

Scientific Data Intellectual Property Research by Literature Statistics

ZHANG Jia, WANG Xirong

University of Shanghai for Science and Technology, Shanghai, China

The main obstacle to the open sharing of scientific data is the lack of a legal protection system for intellectual property. This article analyzes the progress of research papers on intellectual property in scientific data in China through literature search and statistics. Currently, research subjects are unbalanced, research content is uneven, research methods are intellectual single, and research depth is insufficient. It is recommended that different stakeholders engage in deep cross disciplinary cooperation, further improve China's legal and policy protection system for scientific data intellectual property, and promote the open sharing of scientific data.

Keywords: scientific data, intellectual property, bibliometrics, stakeholders, interdisciplinary

Introduction

In the era of big data, with the rise of the fourth research paradigm—data intensive scientific research, scientific data, as a strategic resource, has been widely recognized in the world. Scientific and technological innovation increasingly depends on the analysis, mining, and comprehensive utilization of scientific data. Scientific data mainly includes data generated from basic research, applied research, experimental development, and other fields such as natural science and engineering technology science, as well as original data and its derivative data obtained through observation and monitoring, investigation, inspection, and detection and used in scientific research activities (General Office of the State Council, 2018). Scientific data, also known as research data, is regarded as an important scientific research output like scientific research papers. At present, many high-value scientific data have not been fully shared and used. Only by making scientific data truly flow can it play its potential great value. Intellectual property is legally recognized property rights for the achievements created by people engaged in intellectual labor in the fields of science, technology, culture, art, etc. (Institute of Economics, Chinese Academy of Social Sciences, 2005). Many researchers believe that the obstacle affecting the real flow of scientific data is the lack of protection of intellectual property legal system (Sheng & Wu, 2019a). This paper intends to analyze the progress of research on the subject of intellectual property of scientific data in mainland China through literature retrieval and statistics.

Literature Research

Sample Selection and Literature Sources

This research takes the academic journal database of CNKI as the retrieval source (the research results of

ZHANG Jia, Master, librarian, Library, University of Shanghai for Science and Technology, Shanghai, China.

WANG Xirong, Ph.D., Master tutor, research librarian, Library, University of Shanghai for Science and Technology, Shanghai, China.

general master thesis and conference thesis will also be reflected in academic journals).

Under the “advanced retrieval” item of CNKI “academic journals”, the retrieval condition is set as “subject: scientific data + scientific research data” (hereinafter referred to as “scientific data subject research”), and 8,268 literatures are retrieved. In order to further study the research results related to the intellectual property of scientific data, a parallel search item “subject: intellectual property” was added to the search conditions on the basis of the original set subject, and 254 literatures were retrieved.

Data Processing

This paper focuses on the text analysis of 254 papers on the subject of scientific data intellectual property. After sorting out, the repeated included papers were deleted, and the non-research articles such as notice of solicitation, interviews with people, and news manuscripts were removed. Finally, 232 papers were obtained.

Statistical Results

Publication Time and Quantity

China has a short history of developing the intellectual property system, and scientific data, as an independent discipline, has not been formally established by the scientific community until the 21st century (Sun & Liu, 2003). The term “scientific data” first appeared in 1981, and the research on the subject of intellectual property of scientific data began in 2003. The papers published in the past five years account for more than 50% of the total number of papers, but the research on the theme of intellectual property of scientific data accounts for only 3% of the research on the theme of scientific data.

Analysis of Cooperation Mode

Table 1

Statistical Table of the Number of Paper Authors

Number of authors	Number of published papers	Percentage
1	70	30%
2	83	36%
3	32	14%
4	26	11%
5	13	6%
6-29	8	3%

Table 2

Statistical Table of Paper Cooperation

Cooperation	Number of papers issued	Percentage
Independent author (no cooperation)	70	30%
Multi person cooperation in the same institutions	102	44%
Cross institutions cooperation	55	24%
Interdisciplinary collaboration	5	2%

The number of authors and cooperation of 232 papers are shown in Table 1 and Table 2. There are only 70 papers signed by one author, accounting for 30% of the total; 83 papers were signed by two people, accounting for 36% of the total number; 79 articles with three or more people, accounting for 34% of the total; the article with the largest number of authors was 29.

