Scientific Data Control Strategy and Its Model Optimization

Authors: Wen Yuheng, Xu Kunshan, Wen Yuheng

Date: 2023-11-06T00:00:00+00:00

Abstract

[Purpose/Significance] The proposal and optimization of control strategies and models for scientific data advances the foundational institutional development of scientific data platforms and balances the security protection and shared utilization of scientific data. [Method/Process] Selecting user agreements from national-level scientific data platforms as samples and employing grounded theory methodology to obtain 76 initial concepts, 9 sub-categories, and 5 main categories, thereby refining a control strategy model for scientific data. [Results/Conclusion] China's scientific data platforms have formed a "five-controls-in-one" control strategy model with purpose, behavior, rights, identity, and responsibility as its constituent elements. It is recommended to optimize and form a "three-stage" control strategy model from the aspects of pre-event defense, in-process management, and post-event accountability. Specific measures include: Summarizing and enumerating purpose types to strengthen purpose control; Clarifying data rights ownership to strengthen rights control; Unifying hierarchical and categorical management to supplement identity control; Balancing platform management and user control, transitioning to platform autonomy, promoting unified platform awareness and norms, and introducing industry external regulation; Rationally allocating responsibilities among platforms, users, and scientific data providers.

Full Text

The Control Strategy of Scientific Data and Its Model Optimization

Wen Yuheng[1,2], **Xu Kunshan**[1]

[1]School of Law, Hunan University of Science and Technology, Xiangtan 411105
[2]Hunan Data Governance and Smart Judicial Research Center, Xiangtan 411105

Abstract:
[Purpose/Significance] Proposing and optimizing control strategies and models for scientific data helps advance the foundational institutional development of scientific data platforms while balancing data security protection with sharing and utilization. [Method/Process] Using user agreements from national-level scientific data platforms as samples, this study employs grounded theory methodology to extract 76 initial concepts, 9 subcategories, and 5 main categories, thereby refining a control strategy model for scientific data. [Result/Conclusion] China's scientific data platforms have formed a "five-control-in-one" strategy model comprising purpose, behavior, rights, identity, and responsibility as core elements. We recommend optimizing this into a "three-stage" control strategy model covering pre-defense, in-process management, and post-event accountability. Specific measures include: summarizing and enumerating purpose types to strengthen purpose control; clarifying data rights ownership to enhance rights control; unifying hierarchical classification management to reinforce identity control; balancing platform management with user control, transforming platform autonomy, promoting unified platform awareness and norms, and introducing new industry heteronomy; rationally allocating responsibilities among platforms, users, and scientific data providers.

Keywords: scientific data; control strategy; grounded theory; behavior control; rights control

---

In 2022, the "Opinions of the Central Committee of the Communist Party of China and the State Council on Building Basic Systems for Data to Give Full Play to the Role of Data Resources" (hereinafter referred to as the "Data Twenty Articles") stated that building basic data systems concerns the overall national development and security [1]. In 2023, the establishment of the National Data Administration marked a new phase in China's basic data system construction. Scientific data represents one of the nation's strategic resources, and the foundational institutional development of scientific data platforms (or centers) should receive commensurate attention. As early as 2018, the "Scientific Data Management Measures" (hereinafter referred to as the "Measures") defined scientific data and stipulated that openness should be the norm while non-openness the exception [2], establishing openness and sharing as the value orientation for China's scientific data platform construction. Against this backdrop, scientific data platforms must not only manage existing data rationally and fortify security barriers but also conduct sharing in a safe and orderly manner, making a reasonable and effective scientific data control strategy essential. Only by establishing a practical and systematic control system can we build high-quality scientific data platforms that meet the development needs of domestic advantageous industries and support national scientific innovation, economic and social development, and national security.

As of March 2023, a preliminary review revealed abundant research achievements on scientific data both domestically and internationally. Domestic research, for instance, covers metadata, stakeholders, quality, literacy, management, platform (center) construction, open sharing, security assurance, services, citation, correlation, integration, utilization, publishing, regulation, privacy protection, and coordination with personal data protection—each with further subdivisions. To streamline length and focus the topic, after confirming that existing literature does not address "scientific data control strategies," and considering that our samples derive from scientific data platforms (centers), we narrowed our literature review to focus specifically on scientific data platforms (centers).

