurity systems, with Article 28 mandating robust cybersecurity and protection systems; and (3) standardizing emergency management and disaster backup mechanisms, with Article 29 requiring scientific data centers to "establish emergency management and disaster backup mechanisms, with offsite backup for important data." International practices, such as Princeton University's requirement for triple backups using at least two different media (e.g., hard drives, DVDs, flash drives, or cloud storage), provide valuable references [15].

2.3 Scientific Data Sharing and Utilization

Sharing and utilization represent both the goal of data submission and a critical lifecycle stage. The *Measures* clarify sharing approaches through classification guidance:

(1) **Compiling and publishing scientific data resource catalogs.** Article 19 states that "scientific data from government budget funding shall, following the principle of openness as the norm, be organized by competent authorities into resource catalogs and connected to the national data

sharing exchange platform for public and departmental access." Article 20 requires legal entities to "publish scientific data openness catalogs and provide social sharing through online download, offline sharing, or customized services." Resource catalogs serve as gateways for users to discover and access data.

(2) **Promoting scientific data publication and dissemination.** Article 22 specifies that "competent authorities and legal entities shall actively promote scientific data publication and dissemination, supporting researchers in publishing clear-property, accurate, complete, and high-value data." Data publication signifies reviewed and organized data of high credibility while respecting producers' contributions, thereby encouraging openness [5].

(3) **Encouraging value-added services.** Article 21 states that "legal entities shall analyze and mine scientific data according to demand, forming valuable data products and providing value-added services. Social organizations and enterprises are encouraged to conduct market-oriented value-added services." This establishes a market-oriented development approach that attracts social participation, meets personalized needs, and increases returns for data producers.

3. Interpreting the *Measures* from a Stakeholder Perspective

The concept of "stakeholders" originated in 1960s corporate governance theory, positing that organizational development depends on inputs from various stakeholders (shareholders, creditors, employees, consumers, suppliers) and should pursue overall stakeholder interests rather than those of specific groups [16]. This theory has been widely applied across disciplines, including library and information science. For instance, Huang Ruhua et al. positioned libraries as core actors in scientific data management, analyzing their relationships with stakeholder institutions to propose practical strategies [17]. Guo Shilin examined interest demands and conflicts in scientific data sharing, constructing an interest-balancing mechanism from a policy perspective [18].

The *Measures* notably apply stakeholder theory by clarifying responsibilities for different actors—national science and technology administrative departments, competent authorities, legal entities, and scientific data centers—facilitating policy implementation. The specific responsibilities are outlined in [Figure 1: see original paper].

3.1 State Council Science and Technology Administrative Department

The State Council Science and Technology Administrative Department (primarily the Ministry of Science and Technology) is responsible for implementing national innovation policies, planning and promoting scientific and technological

development, and playing a crucial macro-management and coordination role in scientific data management. Article 7 specifies its responsibilities: (1) organizing research and formulation of national scientific data management policies and standards; (2) coordinating standardized management, open sharing, and evaluation; (3) promoting construction and development of national scientific data centers; and (4) overseeing the national scientific data network management platform. Article 18 further requires strengthening overall planning to optimize and integrate national scientific data centers based on existing high-quality centers. As Deputy Director Wang Ruidan of the National Science and Technology Infrastructure Platform Center noted, "China urgently needs to build influential scientific data centers commensurate with its status as a scientific powerhouse" [6].

3.2 Competent Authorities

Competent authorities are defined in Article 8 as "relevant departments of the State Council and provincial people's governments," including the General Office of the State Council, Ministry of Science and Technology, China Meteorological Administration, Ministry of Water Resources, Ministry of Natural Resources, National Health Commission, National Bureau of Statistics, Ministry of Agriculture and Rural Affairs, and corresponding provincial departments. Their responsibilities include: (1) establishing departmental/regional scientific data management policies and publicizing national policies; (2) guiding affiliated legal entities in strengthening data management; (3) handling data classification according to regulations; (4) planning and constructing departmental/regional data centers and promoting sharing; and (5) establishing incentive mechanisms and conducting performance evaluations.

Additional provisions assign specific duties: Articles 12 and 14 address submission systems; Articles 19 and 22 cover catalog compilation and publication promotion; Articles 26 and 27 regulate confidentiality and security management.

3.3 Legal Entities

Legal entities primarily include research institutes, universities, and enterprises responsible for data collection, submission, preservation, management, and services. Article 9 defines their responsibilities: (1) implementing national and departmental policies and establishing internal management systems; (2) conducting collection, production, and preservation according to standards to ensure quality; (3) ensuring confidentiality and security; (4) establishing data management systems, publishing openness catalogs, and providing sharing services; and (5) guaranteeing necessary infrastructure, funding, and personnel.

