# Strategic Choices for Scientific Data Security Management from an Intellectual Property Perspective (Postprint)

**Authors:** Luo Jiao, Liu Xiwen

**Date:** 2023-04-01T16:02:51+00:00

## Abstract

[目的/意义] Scientific data security is a complex and dynamic concept encompassing sovereignty security, property rights security, and sharing security, among which property rights security constitutes a crucial prerequisite for both sovereignty security and sharing security. Against the backdrop of the absence of data property rights legislation in China, employing intellectual property rights—which exhibit intersecting relationships with scientific data property rights under existing law—as an entry point to analyze property rights relationships in scientific data security management holds significant importance.

[方法/过程] Addressing the issues of absent property rights, deficient sovereignty, and hindered sharing of scientific data, this study utilizes legal norm analysis as its foundation, applies legal interpretation methods to examine the relationship between intellectual property rights and scientific data, clarifies that property rights represent a key measure for resolving scientific data security concerns, and explores potential solutions by incorporating domestic and international practical cases.

[结果/结论] Regarding property rights issues, intellectual property rights should serve as the foundation, supplemented by contracts, to safeguard the property rights security of scientific data through clauses on rights attribution, licensing authorization, confidentiality, and benefit distribution. Regarding sovereignty issues, a top-level national design for scientific data management should be constructed to enhance data sovereignty awareness, and institutional-level policies for scientific data management and sharing should be established to prevent unauthorized data outflow. Regarding sharing issues, obstacles to rights in the open sharing of scientific data should be cleared based on the characteristics of different types of intellectual property rights, and open sharing of scientific data should be realized through licensing agreements.

## Full Text

# Strategies for Scientific Data Security Management from the Perspective of Intellectual Property

**Luo Jiao**[1], **Liu Xiwen**[2,3]
[1]College of Humanities and Development Studies, China Agricultural University, Beijing 100083
[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] Scientific data security is a complex and dynamic concept encompassing sovereignty security, property rights security, and sharing security, with property rights security serving as a crucial prerequisite for both sovereignty and sharing security. Against the backdrop of China's absence of data property rights legislation, analyzing the property rights relationships in scientific data security management through the lens of intellectual property rights—which intersect significantly with scientific data property rights under current law—holds great significance. [Method/Process] Addressing the problems of absent property rights, deficient sovereignty, and obstructed sharing in scientific data, this study employs legal norm analysis and interpretation methods to clarify the relationship between intellectual property and scientific data, identifies property rights as the key to resolving scientific data security issues, and explores potential solutions by examining domestic and international practices. [Result/Conclusion] Regarding property rights, intellectual property rights should serve as the foundation, supplemented by contracts, to safeguard scientific data property rights security through ownership clauses, licensing clauses, confidentiality clauses, and benefit distribution clauses. Regarding sovereignty, top-level national design for scientific data management should be constructed to enhance data sovereignty awareness, and institutional-level scientific data management and sharing policies should be established to prevent illegal data outflow. Regarding sharing, legal barriers to open sharing of scientific data should be removed according to the characteristics of different types of intellectual property rights, and open sharing should be realized through licensing agreements.

**Keywords:** scientific data; intellectual property; contracts; data security

---

In the data era, data defines, connects, and transforms everything. Scientific research activities have also generated massive amounts of data. The research paradigm has evolved from "empirical science" that describes natural phenomena, to "theoretical science" that uses models or induction, to "computational science" that simulates complex phenomena, and now to "data exploration" that integrates theory, experiment, and simulation [1-2], forming a data-intensive

scientific discovery paradigm. Data from the European Large Hadron Collider confirmed the existence of the Higgs boson; genomic big data provides new opportunities for genomics research; and spatiotemporal big data plays a major role in global environmental change research [3]. "Data-driven science" has become a trend. As a strategic core resource [4-5], scientific data is a crucial resource for countries to gain or maintain their research advantages. In China, scientific big data serves as a cornerstone of the national big data strategy [6] and is regarded as an "important strategic resource" with "the fastest dissemination speed, widest influence, and greatest development potential" [7-8]. Its security issues have therefore attracted widespread attention, particularly concerning property rights, sovereignty, and sharing security. Given that China has not yet legislated on data property rights, this paper attempts to use intellectual property rights—which have substantial overlap with scientific data property rights under current law—as an entry point, employing legal norm analysis and interpretation methods to analyze the property rights relationships in scientific data security management.

## 1 Literature Review on Scientific Data Security Issues

Data refers to symbols that record and can identify objective events. In computer science, data means all symbols that can be input into computers and processed by computer programs [9]. Scientific data (Research Data or Scientific Data), also known as scientific and technological data [10], has been defined differently by scholars such as Chen Chuanfu [11] and Wang Xueqin [12], as well as various research institutions [13-19]. To avoid ambiguity, this paper specifically defines scientific data as symbols that record and can identify objective events, obtained and used in scientific research activities, including both original data and derived data.

Data security has different connotations across various dimensions. From an institutional perspective, it refers to protecting data from loss, leakage, illegal acquisition, modification, utilization, and damage through systems. From a morphological perspective, it concerns data storage and transmission security. From a content perspective, it involves hardware, software, system, and content security. From a subject perspective, it encompasses national, social, enterprise, and individual data security [20]. From an operational perspective, it refers to data integrity and utilization security [21]. From foreign legislative practice, confidentiality, integrity, and availability are generally considered the three essential elements that data security must satisfy [22]. From the perspective of national-level legislation in China, neither data security nor scientific data security has been defined.