On CNKI Academic Journals (CSSCI), we conducted precise title searches using "scientific data platform," "scientific data center," "scientific data sharing platform," and "scientific data sharing center." Research on scientific data (sharing) platforms/centers in China remains relatively limited, primarily covering: **Practical conditions**: For example, the "Internet+" environment facilitates deep integration and universal participation in scientific data sharing, providing favorable conditions for improving platform construction and ensuring efficient resource utilization [3]. Existing domestic platforms have explored organizational management, resource construction, and user services to form relatively complete infrastructure [4], yet issues persist such as monolithic service approaches, emphasis on data dominance over data services [5], insufficient normative explanations, shallow content integration, and indirect resource acquisition [6]. **Specialized platforms**: Research covers the National Agricultural Science Data Sharing Center [7], China West Environment and Ecology Science Data Center [8], Open Medical Science Data Platform [9], national science data centers for Yellow River basin thematic data [10], Yangtze River Delta cross-regional scientific data centers [11], field scientific observation and research stations (networks) and scientific data centers [12], Fudan University Social Science Data Platform [13], and Hunan Science and Technology Information and Scientific Data Sharing Platform [14], among others across various domains. **Specific construction aspects**: Research mainly addresses performance evaluation [15] and service effectiveness analysis [16] of scientific data sharing platforms, citation norm investigations [17], citation status analysis [18], data management studies [19], and compliance of user registration agreements [20], with some scholars focusing on applied technologies like open-source technology [21] and knowledge management [22]. **Insights from foreign databases**: Some scholars surveyed German geoscience field scientific data centers to propose development concept maps for China's relevant fields [23], while others studied European and American biomedical science data centers, offering insights from six aspects: basic support, infrastructure, resource construction, technology development, standards and norms, and user services [24].

On Semantic Scholar, we searched English literature since 2014 using "scientific data platform," "scientific data center," "scientific sharing data platform," and "scientific sharing data center" in titles, supplemented by Google Scholar and Web of Science. Over the past decade, foreign research on scientific data (sharing) platforms (centers) has primarily involved: **Practical conditions**: Such as current practices of open science data centers [25] and data management principles of scientific data centers [26]. **Specialized platforms**: Including the open electronic science data sharing platform B2SHARE [27] and the Virtual Molecular Science Data Center (VHERLA) [28]. **Specific construction aspects**: Such as application systems for scientific data [29], cloudification considerations [30], and catalog management [31].

In summary, both domestic and international research emphasizes scientific data platform construction, forming certain achievements in practical conditions, specialized platform building, and specific construction content, with domestic research additionally focusing on foreign database insights. However, insufficient attention has been paid to scientific data control strategies. Therefore, this study selects national-level scientific data centers as research objects, analyzes control strategies in platform practice, refines models, and proposes optimization recommendations, aiming to stimulate academic attention to the "control strategy" theme and advance theoretical development in scientific data control strategies.

2.1 Sample Selection

Based on the National Science and Technology Infrastructure Platform Center, we obtained rosters of 20 national scientific data centers and 31 national resource repositories (hereinafter "national-level scientific data platforms"). We searched each platform's official website for user registration agreements, scientific data sharing policies, or user instructions—any documents potentially involving scientific data management (collectively "user agreements"). However, most platforms lacked such documents, allowing users to directly access data after registration, with nine platforms enabling data downloads without registration. We ultimately obtained user agreements from 11 national scientific data centers and three national resource repositories. Through manual review of full texts, we extracted 51 original text segments involving scientific data (see Table 1). Notably, since most platforms generically combined service terms, disclaimers, and sharing policies into single files with inconsistent formats across platforms, the original segments appeared in different locations—some in privacy policies, others in disclaimers.

2.2 Method Selection

In qualitative research, the "grounded theory" proposed by B. Glaser and A. Strauss is widely applied. Chinese scholars argue that grounded theory's primary purpose is theory-building from empirical data—researchers generally begin without theoretical hypotheses, proceeding directly from observation to induce empirical generalizations from raw materials, then elevating these to theory [32]. Unlike applications that stop at model-building [33][34] or factor induction [35] after coding, this study has dual objectives: objectively presenting the actual status of scientific data control strategies (the "actual model"); and proposing optimized solutions (the "ideal model"). To achieve these, we primarily employ grounded theory supplemented by word frequency statistics.