Further provisions require establishing quality control systems (Article 11), preservation systems with necessary facilities (Article 16), talent development with incentives (Article 17), and value-added services (Article 21). For example, Peking University's Management Science Data Center established the China

Survey Data Archive (CSDA) in 2015, integrating high-quality survey data resources and providing browsing, search, download, and online analysis services, promoting openness in humanities and social sciences.

3.4 Scientific Data Centers

Scientific data centers are crucial institutions for promoting open sharing. Developed countries have established national-level centers that aggregate domestic and global scientific data resources. China is at a critical stage of implementing innovation-driven development and building a scientific powerhouse, making data center construction essential. Article 10 defines their responsibilities: (1) integrating and receiving submitted data; (2) classifying, processing, and mining data; (3) ensuring security and promoting lawful open sharing; and (4) strengthening domestic and international exchanges.

Articles 13 and 15 mandate receiving submitted data; Articles 28 and 29 specify security responsibilities. Some Chinese centers have already explored institutional systems, such as the National Earthquake Data Sharing Center and its sub-centers, which have developed operational regulations based on the *Earthquake Scientific Data Sharing Management Measures*, providing valuable experience.

3.5 Data Producers

Producers are responsible for collecting data and ensuring its quality. The *Measures* define their responsibilities in three areas: (1) complying with standards to ensure quality (Article 11); (2) submitting and depositing data as required to ensure security and sharing (Articles 13 and 14); and (3) actively engaging in data publication and dissemination (Article 22). The National Standardization Administration has issued the *Scientific Data Citation* standard (GB/T 35294-2017) to guide proper citation.

3.6 Data Users

Users serve as both beneficiaries and disseminators. They should exercise responsibility and cite data properly. Article 23 requires users to "comply with intellectual property regulations and acknowledge data sources in papers, patents, and publications."

4. The Two-Dimensional Framework of the *Measures*

The *Measures* establish a management system combining the data lifecycle and stakeholder responsibilities. The framework clarifies stakeholder duties in Chapter 2 ("Responsibilities") while organizing subsequent chapters according to lifecycle stages—collection, submission, preservation, and sharing—further specifying responsible parties. These dimensions integrate and complement each other,

forming China's "national coordination with division of responsibilities among departments and regions" system, as illustrated in .

The table shows that each stakeholder plays three roles: (1) implementing superior policies, (2) establishing departmental regulations, and (3) coordinating and promoting work among subordinate units. This includes guiding, coordinating, incentivizing, and evaluating data collection, processing, submission, preservation, and sharing.

5. Reflections and Recommendations

5.1 Clarify Responsibilities of Research Funding Agencies

Scientific data from national science and technology programs are important research outcomes and foundations for innovation. Since 2008, the Ministry of Science and Technology has required data submission for the National Basic Research Program (973 Program) [19] and, in 2014, mandated submission before project acceptance for basic science and technology work projects [21]. These practices have provided valuable experience.

However, other major programs (e.g., National Natural Science Foundation, National Science and Technology Major Projects, Key R&D Programs) have not yet established submission systems, hindering data integration and sharing. Most developed countries mandate data deposition [22]. Although the *Measures* do not explicitly address funding agencies, they are critical stakeholders. These agencies should promptly formulate data submission regulations to promote sharing.

5.2 Strengthen Domain-Specific Scientific Data Centers

Different disciplines have distinct data characteristics and sharing requirements. Domain-specific data centers should be prioritized to provide specialized services, with national centers formed through optimization and integration. Developed countries have established domain centers like the U.S. National Center for Biotechnology Information's GenBank, NASA's Space Science Data Coordinated Archive, and the UK's Cambridge Crystallographic Data Centre. China has also explored domain-specific centers, such as the 973 Program Resource and Environment Data Submission Management Center established in 2008 [23]. Future efforts should focus on building domain centers based on high-quality units to support data submission and sharing.

5.3 Improve Scientific Data Publication Mechanisms

Data publication is an important sharing method. Article 22 acknowledges its significance, yet China's practice lags behind developed countries. While data journals like *China Scientific Data* and *Geoscience Big Data* have emerged, and the *Scientific Data Citation* standard (GB/T 35294-2017) was issued in

2017, sustainable publication models and mechanisms need strengthening. Relevant departments should explore robust publication mechanisms, particularly sustainable support measures, to promote openness.

5.4 Improve Incentive Mechanisms for Data Producers

Researchers are primary data producers and key to ensuring data quality. Studies show that researchers worry about increased management burdens and misuse of data, and many are unwilling to share data unconditionally [24]. The *Measures* focuses on institutional frameworks rather than producer obligations, mentioning them only in Articles 11 and 13. Improving incentive mechanisms—such as funding for data management, and material or recognition rewards for high-quality data submissions—would encourage sharing and align with Article 17's principles.