Chinese scholars have explored scientific data security issues based on confidentiality, integrity, and availability, proposing security measures. Li Shanqing et al., from a technical perspective, suggest using integrity verification, access control, data encryption, privacy protection, and security auditing technologies to strengthen scientific data security [23]. Yang Yan et al., from a behavioral per-

spective, propose safeguarding data security by establishing behavioral norms for researchers throughout the processes of scientific data generation, collection, storage, description, analysis, utilization, and archiving [24]. Xiao Dongmei et al., from a subject perspective and in combination with cloud environments, suggest responding to scientific data security risks by developing cloud security industries, constructing data protection legal systems, and implementing government data supervision [25]. Sheng Xiaoping et al., from a governance perspective, point out that scientific data security governance requires multiple measures including strengthening legislation, formulating policies, building standards, creating systems, leveraging technology, and constructing institutional mechanisms [22]. Liu Guifeng et al., from a management perspective, emphasize constructing a scientific data security framework from "institutional," "infrastructure," "data literacy," and "implementation" layers based on the scientific data lifecycle [26]. These studies interpret scientific data security from different perspectives and offer profound insights for promoting scientific data security development in China. However, they mostly focus on discussing various measures while lacking analysis of the legal basis and rights foundation for each measure. Even when legal analysis is involved, the interpretation tends to focus on the normative level (what the law should be) rather than the positive level (what the law actually is).

Foreign research and practice attach greater importance to maintaining scientific data security from a property rights perspective. Related research explains the important value of intellectual property rights and contracts for secure data sharing [27] and plays a significant role in protecting researchers' rights and recognizing their contributions [28]. Harvard University uses scientific data property rights policies under the intellectual property framework as an important support for ensuring its scientific data security [29]; Cornell University emphasizes managing data property rights from a copyright perspective [30]; the Finnish Social Science Data Archive uses copyright and contracts for scientific data property rights management [31]; and the European Molecular Biology Laboratory's public database (EMBL-EBI) even directly warns users in its terms of use to pay attention to data property rights issues, particularly the legal uncertainty regarding data property rights that may interfere with scientific data reuse [33].

Given that possessing data control and analytical capabilities are important aspects of scientific data security, and that such capabilities are premised on actual control of scientific data, the direct legal basis for actual control is property rights in data. This paper attempts to use intellectual property rights—the rights closest to data interests under current law—as an entry point to analyze property rights, sovereignty, and sharing issues in scientific data, hoping to fill the research gap in this aspect of scientific data security management.

## 2 Challenges in Scientific Data Property Rights Management

### 2.1 The Problem of Absent Property Rights in Scientific Data

Who owns, uses, and benefits from scientific data concerns the property rights system for scientific data. The lack of a property rights system leading to unclear and lost property rights is the core issue in China's scientific data security management.

**2.1.1 Absence of Property Rights Leads to Unclear Ownership** The absence of data property rights legislation in China means there is no legal basis for determining ownership and distribution of data property rights. Article 127 of China's Civil Code includes data within its scope of protection, but this provision is merely a reference norm, leaving data property rights in a legislative vacuum. In 2018, the General Office of the State Council issued the "Administrative Measures for Scientific Data," which clarified principles and responsibilities for scientific data management but did not address the property rights system, leaving scientific data property rights without top-level design. In judicial practice, Chinese courts often use the Contract Law and Anti-Unfair Competition Law to resolve data property rights issues, but both approaches have limitations. Regarding Contract Law, on the one hand, contract remedies presuppose an existing contract, while scientific data collection, processing, sharing, and utilization often rely on institutional policies, industry guidelines, or academic ethics rather than comprehensive and rigorous contractual arrangements. On the other hand, given the relativity of contracts, they generally do not bind third parties outside the contract, making contracts alone insufficient to address infringement by third parties. Regarding the Anti-Unfair Competition Law, on the one hand, it only applies to market competition entities, while scientific data is not necessarily related to market competition. On the other hand, the law does not characterize data property rights disputes as unfair competition acts; courts often use Article 2, the "general clause," to handle data property rights issues. However, the general clause lacks specific applicable requirements and damage consequences, making it highly uncertain for resolving data property rights issues.

**2.1.2 Unclear Ownership Leads to Loss of Property Rights** The loss of scientific data property rights manifests in three aspects: first, scientific data is not properly preserved; second, scientific data is improperly disclosed; and third, scientific data is incorporated into others' rights control scope. First, the absence of a property rights system makes it difficult to form strong control over the possession, circulation, and use of scientific data through exclusive rights. Therefore, proper data preservation is the direct way to actually control data and the basic condition for subsequent development, utilization, exchange, and transaction. Scientific data not properly preserved means losing actual control, leading to data loss. As early as 2008, China's scientific research administrative

departments proposed mandatory archiving of scientific data formed in national projects, indicating that relevant departments were already aware that scientific data generated by research projects funded by large amounts of public funds were not properly preserved and were being damaged, lost, or flowing abroad [8]. Second, the absence of a property rights system forces some important data to be protected as trade secrets. Protecting scientific data as trade secrets requires the data to possess secrecy, confidentiality, and value. Once data is disclosed, it loses its secrecy and is no longer protected, which directly blocks scientific communication such as data publication alongside papers, data publication as data papers, and open data sharing. Third, the absence of a property rights system means that once scientific data is compiled into others' rights scope, it becomes others' rights. For example, when scientific data is compiled into a database in which others hold copyright, it becomes part of others' database works. Similarly, when data is incorporated into others' patented technical solutions, it becomes part of others' patent rights. In such cases, reusing this scientific data is constrained by others' rights.