Traditional approaches to presenting actual status rely on assumptions, experience, and induction, essentially moving directly from materials to results, often introducing significant error due to researchers' perspectives, understanding, and value orientations. Grounded theory begins without any presuppositions, remaining "faithful" to raw materials from initial concepts, then progressively developing categories, subcategories, and main categories, conducting saturation testing after coding to maximize objectivity and comprehensiveness. Additionally, given that scientific data content is distributed across different platforms' user agreements involving registration, sharing, or usage instructions, grounded theory suits large-scale, fragmented samples for both quantitative and qualitative analysis [36]. Grounded theory has evolved into three paradigms: classic grounded theory (open coding, selective coding, theoretical coding), procedural grounded theory (open coding, axial coding, selective coding), and constructivist grounded theory (initial coding, focused coding, axial coding, theoretical coding). Procedural grounded theory remains mainstream in China's library and information science research [37]. Since inducting scientific data control strategy content requires open, axial, and selective coding, we specifically selected procedural grounded theory.

While grounded theory differs from quantitative empirical research—researchers enter fieldwork without theoretical hypotheses, conducting empirical generalizations directly from data to ultimately develop theory [38]—and word frequency statistics would not be necessary if only refining a theoretical model, we conduct such statistics to reveal the weight of each component in the "actual model," thereby quantifying the degree of attention each element receives in practice. As this analysis uses the full sample, which reflects the overall status of scientific data control strategies, the frequency and proportion of concepts, categories, and main categories can objectively reflect their importance in the overall picture. Since samples come from different national-level scientific data platforms, each with both commonalities and unique characteristics in scientific data provisions, statistical frequency and proportion calculations hold significance. With a fixed total sample size, proportions indicate certain issues—small proportions may not necessarily indicate uniqueness, but large proportions strongly suggest common practices. For example, behavior control appears 71 times (69.61% proportion), indicating it is a universal practice in scientific data control strategies; purpose control appears 7 times (6.86% proportion), indicating it is rarely employed.

Thus, word frequency statistics help clarify optimization approaches for internal elements and their combinations in scientific data control strategies. If an element has high proportion in the "actual model" but experts deem it unimportant, we recommend reduction; if experts deem it important, we recommend maintenance or enhancement. If an element has low proportion but experts deem it important, we recommend appropriate enhancement; if experts deem it unimportant, we recommend maintenance or reduction. Overall, this study attempts to introduce word frequency as a supplementary quantitative method on top of grounded theory's qualitative analysis to maximize conclusion reliability and recommendation reasonableness.

2.3.1 Open Coding

Open coding extracts key statements from original text segments to form initial concepts, ensuring concepts remain "faithful" to raw materials. During coding, one segment often contains multiple key statements, resulting in a one-to-many relationship between segments and initial concepts, rarely one-to-one. Due to space constraints and because Sample 3S covers both one-to-one and one-to-many scenarios with moderate length, we use it as an open coding example (see Table 2). To present the thinking process, Table 2 includes the key statement extraction step.

2.3.2 Axial Coding

Axial coding extracts subcategories from initial concepts and merges similar subcategories into main categories. One initial concept may yield one subcategory, but when aggregating, the relationship is many-to-one. Similarly, subcategories and main categories have a many-to-one relationship. As shown in Table 3 , we extracted 9 subcategories from 76 initial concepts: usage behavior requirements, acquisition behavior requirements, management behavior requirements, usage purpose restrictions, risk responsibility assumption, acquisition identity restrictions, intellectual property ownership, data rights ownership, and user rights restrictions. From these, we obtained 5 main categories: behavior control, purpose control, responsibility control, identity control, and rights control.

2.3.3 Selective Coding

Analyzing relationships among subcategories reveals that acquisition identity restrictions, usage purpose restrictions, risk responsibility assumption, user rights restrictions, intellectual property ownership, and data rights ownership all influence usage behavior requirements, acquisition behavior requirements, or management behavior requirements, essentially achieving control by affecting user or platform behavior (see Table 4). Building on this, we analyzed relationships among main categories (see Figure 1 [Figure 1: see original paper]): behavior control plays a central role when scientific data platforms implement control strategies, while purpose, responsibility, identity, and rights controls also exert effects by influencing behavior.