Of the 232 articles, 174 were from the same research institutions, accounting for 75% of the total. Among the 58 papers on cross institutions cooperation, 53 papers belong to the same research field, accounting for 22% of the total; the authors' interdisciplinary research achievements in different research fields are only five, accounting for 2% of the total. That is to say, 99% of the researchers are from the same field.

Analysis of Research Institutions

Referring to the classification of different stakeholders in Sheng Xiaoping and Wu Tong (2019a) and combining with the actual situation of this study, author institutions are classified and counted according to library and information institutions, other research institutions, data centers, publishing houses, and legal institutions. The same names are consolidated. The statistical results are shown in Table 3. The data show that a total of 263 research institutions participate in the theme research of intellectual property of scientific data. Among them, the number of institutions belonging to library and information institutions is the largest, with 115; other research institutions took the second, with 97; 19 in the field of law, six in data centers, nine in publishing houses, and 17 in other fields.

Table 3

Author Organization Classification Statistics

Ranking	Stakeholder types	Number
1	Library and information institutions	115
2	Other research institutions	97
3	Legal institutions	19
4	Data centers	6
5	Publishing houses	9
6	Others	17
	Total	263

Researcher Analysis

Table 4

Statistics on the Number of Published Papers and the Number of People

Number of published papers	Number of authors	Percentage
1	446	85.6%
2	57	10.9%
3	6	1.2%
4	5	1%
6-8	7	1.3%

There are 521 authors in 232 papers. Ranked by the number of papers published by the authors, 18 authors publish three or more scientific data on intellectual property topics; 57 authors have published two papers; there were 446 authors who publish only one article, accounting for 85.6% of the study population (Table 4). There are many authors conducting research on the topic of scientific data intellectual property, but not many have conducted in-depth research. The Chinese Academy of Sciences system and universities are the two main camps for conducting research on scientific data intellectual property.

Analysis of Fund Literature Distribution

Of the 232 papers, 163 were funded projects, accounting for 70% of the total. Among the 163 funded papers, 53 were funded by the National Social Science Fund of China (NSSFC) and the National Natural Science

Foundation of China (NSFC), 14 were funded by the National Science and Technology Infrastructure Platform Project (including six funded by NSFC), and 102 were funded by provincial and ministerial-level projects and associations. It can be seen that the Chinese government and science and technology management departments at all levels have promoted the overall level of scientific data openness and sharing in China by formulating national laws and regulations as well as government-backed fund support, providing macro guidance for scientific data management.

Table 5 shows the classification statistics of 53 project papers funded by NSFC and NSSFC. The most funded papers are based on the communication link of scientific data life cycle, involving data citation, data publication, data supervision and traceability, data security content, institutional alliances, institutional warehousing, as well as scientific data research in different disciplines such as materials and ocean.

Table 5

Classification and Statistics of the Content of Papers Funded by NSFC and NSSFC

Description of research content	Number of papers	Percentage
Scientific data management experience in other countries	10	19%
Research based on the current intellectual property legal framework	14	26%
Research on the specific implementation level of open sharing of scientific data	23	43%
Others	6	11%

Based on Text Content Analysis

Based on the classification of different stakeholders, the 232 papers were divided into five fields according to the nature of the first author's institution: library and information institutions, other research institutions, data centers (national infrastructure platforms), publishing houses, and legal research institutions. The research priorities of different stakeholders in the field of scientific data intellectual property research were analyzed.