### 2.2 The Problem of Deficient Sovereignty in Scientific Data

Scientific data sovereignty reflects a state's control capability and jurisdictional power over scientific data. A state's claim to sovereignty is premised on its nationals' actual possession of data or enjoyment of data property rights. For data that is neither actually possessed nor under property rights control, the state lacks legal basis to claim sovereignty. Therefore, data sovereignty is the foundation and prerequisite for realizing data property rights, while data property rights provide a strong basis for maintaining data sovereignty [34]. In China, scientific data sovereignty issues manifest as sovereignty deficiency caused by property rights absence, specifically shown in the outflow of scientific data abroad.

Scientific data outflow has attracted scholars' attention [35-36]. Existing research shows that scientific data outflow is mainly reflected in three aspects: first, the "siphon" effect of foreign academic journals, data journals, data platforms, and data centers is evident; second, cross-border data transmission in international scientific cooperation is non-compliant and illegal; third, cloud storage and computing make national boundaries increasingly blurred, posing challenges to data sovereignty. For example, submission policies of some top foreign academic journals require authors to submit data supporting their papers to designated databases (usually foreign databases), and researchers must comply to publish [37]. Similarly, some well-known foreign scientific data platforms or data centers are attracting Chinese scientists to store their data on these platforms [37]. Taking China's life and health big data as an example, Chinese scholars have published numerous SCI papers, most of whose data can only be submitted to internationally renowned databases such as NCBI and EBI. Over 25% of data in the NCBI database comes from China. Due to the lack of top-level design in scientific data management in China, such scientific data is seriously lost [38]. Finally, if scientific data is not archived in designated

domestic institutions, with the wide application of cloud technology, users can store and access cloud data via the Internet, but the ones who truly control the data are cloud storage service providers. Once cloud services are terminated for various reasons, the data may also be lost.

The security issues caused by massive scientific data outflow cannot be underestimated. First, massive outflow affects the security of scientific data itself. It creates obstacles for Chinese scientists to reuse such data, such as being constrained by others' intellectual property rights (e.g., copyright, database rights, patent rights) or being subject to others' policy conditions (e.g., requiring complex and lengthy ethical reviews, being forced to accept unreasonable and overly strict usage conditions). Second, massive outflow affects the security of scientific data utilization. Scientific and technological knowledge occupies a prominent position among various production factors, and its destructive power can pose great risks if improperly used by extremists [39]. For example, using genetic data and technology to amplify genetic defects in certain ethnic groups [40].

### 2.3 The Problem of Obstructed Sharing in Scientific Data

Research is not an island; data lives through flow, and data sharing is the foundation of science [41]. "Scientific data resources are important national wealth and productive forces; scientific data must be made to flow and be shared and applied" [8]. As early as 2001, China's Ministry of Science and Technology proposed the suggestion of "implementing a scientific data sharing project to enhance national scientific and technological innovation capability," and scientific data sharing projects in various fields were subsequently launched [42]. To date, China has achieved remarkable success in scientific data sharing, but further development remains constrained by root issues in culture, mechanisms, and technology [43]. Abroad, even NIH, considered a successful model of scientific data sharing, acknowledges in its latest data sharing policy that legal, ethical, and technical factors constrain the ability to preserve and share data [32].

The absence of property rights is one of the main obstacles to scientific data sharing, such as the difficulty in defining intellectual property rights [44] and the conflict between the exclusivity of intellectual property rights and the sharing nature of scientific data [45]. This paper argues that the absence of a property rights system hinders scientific data sharing for three reasons: first, scientific data holders worry that open sharing will lose control over data access and their competitive research advantage [22]; second, the effectiveness of data reuse is disrupted because data ownership uncertainty can easily lead to disputes; third, contributors to scientific data do not receive proper recognition, thereby losing potential opportunities to enhance influence and obtain research funding or awards. As a result, while China's total database resources rank among the world's top, the sharing and utilization efficiency is extremely low, lacking large-scale, high-quality data centers that lead internationally; databases cited more than 500 times are rare, and those cited more than 1,000 times are nonexistent

[38].

Data that should be shared but is not creates resource waste; data that should not be shared but is shared illegally affects data security. In 2015, BGI and Huashan Hospital of Fudan University illegally shared China's genetic data in international cooperative research with Oxford University [46-47]; in 2016, Suzhou WuXi AppTec illegally shared China's serum data with foreign parties [48]. These incidents not only caused data loss but also sounded the alarm for data sharing security. Therefore, some scholars point out that "science has no borders, but scientific data has borders. Balanced and moderate openness, maintaining the legitimate boundaries of scientific data openness, and establishing security review mechanisms for data sharing and foreign exchange are also important issues in scientific data open sharing" [49]. Thus, while sharing scientific data is a trend, it does not mean abandoning property rights and sovereignty. Clarifying property rights and asserting sovereignty are important for ensuring secure sharing of scientific data.