2.3.4 Saturation Testing

When applying procedural grounded theory coding: saturation testing is essential for sampling-based coding but optional for full-sample coding [39]. Since this study requires understanding the status of scientific data control strategies—analyzing model components (represented by subcategories and main categories) and their proportions—full-sample coding is necessary, which can be considered saturated in this sense. Nevertheless, we conducted saturation testing: after coding the 10th sample, no new subcategories or main categories emerged; continuing through samples 11–13 likewise revealed no new categories. Additionally, we reserved Sample 14S for theoretical saturation testing, coding it separately after completing all other samples. Its coding results remained consistent with previously obtained subcategories and main categories (see Table 5). Through this dual verification, we conclude that coding saturation was achieved.

3.1 Three Behavior Strategies Involving Two Types of Actors

Behavior control refers to strategies that achieve control by stipulating or restricting actor behavior, involving two types: **Users**: Restricting usage or acquisition behavior through conditional constraints, summarized as "usage behavior requirements" and "acquisition behavior requirements"; **Platforms**: Detailing measures for managing scientific data to achieve orderly control, reflecting a positive control philosophy toward existing data, summarized as "management behavior requirements."

3.1.1 User Usage Behavior Control: Minimal Choice Space

"Usage behavior requirements" appears 46 times, accounting for 64.79% of behavior control strategies, indicating it is the universal practice and focal point of control strategies. Secondary classification reveals three types: prohibitions (24 occurrences, over half), obligations (18 occurrences), and permissions (minimal occurrences) (see Table 6). This shows platforms establish clear boundaries for usage behavior, with mandatory "prohibitions" and "obligations" comprising 91.30%, leaving users with minimal "permission" space.

3.1.2 User Acquisition Behavior Control: Achieved Through Application Procedures

"Acquisition behavior requirements" appears only 5 times (7.04% of behavior control strategies), indicating rare practical application. This likely stems from acquisition being merely one step in data usage, with user intentions focused on use. Most platforms, considering cost savings or avoiding redundancy, emphasize usage requirements over acquisition. Existing provisions reflect control through pre-application procedures: for example, "protocol users acquire agreement-shared data through procedures" sets requirements based on user levels; "confidential data is acquired after approved application" and "post-protection-period data is acquired through application" base requirements on data classification; "users may apply to member units for data" and "users may apply to the platform for data" base requirements on provider type. Despite different bases, all stipulate "apply before acquisition" for special circumstances, preventing direct data access without prior application.

3.1.3 Platform Management Behavior Control: Forming a Relatively Complete System

"Management behavior requirements" appears 20 times (28.17% of behavior control strategies), indicating some practice but insufficient application. Nevertheless, it forms a relatively complete system: numbering management, dedicated personnel, specialized equipment, regular review, publication decisions, update mechanisms, and preservation methods span collection, transmission, storage, processing, and exchange stages, with some detailed provisions facilitating daily management.

3.2 Other Control Strategies

3.2.1 Purpose Control: Commerciality as Criterion

Purpose control uses commerciality as its standard, permitting non-commercial users to access and use scientific data while excluding commercial purposes. Appearing 7 times (6.86% of all samples), it manifests in three patterns: **Reverse stipulation** (most common): "not for commercial purposes," which prohibits commercial use without specifying acceptable purposes; **Forward stipulation** (once): "must be for information acquisition purposes," which should be narrowly interpreted as "for knowledge only"; **General stipulation** (once): "strictly prohibited for any purpose not explicitly permitted," without providing an approved purpose list. Overall, reverse and forward stipulations are specific, with reverse offering maximum user space and forward offering minimum, while general stipulations are ambiguous and inconvenient for user judgment.

3.2.2 Rights Control: Ownership and User Rights Limitations

Rights control includes: intellectual property or data rights ownership of platform-provided scientific data; and what rights users have over acquired data. Appearing 9 times (8.82% of all samples), it comprises intellectual property ownership (5 occurrences) and data rights ownership (2 occurrences) across different platforms. Regarding actors, it involves platforms, original units, and original authors, with no consensus on ownership—some assign rights to platforms, others to original units, and some to joint platform-author ownership. The basis for ownership determination (platform investment) represents the platform's "unilateral statement" requiring further clarification of legitimacy. Additionally, user rights limitations appear twice, indicating minimal granted rights—essentially only usage rights with strict constraints, such as limited, non-exclusive rights to earthquake and meteorological data. Limitations like "limited" and "non-exclusive" prevent abuse or monopolization.