Table 6

Statistics of the Number of Papers From Different Stakeholders

Classification of different stakeholders	Number of papers by first author's institution	Percentage
Library and information institutions	146	63%
Other research institutions	65	28%
Legal institutions	14	6%
Publishing houses	5	2%
Data centers	2	1%

As can be seen from Table 6, library and information institutions are the main force in the research on the subject of intellectual property of scientific data, accounting for 63% of the total research results; other major research forces from high to low are other research institutions accounting for 28%, law research institutions accounting for 6%, publishing houses accounting for 2%, and data centers accounting for 1%.

Library and Information Institutions

From the 146 papers of the library and information institutions, the research contents mainly include: (1) 42 papers, accounting for 29% of the research results of the library and information institutions, investigated the foreign macro legal system, sharing policies, and scientific data services of foreign university libraries. In particular, the research results on the current situation of scientific data management and service in foreign university libraries are the most. (2) Research on China's macro level legal policy system: This paper sorts out

the ownership of scientific data from the legal level, studies the monopoly of scientific data and the outflow of scientific data, and suggests improving the legal environment for public access to information resources and building a special scientific data management and sharing intellectual property system. (3) Carry out research on the practice of scientific data sharing, and explore the licensing and sharing mechanism of scientific data based on the current intellectual property legal system. (4) Study the common problems of scientific data life cycle: The content involves scientific data rights management, data publishing, data citation, knowledge link system, data monitoring model, scientific data security content framework, etc. (5) University libraries seek transformation and development, actively expand the service field, propose scientific data management strategies, and explore the development path of library scientific data services in the aspects of subject services, scientific data literacy education, small data and long tail data research, blockchain technology application, institutional repository alliance, and institutional storage.

Other Research Institutions

65 papers from other research institutions account for 28% of the total number of papers, and they are also an important force in the subject research of scientific data intellectual property. From the content of 65 papers from other research institutions, the research shows the following characteristics. (1) Research institutions have conducted in-depth research in specific disciplines, including materials science data, medical science data, environmental and ecological science data, geological science data, agricultural science observation data, marine environment monitoring and prediction data, military research self-produced data, chemical professional data, national population and health science data, etc. (2) The research on the subject of intellectual property of scientific data started early and from a high starting point. In 2002, China proposed the scientific data sharing project. Academicians SUN Jiulin and SUN Honglie and other scholars actively promoted the opening and sharing of scientific data in China. Academician Sun Honglie and Liu Chuang (2003) proposed to establish the rules and regulations and legal system for scientific data management and sharing in China. Academician Sun Jiulin (2003) proposed 10 strategic measures for scientific data sharing, including the policy and regulation system for scientific data sharing. (3) In the related research on the opening and sharing of scientific data, some researchers proposed to build the framework of scientific data publishing system and carry out the research on scientific data publishing and citation. In general, other research institutions have less research on foreign macro legal policies, and there are not many papers related to the policies of foreign science funding organizations. The focus is to carry out relevant research in the field of professional disciplines.

Data Centers

Data center is sometimes called data sharing center or data sharing platform (Sheng & Wu, 2019b). Two papers with data center or National Science and Technology Infrastructure Platform Center as the first author organization accounted for 1% of the research sample. The National Science and Technology Infrastructure Platform Center and its funded research projects are mainly carried out at the macro legislative level. In 2003, the National Science and Technology Infrastructure Platform Center launched the project of “Research on legislation for the protection and sharing of science and technology resources”. Some researchers proposed to formulate a unified law on the sharing of science and technology resources (Ma, Zhang, & Gao, 2007). According to the current situation of scientific data management of NSFC, some researchers put forward policy suggestions to improve the data management and sharing of NSFC projects (Zhou, Liu, & Wang, 2023). Other research contents include not only the overall thinking of the strategic orientation of scientific data sharing project, but

also the exploration of the legislation of scientific and technological resources sharing and the construction of scientific data resources sharing system. Some researchers have studied the data authorization and data publishing system framework in the practice of scientific data open sharing.