## 3 Using Intellectual Property as the Core Means to Ensure Scientific Data Property Rights Security

### 3.1 Using Intellectual Property Rights to Define Scientific Data Property Rights

The key to ensuring scientific data security is enhancing actual control over scientific data. Clarifying property rights, asserting sovereignty, and preventing illegal sharing will help strengthen the initiative and dominance in controlling, developing, and disseminating scientific data, gaining competitive advantages in international scientific research games. Under the current absence of a data property rights system, using intellectual property as a foundation to define the boundaries of scientific data property rights is a feasible measure. Under China's current legal environment, three types of intellectual property rights can serve as the basis for scientific data protection: copyright, patent rights, and trade secrets.

**3.1.1 Using Copyright as the Basis for Defining Scientific Data Property Rights** Copyright is an important right type for granting property rights to scientific data. Under China's Copyright Law, protection can be provided at three levels: first, data can be given original expression to obtain copyright protection. For example, drawings or photographs can reflect originality through composition, lighting, and focus, allowing the corresponding drawings or photos to obtain copyright. Otherwise, raw observation and experimental data belong to "facts" and are not protected by copyright. Second, copyright can be obtained through original selection and arrangement of data. For instance, if researchers demonstrate originality in selecting field names and arranging their order, even data in Excel spreadsheet form will receive compilation work copyright, but the copyright is limited to the dataset level and does not extend to the

data content itself. This means that if another researcher rearranges these data with different field names, it does not infringe the original dataset's copyright. Third, interpretations of data, data visualizations, metadata, and other products with sufficient originality can obtain copyright. Creating visualizations, graphics, charts, pictures, or other products "processed" from scientific data typically involves the processor's original expression and thus receives copyright protection, such as "black hole photos."

**3.1.2 Using Trade Secrets as the Basis for Defining Scientific Data Property Rights** Scientific data that constitutes secret information can be protected as trade secrets. According to China's Anti-Unfair Competition Law, trade secrets must possess secrecy, confidentiality, and value. Therefore, scientific data creators or controllers must adopt appropriate protective measures to keep data in a secret state to protect it as trade secrets. In practice, physical controls (e.g., access control) or technical controls (e.g., encryption or password protection) are typically used to protect the secret status of scientific data. Additionally, data secrecy and confidentiality can be guaranteed through confidentiality agreements. It should be noted that protecting scientific data as trade secrets conflicts with open sharing of scientific data, as open sharing means losing trade secret rights.

**3.1.3 Using Patent Rights as the Basis for Defining Scientific Data Property Rights** When scientific data has practical utility, it may constitute part of a patent or be patentable itself, such as genetic data or related patents. There are two ways to protect scientific data through patent rights: first, applying for a patent to include scientific data within the patent's protection scope. For such data that can constitute a patent or part of a patent, it is necessary to ensure the data is not disclosed. Confidentiality agreements should be established before obtaining the patent, and data should be strictly disclosed to specific parties according to the confidentiality agreement to prevent the data from becoming prior art due to public disclosure. Second, preventing others from applying for patents. Some researchers want their data to be free from legal restrictions, including patents, so that data can continue to be shared and reused. In this regard, publishing data to the public domain is sufficient to make the relevant data prior art, thereby preventing others from successfully applying for patents. However, even when data is publicly released, others may still use the disclosed data as part of their patent application technical solutions. If the invention is granted patent rights, the subsequent patent rights may be sufficient to prevent the actual use of that portion of scientific data constituting the invention [50-52]. In practice, contracts are generally used to prevent scientific data from being restricted by such patent rights. For example, publishing data through online databases requires users accessing the database to sign a "click-to-agree data use agreement," which can prohibit patent applications based on certain data or allow patent applications on the condition that the patent is not restrictive and must permit further use of the data. Alternatively,

agreements can stipulate that inventions or designs obtained using scientific data must license the included scientific data to others for free use according to open licensing agreements such as "CC licenses" [53-54].

### 3.2 Using Contracts to Perfect Scientific Data Property Rights Arrangements

Although scientific data can be brought within the protection scope of intellectual property rights such as copyright, patent rights, and trade secrets through certain forms, intellectual property rights and scientific data security have a cross-cutting rather than overlapping relationship. Intellectual property rights cannot cover all scientific data, and the property rights boundaries of scientific data need to be further clarified through contracts, as shown in Figure 1 [Figure 1: see original paper]. Confidentiality agreements, licensing agreements, and access agreements are basic contract types for protecting scientific data security.

Confidentiality agreements protect confidential data by agreeing on confidentiality obligations and measures to control information disclosure, thereby keeping relevant data in a secret state. Confidentiality agreements typically need to specify the rights holder of confidential information, what is considered confidential, the permitted scope of information use, obligations imposed on the disclosing party, and consequences for non-compliance.

Licensing agreements grant others permission to process databases or datasets and are key to secure sharing of scientific data. Without licensing, database managers cannot legally copy and provide any scientific data or datasets protected by exclusive rights to the public, nor can users use or process data in other ways. Achieving secure sharing of scientific data requires at least two licensing agreements: first, authorization from data rights holders to platforms (mostly databases) for storing and providing data; second, authorization from data rights holders to end users for using data. Licensing agreements can specify restrictions on license duration, licensed rights, geographical scope, and purposes.