3.2.3 Identity Control: Qualification Levels Determine Accessible Data

Identity control refers to required identity qualification levels for accessing scientific data, such as registered users, real-name authenticated users, or protocol users. Appearing 9 times (8.82% of all samples), all provisions express acquisition identity restrictions—users with corresponding qualifications can only access corresponding data levels. These "levels" do not follow strict standards but rather platforms' self-defined classification systems, inapplicable to non-platform data. Obtaining corresponding identity requires secondary applications; for example, "protocol open/shared data limited to protocol users" requires users to apply for upgrade after becoming regular users, pending platform approval. Essentially, identity control strengthens platform control through multiple reviews.

3.2.4 Responsibility Control: Imbalanced Allocation Between Platform and Users

Responsibility control determines which party bears responsibility when adverse consequences occur. Appearing 6 times (5.88% of all samples), it divides into: **General provisions** like "users bear their own risks," assigning all risk to users; and **Specific provisions** like "bearing copyright infringement liability," explicitly informing users of consequences for unauthorized copyright information deletion. Analysis suggests general provisions aim to reduce platform review obligations; for example, "users bear all risks in receiving, relying on, or using platform data; NGDC assumes no legal liability for errors, inaccuracies, or omissions." Given that scientific data is provided by domain-specific institutions or scholars, platforms as intermediaries have limited capacity and cost for substantive professional review, making such provisions understandable. Transforming general into specific provisions that clarify responsible actors under different circumstances would help prevent adverse outcomes.

4 Optimization Recommendations

The "five-control-in-one" model (see Figure 1) is an "actual model" built from full-sample grounded theory extraction of behavior, rights, identity, purpose, and responsibility elements. While reflecting current realities, it offers optimization potential from a normative perspective. Examining the model through "pre-event, in-process, post-event" three-stage control theory reveals these five elements can be distributed across stages: rights, identity, and purpose controls focus on constraining actors before action, belonging to pre-event control; behavior control targets actions themselves, belonging to in-process control; responsibility control clarifies adverse consequence allocation after action, belonging to post-event control. Therefore, optimization could shift from the "five-control-in-one" to a "three-stage" model (see Figure 2 [Figure 2: see original paper]), achieving integrated pre-event prevention, in-process management, and post-event accountability. The "three-stage" model retains the five internal elements but adjusts their positional relationships (except for behavior) and proposes content optimizations accordingly.

The shift from scattered distribution of purpose, identity, rights, and responsibility around "behavior" to a three-stage distribution—pre-event (purpose, identity, rights), in-process (behavior), and post-event (responsibility)—does not change the core status of behavior control or the basic relationships among strategies. Other strategies still achieve control by influencing behavior (e.g., pre-event rights limitations affect in-process user behavior; post-event responsibility affects in-process platform management). The stage division aims to form a more orderly, traceable, and integrated scientific data control strategy system.

4.1 Pre-Event Defense: Strengthening Purpose, Rights, and Identity Controls

4.1.1 Strengthen Purpose Control: Summarize and Enumerate Purpose Types Users' true motivations and purposes for acquiring or using scientific data are highly implicit. Strengthening purpose control cannot simply prohibit commercial use but must specify concrete purpose types that are permissible or impermissible, enabling users to pre-determine whether they can access or use specific data for particular purposes. Simple reverse stipulations like "not for commercial purposes" are difficult for ordinary users to interpret. Commercial purpose cannot be determined solely by whether an entity has commercial status. We recommend adopting a "general + enumerated" model to list purpose types. The "general" component enhances interpretive space, while "enumeration" facilitates analogical judgment. Using profitability as the criterion, purposes can be divided into direct and indirect profit-making. Commercial purposes include but are not limited to paid sharing, marketing, or trading activities, directly or indirectly. Permissible purposes include non-profit research, experimentation, teaching, official duties, or public welfare.