Publishing Houses

Publishing houses are another very important stakeholder in the open sharing of scientific data, but they have little research in the field of intellectual property of scientific data, accounting for only 2% of the research papers. Scientific data publishing is one of the important ways to promote scientific data sharing. At present, there are three types of data publishing: independent data publishing, data publishing as an attachment to the paper, and data paper publishing. China has two journals publishing scientific data papers, namely, *China Scientific Data* and *Journal of Global Change Data & Discovery*. Some publishing houses have proposed the integrated publishing mode of enhancing publishing. Scientific data is a comprehensive and interdisciplinary subject. Traditional academic paper publishers are limited in data storage, management, data integration, and secondary processing. Data publishing involves five basic links: data submission, peer review, data release and permanent storage, data citation, and impact evaluation. Therefore, it needs to be jointly carried out by publishers, research institutions, data centers, and other institutions, protect the intellectual property of scientific data, and promote value-added services of data.

Legal Institutions

The number of first authors in the field of law was 14, accounting for 6%. From the perspective of the breadth of the research, only one paper is a study of foreign policies, the relevant intellectual property rules and open access policies of the EU “Horizon 2020 Plan” (Han, 2015). From the perspective of the research content of the paper, it mainly involves the research on the legislation and guarantee mechanism of government data, government information sharing, and scientific and technological resources. Although scientific data and government data have something in common, they are different from each other in terms of professionalism. From the perspective of the depth of the research, there is little content related to the specific links of scientific data open sharing. Researchers in the field of law have theoretically demonstrated the sharing of scientific data and the proprietary nature of intellectual property, and proposed to establish a government guidance mechanism, improve the sharing legal system, clarify the definition of property rights, and standardize the property rights trading mechanism. Other researchers take the intellectual property which has a cross relationship with the scientific data property right in the current law as the starting point, and put forward the strategic choice of data property right, data sovereignty, and data sharing. The third author of Institute of Law, Chinese Academy of Social Sciences participated in the research on the subject of intellectual property of scientific data, and studied the legal system, policies and legal norms of scientific data sharing from a macro perspective. This is one of the few research institutions and researchers in the field of law to carry out interdisciplinary fusion research.

Conclusion

At present, there is an obvious imbalance in the subject research on the intellectual property of scientific data. First, the research topic is unbalanced. The scope of intellectual property includes copyright and neighboring rights, patent rights, trademark rights, trade secret rights, integrated circuit layout design rights, etc. At present, most of the research on the subject of intellectual property of scientific data only involves copyright, while the research on patent rights and other aspects is rarely involved. Second, the research subject is unbalanced. Library

and information institutions are the absolute main force in the subject research of scientific data intellectual property, while other research institutions mainly carry out research in different disciplines. The research results of data centers and publishing houses are not many. Third, the research method is single. It mainly focuses on individual combat and cooperation with researchers from the same institution and the same field, and really carries out interdisciplinary research, especially rarely with the legal institutions. Fourth, the research depth is not enough. The research on intellectual property of scientific data involves both professional technical knowledge and intellectual property knowledge. Only through the in-depth cooperation of researchers from different disciplines and different research fields can we promote the opening and sharing of scientific data in China from the top to the bottom.

From the perspective of macro policy environment, different stakeholders have paid unprecedented attention to the intellectual property of scientific data. In December 2021, the Standing Committee of the National People's Congress passed the newly revised *Law of the People's Republic of China on the Progress of Science and Technology*, clearly proposing to “promote the reform of intellectual property ownership and rights distribution mechanism” and “establish a science and technology data information system and resource database”. In September, 2021, the General Office of the State Council of the CPC Central Committee issued the *Outline for Building a Strong Intellectual Property Country 2021-2035*, which pointed out that “we should establish data intellectual property protection rules, improve open source intellectual property and legal system” and “establish data standards, resource integration and efficient information service mode”. In June 2021, China promulgated *The Data Security Law*. In April, 2018, the General Office of the State Council issued *Scientific Data Management Measures*, which is the highest administrative regulation at the national level in the management of scientific data. According to the sharing concept of “openness as the norm and non-openness as the exception”, it defined the policy guidance and action program.