Access agreements are typically used where researchers or research organizations can control data storage. Access agreements usually include: the scope of data to be accessed; the scope of personnel permitted to access data; whether access rights can be transferred to third parties (data providers generally declare that transfer is not allowed); whether accessed data is restricted in purpose (e.g., data cannot be used for commercial purposes, or commercial use requires compensation to data creators); disclaimers (data providers are not responsible for any errors in data); and consequences for access parties' non-compliance. Additionally, access agreements serve an important function—when data loses its secrecy and confidentiality agreements cannot control data access and use, access agreements can be used to control others' access and use of data.

# 4 Management and Policy Recommendations

## 4.1 Strengthening Scientific Data Property Rights Management

Using intellectual property rights as a foundation and contracts as a supplement to protect scientific data property rights security can be achieved by designing scientific data ownership clauses, licensing clauses, confidentiality clauses, and benefit distribution clauses.

**4.1.1 Distinguishing Between Basic Data and Derived Data in Ownership** For basic data, it should be clear who owns the content of basic data, the database, and all algorithms contained therein. Failure to declare this in advance leaves an obvious contractual loophole—ownership of relevant data may be challenged, such as the other party claiming the database is not legally protected. For derived data generated from secondary use of basic data, failure to clarify ownership not only affects the basic data owner's ability to authorize other parties but also causes data reusers' investment in data development to be wasted. Therefore, data reuse agreements can negotiate whether derived data belongs to the licensor, licensee, or is co-owned. Regardless of the ownership arrangement, care must be taken to avoid merging data owned with data merely authorized, otherwise one may inadvertently transfer one's own data to others.

**4.1.2 Clarifying Authorization Types, Scope, and Limitations** For licensors, consideration should be given to how to authorize data, such as exclusive licenses, sole licenses, or ordinary licenses. In some cases, the licensee's technology cannot achieve data development and utilization, and the licensee may wish to grant sublicenses to third parties to assist in data development. In this regard, licensors can restrict the licensee's sublicensing, stipulating that sublicensing is only for the licensee's benefit and only on the licensee's behalf. Additionally, since different datasets have different usage restrictions, data merging brings new risks. Therefore, licensors should also consider whether to allow licensees to merge licensed data with any other unauthorized data. Finally, to avoid omissions, licensors can stipulate that licenses for relevant data are limited to the authorization scope listed in the agreement, with all unlisted rights remaining with the licensor.

For licensees, they should first clarify whether authorization clauses can and how they can enable them to achieve currently envisioned and future foreseeable goals. If the licensee obtains an ordinary license, they need to consider whether to require the licensor not to re-license the data to the licensee's competitors, because once competitors master the same data, the licensee's competitiveness will be affected.

**4.1.3 Setting Confidentiality Clauses in Data Access and Reuse Agreements** Proper confidentiality clauses should at least include: first, requiring all parties to acknowledge data secrecy and not disclose it without explicit authorization; second, requiring all parties to adopt confidentiality measures for

data, such as deploying corresponding physical, management, and technical protective measures, to maintain data secrecy; third, limiting the licensee's ability to grant sublicenses, such as requiring the licensee to only provide data they have the right to disclose to third parties and requiring third parties to bear the same confidentiality obligations as between licensor and licensee.

#### 4.1.4 Balancing Interests Between Data Owners and Data Reusers

Both scientific data itself and its development and utilization have significant value and are resource-intensive work requiring substantial investment and resources. Therefore, a licensing fee distribution mechanism is crucial for maintaining balance. On the one hand, licensing fees cannot be too high, so that new licensees will continuously join the data development ecosystem. On the other hand, licensing fees cannot be too low, to ensure that licensors receive returns on their data investment. Benefit distribution design should maintain a balance that neither excessively raises entry costs for data development and utilization, causing related R&D to stagnate, nor excessively reduces data owners' profit expectations, making them unwilling to continue collecting, maintaining, and providing data. Specific measures can include innovative licensing fee payment models, such as establishing prepayment models, installment payment models, and setting sublicensing profit-sharing clauses. Additionally, data development and utilization require not only substantial investment but also considerable time. Once developers become dependent on data, a "seller's market" is formed, and the possibility of developers obtaining data at reasonable costs will be greatly reduced. Therefore, it is necessary to include price protection clauses in licensing fee terms, such as requiring data owners not to increase licensing fees within a fixed period; if fees are increased after the fixed period, sufficient written notice must be given, and the fee increase must not exceed a certain percentage.

### 4.2 Enhancing Scientific Data Sovereignty Awareness

#### 4.2.1 Constructing National-Level Top-Level Design for Scientific Data Management and Emphasizing Data Property Rights Security

Formulating national principles for scientific data property rights management at the national level is an effective experience in foreign scientific data management. For example, the Australian Research Council (ARC), together with the Australian Government Department of Industry, Science and Resources and IP Australia, formulated the "National Principles of Intellectual Property Management for Publicly Funded Research," emphasizing maximizing national interests and return on investment in public research [55]. Similarly, the OECD Committee for Scientific and Technological Policy's "Declaration on Access to Research Data from Public Funding" emphasizes that scientific data access mechanisms should fully attend to national security, privacy, and trade secret requirements under domestic law, and clarify the main responsibilities of all parties involved in data activities [56].