4.1.2 Strengthen Rights Control: Clarify Data Rights Ownership Rights control involves intellectual property and data rights ownership. While intellectual property has legal stipulations, data rights lack legal provisions and remain in exploratory stages. Scientific data published on platforms does not automatically belong to platforms, users, or providers. Using the "Data Twenty Articles" [1] as policy guidance, we can define legitimate rights of providers, platforms, and users based on data sources and generation characteristics, establishing a data property rights operational mechanism covering holding, processing, and product management. Without considering the legitimacy of providers' data holding rights (providers do not lose holding rights by uploading data), platforms should be granted processing rights based on aggregated data and management rights over processed datasets; users should be granted processing and usage rights for platform-acquired data based on agreed purposes. This structural separation of property rights can stimulate the enthusiasm of different actors.

4.1.3 Reinforce Identity Control: Unify Hierarchical Classification Management

User classification holds value; for example, the TIS model framework provides theoretical guidance for user management in online user innovation communities, maximizing innovation value [40]. Identity control reflects platforms' hierarchical classification management philosophy, setting different user access permissions corresponding to different data levels to achieve limited, layered openness or sharing. Platforms without identity control strategies have imperfect internal data classification and user-level matching. Scientific data covers broad content with high specialization, diverse interests, and typically spans disciplines, departments, or industries, leading to inconsistent sensitivity and risk perceptions across platforms. Therefore, identity control improvement should proceed bottom-up: each platform should first implement classification methods suited to its actual conditions, continuing existing identity requirements to differentiate permissions among registered, real-name, and protocol users. The key to reinforcing identity control lies in unified hierarchical classification management, helping implement the "Data Twenty Articles" requirement to "control what should be controlled and open what should be opened" [1] and forming unified industry norms.

4.2 In-Process Management: Balancing Platform Management and User Control

4.2.1 Transform Platform Autonomy: Balance Management and User Control

User usage behavior requirements, user acquisition behavior requirements, and platform management behavior requirements, despite different actors, share the commonality of being self-defined by national-level scientific data platforms—collectively termed platform autonomy. The current focus of platform autonomy restricts user acquisition and usage behavior. However, data value lies in utilization, and the state actively promotes open sharing. More user behavior restrictions hinder greater openness. Platform autonomy should adhere to "control what should be controlled and open what should be opened," minimizing constraints and maximizing openness while ensuring security. Platforms bear responsibilities for maintaining platform and data security, necessitating user behavior control; simultaneously, as important carriers for promoting scientific data openness, they should impose fewer constraints. Therefore, platform autonomy should shift from "heavy user behavior control, light platform management" to "balanced platform management and user control."

4.2.2 Introduce Industry Heteronomy: Promote Unified Platform Awareness and Norms

Overall, only 14 national-level scientific data platforms provided user agreements on their official websites during sampling; others were either inaccessible or lacked relevant documents. Individually, platform behavior control strategies show more "individualization" than "standardization."

For example, some platforms stipulate acquisition behavior requirements while others do not; some concentrate on one or two control strategies while others include multiple strategies. Even with the common emphasis on behavior control, no consensus has formed, let alone unified norms. Therefore, we propose that the National Science and Technology Infrastructure Platform Center—a public institution directly under the Ministry of Science and Technology—take the lead within its "construction of normative standards and management methods" mandate to promote unified behavior management consensus and develop general behavior management norms. This industry heteronomy can ensure minimum requirements for platform behavior strategies while allowing individual platforms to add "personalized" elements atop general norms.

4.3 Post-Event Accountability: Improve Responsibility Strategy Allocation

Rational responsibility allocation means each actor bears responsibility corresponding to its actions, without excluding its own responsibilities or transferring them to others. Scientific data service providers (platforms), as bridges between data providers (researchers or institutions) and data users, connect the entire process from data generation and submission to usage. As previously discussed, providers, platforms, and users should enjoy corresponding data rights and bear corresponding responsibilities. Responsibility allocation is a crucial mechanism ensuring responsibility control effectiveness, requiring rational distribution among providers, platforms, and users. Data quality is closely related to providers, who should bear responsibility for ensuring data authenticity, validity, and format standardization. Platforms should ensure scientific data is findable, accessible, interoperable, and reusable [41], bearing fundamental responsibilities including data review, management, and security assurance. Users must acquire and use scientific data in accordance with laws, regulations, and platform rules, bearing responsibility for violations such as unauthorized transfer to third parties, failure to cite sources, or arbitrary data modification.