The current positive macro policy environment in China has brought opportunities for the work of intellectual property of scientific data. Research institutions, data centers, publishing houses, and legal institutions should work together to carry out cross-border integration research and jointly promote the establishment of scientific data intellectual property laws and regulations at the macro level by the country; as the meso level, government funded institutions and competent departments of various industries should establish shared policies and action guidelines in various professional fields; research institutions involved in the life cycle of scientific data, as the micro level, should actively develop operable incentive mechanisms and assessment methods such as post setting, performance income, and professional title evaluation. Through the establishment of laws, regulations, policies, and system guarantee systems at the macro, meso, and micro levels, we aim to clear obstacles for achieving open sharing of scientific data in the future. It is reasonable to believe that the cooperative research between different stakeholders and the legal institutions will actively promote the legislation of open and shared intellectual property of scientific data in China, and the library and information research institutions will further highlight their core position in the future.

References

- General Office of the State Council. (2018-03-17). Circular of the General Office of the State Council on printing and distributing notice on measures for scientific data management. Retrieved from https://www.gov.cn/gongbao/content/2018/content_5283177.htm
- Han, Y. (2015). Research on intellectual property rules and open access policies related to the EU horizon 2020 plan. *Intellectual Property*, 29(3), 92-96. Retrieved from <https://x.cnki.net/kcms/detail/detail.aspx?FileName=ZSCQ201503015&DbName=CJFQ2015>

- Institute of Economics, Chinese Academy of Social Sciences. (Ed.). (2005). *Dictionary of modern economics*. Nanjing: Phoenix Publishing House; Jiangsu People's Publishing House.
- Ma, D. H., Zhang, H., & Gao, C. N. (2007). Issues on legislation on share of scientific resources. *Journal of Beijing Institute of Technology (Social Sciences Edition)*, 9(2), 8-13. Retrieved from <https://kns.cnki.net/kcms/detail/detail.aspx?FileName=BLDS200702001&DbName=CJFQ2007>
- Sheng, X. P., & Wu, H. (2019b). An analysis of different stakeholders' motivations in open sharing of scientific data. *Library and Information Service*, 63(17), 40-50. Retrieved from <https://kns.cnki.net/kcms/detail/detail.aspx?dbname=DKFX2019&filename=TSQB20190904005&dbcode=DKFX>
- Sheng, X. P., & Wu, T. (2019a). Review on open sharing of scientific data across the world. *Library and Information Service*, 63(17), 6-14. Retrieved from <https://x.cnki.net/kcms/detail/detail.aspx?FileName=TSQB201917003&DbName=DKFX2019>
- Sun, H. L., & Liu, C. (2003). Research on the development of international science and technology data frontier. *China Basic Science*, 5(1), 15-20. Retrieved from <https://kns.cnki.net/kcms/detail/detail.aspx?FileName=ZGJB200301003&DbName=CJFQ2003>
- Sun, J. L. (2003). Scientific data resources and sharing. *China Basic Science*, 5(1), 32-35. Retrieved from <https://kns.cnki.net/kcms/detail/detail.aspx?dbname=CJFD2003&filename=ZGJB200301006&dbcode=CJFD>
- Zhou, W. N., Liu, Y., & Wang, G. B. (2023). Policy analysis of domestic and foreign scientific data management and sharing and its enlightenment to National Natural Science Foundation of China. *Bulletin of National Natural Science Foundation of China*, 37(1), 150-160. Retrieved from <https://kns.cnki.net/kcms/detail/detail.aspx?dbname=CJFD2023&filename=ZKJJ202301026&dbcode=CJFD>