**4.2.2 Establishing Institutional-Level Scientific Data Management and Sharing Policies to Avoid Illegal Data Outflow** Institutions can establish scientific data management and sharing policies according to data categories as appropriate. For example, the World Health Organization (WHO) has formulated separate "Data Sharing Policy for Public Health Emergencies" [57] and "Data Sharing Policy for Non-Public Health Emergencies" [58]. Simultaneously, institutions should stipulate scientific data localization storage and other security management measures in their data management and policies. For instance, the U.S. National Institutes of Health (NIH) strongly encourages using its established repositories to preserve and share scientific data in its latest data management policy and requires researchers to explain in data management plans how to manage scientific data generated by research projects and which scientific data and accompanying metadata to share [59].

### 4.3 Promoting Open Sharing of Scientific Data

**4.3.1 Using Licensing Agreements to Achieve Secure Sharing of Scientific Data** Licensing agreements can apply in two situations: first, when one or more intellectual property rights cover scientific data, rights holders can authorize others to reuse data through licensing agreements. Second, although scientific data is not within the protection scope of intellectual property rights—for example, large sensor data collections arranged in chronological order—the acquisition of such data requires access to or download from data platforms associated with "access agreements" and "terms of use." In this case, these "access agreements" and "terms of use" can still serve as "contracts" to restrict data acquisition and reuse. The second situation is equivalent to imposing contractual constraints on data in the public domain. Therefore, the design of such "access agreements" and "terms of use" should take open scientific data as the principle and non-open as the exception.

**4.3.2 Using Legal Mechanisms to Remove Rights Barriers to Open Sharing of Scientific Data** For trade secret rights, public disclosure can eliminate trade secret rights on data. Other forms of sharing can only share data within agreed scopes under specific confidential forms or confidentiality agreements. For patent rights, as mentioned above, public disclosure can make data lose novelty, thereby limiting or destroying the patentability of inventions related to the data. When patents cover scientific data, although the corresponding data is disclosed with the patent application documents, reusing the data if it constitutes implementation of the relevant patent requires the patentee's permission. For copyright, in open data practice, "CC licenses" and other open licensing agreements are generally used to set access and use permissions and conditions for data or datasets protected by copyright. The essence of "CC licenses" is to grant users the right to engage in behaviors controlled by copyright without imposing additional obligations on users, thereby achieving open sharing of scientific data. The Open Knowledge Foundation (OKFN) describes the benefits of such agreements as "allowing others to freely and broadly license

works; others do not need to seek permission time-consumingly each time they wish to use or disseminate works; encouraging others to continuously add value to works; and encouraging others to create new works based on original works or from original works" [60]. In addition to CC licenses, the GNU Free Documentation License (GFDL) and Open Data Commons (ODC) licenses [61] are also widely used in open data practice. Most open licensing agreements support online access and can be easily attached to data.

Open sharing of scientific data in specific fields, such as medicine, also needs to address privacy and personal information issues. For data containing personal information, privacy, and other sensitive information, publication and reuse may be restricted by privacy rights and personal information rights [62]. The best approach is to adopt the highest level of privacy compliance standards where feasible, supervise data collection and sharing throughout the process to ensure security and privacy, and take licensing measures or desensitization measures before sharing, such as seeking consent from data subjects and anonymizing or fuzzifying data. Once desensitization measures are found to be ineffective or potentially ineffective, relevant data should be immediately withdrawn. Additionally, as data analysis technology advances, traditional anonymization or fuzzification measures may become ineffective. It is recommended to only open data analysis results for such data, not opening or cautiously opening raw data.

## Conclusion

In the data era, the fourth paradigm of scientific research—data-intensive scientific research—has emerged, and scientific data has become an important strategic resource of the nation, making it imperative to maintain its security. The core of scientific data security lies in enhancing the ability to control and apply scientific data, premised on ensuring secure storage, secure use, secure sharing, and secure services of scientific data. However, long-term absence of property rights, sovereignty, and sharing has laid hidden dangers for scientific data security.

Under the current situation where data property rights legislation is still blank, intellectual property rights, as the rights closest to data interests, using intellectual property as a foundation and contracts to clarify scientific data property rights, maintain scientific data sovereignty, and regulate scientific data sharing has certain operability. However, intellectual property rights are essentially private rights, with "rational economic man" as their logical starting point and "rights-based" as their value foundation, which is far from the internal mechanism of "decentralization, flattening, and borderlessness" in the digital era and the scientific spirit of "openness, cooperation, and sharing." Using intellectual property rights to maintain the security boundaries of scientific data is only a temporary measure. "Laws change with time to achieve good governance." Creating "data rights" and constructing a brand-new system may be necessary to truly achieve "making the best use of data and resolving disputes" and defending scientific data security.

## References

[1] Deng Zhonghua, Li Zhifang. The evolution of scientific research paradigms—The fourth paradigm of scientific research in the big data era [J]. Information and Documentation Services, 2013(4): 19-23.

[2] HEY T, TANSLEY S, TOLLE K. The fourth paradigm: data-intensive scientific discovery [M]. Redmond: Microsoft Research, 2009.

[3] Guo Huadong, Wang Lizhe, Chen Fang, et al. Scientific big data and digital earth [J]. Chinese Science Bulletin, 2014, 59(12): 1047-1054.

[4] CUKIER K, SCHOENBERGER M V. The rise of big data: how it's changing the way we think about the world [J]. Foreign affairs, 2013(92): 28-33.