Scientific data neither involves complex personal interests like personal data nor carries rich property interests like enterprise data—perhaps due to its public welfare nature, it receives far less attention. However, scientific data constitutes important strategic resources for national scientific development, and scientific data platforms bear the vital mission of integrating data and providing services as crucial carriers for openness and sharing [17]. China has issued policy documents like the "Scientific Data Management Measures" and "National Science and Technology Resource Sharing Service Platform Management Measures," as well as national standards like the "Technical and Management Specifications for Scientific Data Submission from Science and Technology Plans," forming a basic institutional framework for scientific data governance. Establishing and improving scientific data control strategies is central to platform institutional construction and key to foundational data system development. Although academic research on scientific data is abundant, studies on control strategies remain insufficient.

This study employs grounded theory methodology, using user agreements from national-level scientific data platform practices as samples to investigate scientific data control strategies. Potential innovations and future research directions include: expanding the thematic breadth of scientific data research, with subsequent studies advancing theoretical construction of control strategies; extracting five constituent elements (purpose, behavior, rights, identity, responsibility) for future research to refine or propose new elements; summarizing the "five-control-in-one" model and optimizing it into a "three-stage" model for future refinement. Limitations include focusing only on national-level scientific data centers and databases, excluding local platforms like Shanghai R&D Public Service Platform or Heilongjiang Science and Technology Resource Sharing Service Center.

References

[1] The Central People's Government of the People's Republic of China. Opinions of the Central Committee of the CPC and the State Council on building basic systems for data to give full play to the role of data resources[EB/OL].[2023-10-01].https://www.gov.cn/zhengce/2022-12/19/content_{5732695}.htm.

[2] The Central People's Government of the People's Republic of China. The scientific data management rules[EB/OL].[2023-10-01].https://www.gov.cn/zhengce/content/2018-04/02/content_{5279272}.htm.

[3] DENG Z H, HUANG Y T. Research on the development of scientific data sharing platform at home in the "Internet+" environment[J]. Information studies: theory & application,2017,40(2):128-132.

[4] ZHANG X X. Research on development and service of scientific data sharing platform[J]. Journal of modern information,2016,36(11):109-113.

[5] LI Z C. Study on the current situation and development strategy of domestic scientific data sharing platform[J]. Library theory and practice,2018(8):108-112.

[6] SI L, WANG Y W. Current situation and suggestions of data organization on domestic scientific data sharing platforms—a study based on national science and technology infrastructure platform[J]. Library development,2018(10):52-58.

[7] ZHU L, MENG X X, ZHAO R X, et al. Research on resource construction for National Agricultural Science Data Sharing Center[J]. Digital library forum,2017(11):15-20.

[8] NAN Z T, WANG L X, WU L Z, et al. Experiences in data and knowledge integration of the environmental and ecological science data center for West China[J]. China science & technology resources review,2010,42(5):15-21,36.

[9] LI Y, WU D. Research on open medical scientific data platform[J]. Library and information service,2015,59(18):24-29,50.

[10] SUN Y H, LIU G F, CHEN S Y. Research on the construction of association model of national science data center based on the thematic data of Yellow River basin[J]. Digital library forum,2022(8):2-10.

[11] CHU J W, YANG T T. Study on the construction of cross-regional scientific data center in Yangtze River Delta[J]. Digital library forum,2022(6):58-64.

[12] LIAO X G, FENG Z M, GAO X, et al. The development of field scientific observation and research stations (networks) scientific centers[J]. Acta geographica sinica,2020,75(12):2669-2683.

[13] ZHANG J L, YIN S Q, ZHANG Y, et al. Social scientific data sharing and serving—an example of Fudan university social scientific data platform[J]. Journal of academic libraries,2015,33(1):74-79.

[14] XIAO X K. Research report on the construction of Hunan science and technology information scientific sharing platform[J]. Science & technology progress policy,2007(2):124-127.

[15] SI L, LI Y T, XING W M, et al. An empirical study on the performance evaluation of scientific data sharing platforms in China[J]. Library theory and practice,2014(9):30-35.

[16] SI L, HUA X Q. Analysis of the service effectiveness of national science and technology infrastructure platforms[J]. Library work and study,2014(4):24-26.

[17] WANG C P, SONG W Q, JIANG X Y, et al. Investigation and research on scientific data description and citation standards of domestic scientific data sharing platform[J].Research on library science,2022(11):76-85.