References

[1] Huang Ruhua, Wang Bin, Zhou Zhifeng. Strategies for promoting scientific data sharing in China[J]. Library and Information Service, 2014(3): 7-13.
[2] Zhang Lili, Wen Liangming, Shi Lei, et al. Latest developments in scientific data management and open sharing at home and abroad[J]. Bulletin of Chinese Academy of Sciences, 2018, 33(8): 774-782.
[3] General Office of the State Council issues the *Scientific Data Management Measures*[J]. China Work Safety Science and Technology, 2018, 14(4): 74.
[4] Zhang Baogang. General Office of the State Council issues the *Scientific Data Management Measures*[J]. Beijing Surveying and Mapping, 2018, 32(5): 577.
[5] Wu Yan. *Scientific Data Management Measures* released[J]. Fujian Light & Textile, 2018(5): 2.
[6] State Council Information Office of China. Press conference on the *Scientific Data Management Measures*[EB/OL]. [2019-01-27]. http://www.scio.gov.cn/xwfbh/gbwxwfbh/xwfbh/kjb/Doc
[7] Yuan Yufei. Making open sharing of scientific data the norm[N]. Guangming Daily, 2018-04-05(2).
[8] 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(1): 117-124.
[9] Policy Research Institute, CCID. Highlights of the *Scientific Data Management Measures*[J]. China Industry and Information Technology, 2018(9): 12-14.
[10] Ma Feicheng, Wang Juncheng, Zhang Yutao. Mapping the knowledge domain of lifecycle theory research in China[J]. Information Science, 2010, 28(4): 481-487, 506.
[11] Wei Yue, Liu Guifeng. Analysis of foreign university scientific data management and sharing policies based on data lifecycle theory[J]. Journal of Intelligence, 2017, 36(5): 153-158.
[12] Huang Ruhua, Lai Tong. Obstacles to China's government data openness from a data lifecycle perspective[J]. Information Studies: Theory & Application,

2018, 41(2): 7-13.
[13] Cao Yanrong, Bi Jiantao, Chi Tianhe, et al. Data submission issues in scientific data sharing[C]//Proceedings of the National Conference on Cartography and GIS. 2004.
[14] ScienceNet. Expert interpretation of the background to the *Scientific Data Management Measures*[EB/OL]. [2019-01-27]. http://news.sciencenet.cn/htmlnews/2018/4/40815.shtm.
[15] Wei Yue, Liu Guifeng. Analysis of foreign university scientific data management and sharing policies based on data lifecycle theory[J]. Journal of Intelligence, 2017, 36(5): 153-158.
[16] Zhang Xu. Mapping international think tank evaluations: An evolutionary review based on stakeholder theory[J]. Information Studies: Theory & Application, 2019, 42(3): 49-54.
[17] Huang Ruhua, Lai Tong. Analysis of library participation in scientific data management from a stakeholder perspective[J]. Library and Information Service, 2016, 60(3): 21-25, 89.
[18] Guo Shilin. Research on interest balancing mechanisms for scientific data open sharing from a policy perspective[D]. Harbin: Heilongjiang University, 2018.
[19] Interim Measures for Data Submission of Resource and Environment Projects under the National Basic Research Program[EB/OL]. [2019-01-27]. http://www.973geodata.cn/files/1.pdf.
[20] Ministry of Science and Technology of China. Notice on strengthening management of project acceptance for the 11th Five-Year Science and Technology Program[EB/OL]. [2019-02-03]. http://www.most.gov.cn/tztg/201006/t20100629_{78173}.htm.
[21] Department of Basic Research, Ministry of Science and Technology. Notice on acceptance of basic science and technology work projects[EB/OL]. [2019-02-03]. http://www.most.gov.cn/tztg/201405/t20140521_{113311}.htm.
[22] Zhang Yao, Gu Liping, Yang Yunxiu, et al. Investigation and analysis of foreign research funding agencies' data policies: Taking British and American research councils as examples[J]. Library and Information Service, 2015(6): 53-60.
[23] 973 Program Resource and Environment Data Submission Service Platform[EB/OL]. [2019-02-03]. http://www.973geodata.cn/webinfo.
[24] Xing Wenming, Yang Ling. Investigation of scientific data management status in China's research institutions[J]. Digital Library Forum, 2018(12): 27-33.

Author Contributions:

Xing Wenming: Research design and framework, content revision and finalization
Hong Fanglin: Initial draft writing and revision
Li Xiaoyan: Data supplementation

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

Source: ChinaXiv — Machine translation. Verify with original.