[5] Obama White House. Big data research and development initiative [EB/OL]. [2020-11-01]. https://obamawhitehouse.archives.gov/blog/2012/03/29/big-data-big-deal.

[6] Guo Huadong. Scientific big data—The cornerstone of the national big data strategy [J]. Bulletin of Chinese Academy of Sciences, 2018, 33(8): 768-773.

[7] Chinese Government Network. Ensure security, highlight sharing, support innovation—How to manage scientific data scientifically [EB/OL]. [2020-11-01]. http://www.gov.cn/zhengce/2018-04/08/content_{5280429}.htm.

[8] Chinese Government Network. Scientific data is an important strategic resource [EB/OL]. [2020-11-01]. http://www.gov.cn/zhengce/2018-04/06/content_{5280211}.htm.

[9] Ye Bin, Huang Hongqiao, Yu Yang. Fundamentals of information technology [M]. Chongqing: Chongqing University Press, 2017.

[10] Sun Jiulin, Huang Dingcheng, Li Xiaobo. New progress in scientific and technological data management and sharing services in China [J]. World Sci-tech R&D, 2002, 24(5): 15-19.

[11] Chen Chuanfu. Mechanism for public access to scientific data in China: Characteristics, obstacles and optimization suggestions [J]. China Soft Science, 2004(2): 8-13.

[12] Wang Xueqin, STOUT A, SILVER H. Building data-driven e-Science library services: Opportunities and challenges [J]. Library and Information Service, 2011, 55(3): 80-83.

[13] NSF. NSB-05-40 Long-lived digital data collections: enabling research and education in the 21st century [EB/OL]. [2020-11-01]. http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf.

[14] ANU. ANU data management manual: managing digital research data at the Australian National University [EB/OL]. [2020-11-01]. http://regnet.anu.edu.au/sites/default/files/files/AN

[15] OECD. OECD principles and guidelines for access to research data from public funding [EB/OL]. [2020-11-01]. http://www.oecd.org/dataoecd/9/61/38500813.pdf.

[16] University of Cambridge. Explanation of terms [EB/OL]. [2020-11-01]. http://www.lib.cam.ac.uk/preservation/incremental/glossary.html.

[17] Agricultural Science Data Sharing Center. Introduction [EB/OL]. [2020-11-01]. http://www.agridata.cn/homepage/chintro.asp.

[18] Surveying and Mapping Science Data Sharing Service Network. Introduction [EB/OL]. [2020-11-01]. http://sms.webmap.cn/.

[19] Scientific data sharing concepts and terminology [EB/OL]. [2020-11-11]. http://www.sciencedata.cn/pdf/2.pdf.

[20] Qi Aimin, Pan Jia. Data rights, data sovereignty and basic principles of big data protection [J]. Journal of Soochow University (Philosophy & Social Science Edition), 2015, 36(1): 64-70, 191.

[21] Zhang Ni, Zhang Yunyong, Hu Kun. Big data security technology and application [M]. Beijing: People's Posts and Telecommunications Press, 2014.

[22] Sheng Xiaoping, Guo Daosheng. Research on data security governance in scientific data open sharing [J]. Library and Information Service, 2020, 64(22): 25-36.

[23] Li Shanqing, Zheng Yanning, Xing Xiaozhao, et al. Research on security management issues in scientific data sharing [J]. China Science & Technology Resources Review, 2019, 51(3): 11-17.

[24] Yang Yan, Ruan Jianhai. Research on scientific data security behavior based on the research process [J]. Knowledge Management Forum, 2019, 4(4): 218-231.

[25] Xiao Dongmei, Sun Lei. Security risks and governance countermeasures of scientific data in cloud environments [J/OL]. Library Tribune, 2021(2): 1-10.

[26] Liu Guifeng, Ruan Bingying, Bao Xiang. Construction of a scientific data security content framework in universities from the perspective of data lifecycle [J]. Journal of Intelligence, 2021, 40(2): 1-8.

[27] CARROLL M W. Sharing research data and intellectual property law: a primer [EB/OL]. [2020-03-12]. https://doi.org/10.1371/journal.pbio.1002235.

[28] CONTRERAS J L. Data sharing, latency variables and science commons [J]. Berkeley technology law journal, 2010, 25(11): 1601-1672.

[29] Harvard University. Research data ownership policy [EB/OL]. [2020-03-12]. https://vpr.harvard.edu/data-ownership.

[30] Cornell University. Research data management service group data management planning best practices [EB/OL]. [2020-03-12]. https://data.research.cornell.edu/content/intellectual-property.

[31] FSS. Finnish social science data archive data management guidelines [EB/OL]. [2020-03-12]. https://www.fsd.tuni.fi/en/.

[32] NIH. Final NIH policy for data management and sharing [EB/OL]. [2021-03-12]. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html.

[33] EMBL-EBI. Terms of use of the EBI Services [EB/OL]. [2020-03-12]. http://www.ebi.ac.uk/about/terms-of-use.

[34] TRACHTMAN P J. Cyberspace, sovereignty, jurisdiction and modernism [J]. Indiana journals of global legal studies, 1998, 5(2): 561-582.

[35] Zeng Jianxun, Yang Daiqing. Policy thinking on reversing the outflow of Chinese scientific papers [J]. Acta Editologica, 2020, 32(6): 600-604.

[36] Li Yang, Wen Liangming. Outflow of scientific data in China: Performance, problems and countermeasures [J]. Library Journal, 2019, 38(12): 72-81, 115.