[18] LIU X Y, JI X M, WANG F. How does academic literature cite scientific data on the national science &technology infrastructure platform[J]. Journal of intelligence,2020,39(9):168-175.

[19] SUN W, LUO Z L, ZHANG W. Research on management of science data sharing platform[J]. Science of science and management of S.&T.,2005(12):32-36.

[20] WEN Y H, JIN M. Compliance assessment of user registration agreements of scientific data centers[J]. Documentation, information & knowledge,2022,39(1):119-129.

[21] WANG L X, WU L Z, NAN Z T, et al. Application of open source technologies in geoscientific data centers[J]. China science & technology resources review,2010,42(3):17-23,35.

[22] GAO F, WANG J. Study on the application of knowledge management in project management[J]. Science and technology management research,2012,32(16):184-187.

[23] LIU J Y, JIANG H, LIAO Y. Research and enlightenment from Germany scientific data center for the field of geoscience[J]. Digital library

forum,2019(12):52-58.

[24] WU S Z, WANG A R, XIU X L, et al. Construction and implications of biomedical scientific data centers in Europe and the United States[J]. Digital library forum,2022(4):2-10.

[25] ARORA R. Data management: state-of-the-practice at open-science data centers[M]. Handbook on data centers. New York: Springer,2015:1095-1108.

[26] DOWNS R R. Implementing the group on earth observations data management principles: observations of a scientific data center[J]. The international archives of the photogrammetry, remote sensing and spatial information sciences,2017,XLII-3/W2:51-55.

[27] ARDESTANI S B, HAKANSSON C J, LAURE E, et al. B2SHARE: an open eScience data sharing platform[C]//IEEE 11th international conference on e-Science. Munich:IEEE,2015:448-453.

[28] VIRT&L-COMM. VHERLA: a virtual molecular science data center allocated on the GARR cloud[EB/OL].[2023-08-19].http://services.chm.unipg.it/ojs/index.php/virtlcomm/article/

[29] KIM C W, YOON H, JIN D, et al. Integrated management system for a large computing resources in a scientific data center[J]. The journal of super computing,2016,72(9):35.

[30] M CINQUE, D COTRONEO, F FRATTINI, et al. To cloudify or not to cloudify: the question for a scientific data center[J]. IEEE transactions on cloud computing,2016,4(1):9.

[31] CAROSI A, LUCARELLI F, ANTONELLI L A, et al. The very high energy source catalog at the ASI science data center[C]//Software and Cyberinfrastructure for Astronomy IV. Edinburgh, SPIE Digital Library,2016:781-785.

[32] CHEN X M. Grounded theory: its train of thought and methods[J]. Educational research and experiment,1999(4):58-63,73.

[33] ZHU M, LIAO X Z. Theoretical construction of internalization of university library management system: an exploratory study based on grounded theory[J]. Library and information service,2019,63(5):32-41.

[34] FENG Y F, HU C P, QIU R R. Model construction of user privacy concerns for digital academic resources: a grounded theory study[J]. Information science,2019,37(3):3-8.

[35] WU X Y. Research on the influencing factors of college students' deep reading behavior based on mobile terminal: qualitative analysis based on grounded theory[J]. Library and information service,2021,65(24):80-88.

[36] WEN Y H, HE Y F. A survey of market entities' conception of data property rights and its correction[J]. Library tribune,2022,42(3):31-42.

[37] SHENG D F. The application of grounded theory in library and information science research in China[J]. Library tribune,2020,40(8):78-86.

[38] ZHANG J W, MA D J. Grounded theory research method and management research[J]. Modern management science,2009(2):115-117.

[39] WEN Y H, FU Z Y. Local legislation strategy and adjustment on the right of data elements in China based on grounded theory[J]. Library and information service,2023,67(7):53-66.

[40] YANG J. A framework of user classification model of online user innovation communities based on user innovation value[J]. Open journal of social sciences,2020,8(5):232-244.

[41] WILKINSON M D, DUMONTIER M, AALBERSBERG I J, et al. The FAIR guiding principles for scientific data management and stewardship[J]. Scientific data,2016,3(1):1-9.

Author Contributions:
Wen Yuheng: Proposed research topic and framework, drafted and revised manuscript.
Xu Kunshan: Collected and analyzed data, drafted initial manuscript.

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