[37] National Science and Technology Infrastructure Center. National scientific data resources development report (2016) [M]. Beijing: Scientific and Technical Documentation Press, 2016.

[38] Bao Yiming, Xue Yongbiao. Status and prospects of life and health big data [J]. Bulletin of Chinese Academy of Sciences, 2018, 33(8): 861-865.

[39] Zhong Cantao. Openness and confidentiality: Control of scientific and technological information dissemination and its impact on innovation—Taking the U.S. scientific and technological information dissemination control mechanism as an example [J]. Studies in Science of Science, 2013, 31(3): 335-343.

[40] VOGEL G. German law allows use of DNA to predict suspects' looks [J]. Science, 2018, 360(6391): 841-842.

[41] RAIK A. Regulating scientific research: intellectual property rights and the norms of science [J]. Northwestern university law review, 1999, 94(1): 77-152.

[42] Yang Lan. Open scientific data for common development [EB/OL]. [2020-11-01]. http://finance.china.com.cn/roll/20140424/2358833.shtml.

[43] PENG C, SONG X, JIANG H, et al. Towards a paradigm for open and free sharing of scientific data on global change science in China [EB/OL]. [2021-02-11]. https://esajournals.onlinelibrary.wiley.com/doi/full/10.1002/ehs2.1225.

[44] Liu Runda, Sun Jiulin, Liao Shunbao. Preliminary study on data authorization issues in scientific data sharing [J]. Journal of Wuhan University of Technology, 2007(7): 164-167.

[45] Xu Xiandong, Zhu Xuezhong. Research on the conflict between intellectual property exclusivity and scientific data sharing [J]. Journal of Wuhan University of Technology, 2007(7): 164-167.

[46] Ministry of Science and Technology of the People's Republic of China. Administrative penalty decision: Guo Ke Fa [2015] No. 1 [EB/OL]. [2020-11-01]. http://www.most.gov.cn/bszn/new/rlyc/xzcf/201810/t20181011_{142042}.htm.

[47] Ministry of Science and Technology of the People's Republic of China. Administrative penalty decision: Guo Ke Fa [2015] No. 2 [EB/OL]. [2020-11-01]. http://www.most.gov.cn/bszn/new/rlyc/xzcf/201810/t20181011_{142043}.htm.

[48] Ministry of Science and Technology of the People's Republic of China. Administrative penalty decision: Guo Ke Fa [2016] No. 1 [EB/OL]. [2020-11-01]. http://www.most.gov.cn/bszn/new/rlyc/xzcf/201810/t20181011_{142045}.htm.

[49] Wang Ruidan, Yang Jing, Gao Mengxu. Thoughts on strengthening and standardizing scientific data management in China [J]. China Science & Technology Resources Review, 2018, 50(2): 1-5.

[50] GITTER M D. Resolving the open source paradox in biotechnology: a proposal for a revised open source policy for publicly funded genomic databases [EB/OL]. [2020-11-01]. http://ssrn.com/abstract=901994.

[51] EISENBERG S R, RAY K A. Harnessing and sharing the benefits of state-sponsored research: IP rights and data sharing in California's stem cell initiative [J]. Berkeley technology law journal, 2006, 21(11): 1187-1213.

[52] DRISCOLL C T. NIH data and resources sharing, data release and intellectual property policies for genomics community resource projects [J]. Expert opinion on therapeutic patents, 2005, 15(1): 1-6.

[53] CAMBIA. About biOS (biological open source) licenses and MTAs [EB/OL]. [2020-11-01]. http://www.bios.net/daisy/bios/licenses/398.html.

[54] JEFFERSON R. Science as social enterprise: the CAMBIA biOS initiative [J]. Innovations: technology, governance, globalization, 2006, 1(4): 13-44.

[55] The Australian research council et al. IP Australia, the national health and medical research council, national principles of intellectual property management for publicly funded research [EB/OL]. [2020-11-01]. http://www.arc.gov.au/pdf/01_{01}.pdf.

[56] OECD. Declaration on access to research data from public funding [EB/OL]. [2020-11-01]. https://legalinstruments.oecd.org/public/doc/157/157.en.pdf.

[57] WHO. Policy statement on data sharing by the world health organization in the context of public health emergencies [EB/OL]. [2021-02-10]. https://www.who.int/about/who-we-are/publishing-policies/data-policy.

[58] WHO. Policy statement developing global norms for sharing data and results during public health emergencies (as of 13 April 2016) [EB/OL]. [2021-02-10]. https://www.who.int/wer/2016/wer9118/en/ or www.who.int/ihr/procedures/SPG_{{{data}}_{{sharing}}}.

[59] NIH. Policy for data management and sharing [EB/OL]. [2021-02-10]. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html.

[60] OKFN. Open knowledge foundation [EB/OL]. [2020-11-01]. http://www.opendefinition.org/guide?action=

[61] Luo Jiao. Copyright solution for open data—Research on ODC licenses [J]. Library and Information Service, 2017, 61(9): 23-32.

[62] Luo Jiao. Research on legal issues of personal information protection in big data environment [J]. Library, 2018(5): 31-36.

**Author Contributions:**
Luo Jiao: Designed the paper framework and wrote the initial draft.
Liu Xiwen: Proposed the research topic, designed the research framework, revised the manuscript, and finalized the paper.

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