Research on the Characteristics of Scientific Researchers' Demand for Scientific Data and Its Implications for the Construction and Service of Literature Resources

——Taking the customers of the National Science and Technology Library as an example[1]

LIU YuWei[1], ZHOU Qin*[1], YANG ChenLu[2]

[1]Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing 100005, P. R. China

[2] Institute of Basic Medical Sciences,Chinese Academy of Medical Sciences, Beijing 100005, P. R. China

zhou.qin@imicams.ac.cn

Abstract: The rapid development of new artificial intelligence and data processing technologies has made scientific data an indispensable and crucial resource in scientific research. This study conducted a survey and analysis of scientific data needs among users of the National Science and Technology Library (hereinafter referred to as "NSTL"), aiming to explore the demands and usage characteristics of scientific data resources in different research fields. The goal is to provide insights for the development and service of new literature resources. Through questionnaire surveys and interviews, researchers' needs for scientific data resources, their usage patterns, acquisition methods, difficulties in obtaining and utilizing scientific data, and the application scenarios of scientific data were investigated. A total of 477 questionnaires were collected, and statistical analysis methods were employed to summarize the characteristics of researchers' needs and utilization of scientific data, including: (1) Overall high demand, with variations in demand intensity across different disciplines; (2) Predominantly free sources, with variations in source distribution between different disciplines; (3) Widespread application scenarios for scientific data; (4) Lack of effective platforms for scientific data sharing and information channels. Based on these characteristics, implications for literature resource development and services are proposed: (1) The significant

1 This study is founded by CAMS Innovation Fund for Medical Sciences (CIFMS) 2021-I2M-1-033.

C. Bai et al. (eds.), *Proceedings of 2023 China Science and Technology Information Resource Management and Service Annual Conference (COINFO 2023)*, Advances in Economics, Business and Management Research 293,
https://doi.org/10.2991/978-94-6463-498-3_15

demand for scientific data by researchers highlights the importance of developing and providing services for scientific data resources; (2) The distinct disciplinary differences in the characteristics of scientific data demand necessitate personalized resource development and services; (3) Some researchers lack sufficient awareness of scientific data, emphasizing the necessity for relevant promotional and training services.

Keywords: Scientific Data; Needs and Utilization; Literature Resource Development and Services.

In the era of big data, data-intensive research has emerged as a new research paradigm, with the research value inherent in data continuously being explored. Data plays an increasingly crucial role in various stages of the research process. Scientific data refers to original and fundamental data produced in scientific and technological activities, as well as datasets formed through processing and organization1. Scientific data not only forms the foundation of scientific research but also serves as a powerful driver for advancing scientific inquiry. As researchers shift their attention from traditional sources such as journals and books to information resources, scientific data resources have gradually become a new focus in resource development efforts. Existing literature survey results indicate that researchers universally recognize the significant role of scientific data, making it an indispensable resource in their research processes. Understanding how to develop and provide services for scientific data resources based on the characteristics of researchers' needs and usage has become a crucial topic in the advanced construction and services of scientific and technological literature resources in the era of data.

In order to analyze the characteristics of researchers' demands for scientific data resources and their current usage, and to provide references for the development of advanced construction and services of scientific and technological literature resources, this study conducts a survey and analysis of scientific data needs among customers of various member units of the NSTL. The aim is to explore the demands and usage characteristics of scientific data resources among researchers in different research fields.

1. Current Research Status on the Demand and Utilization of Scientific Data Resources

Research conducted by Ma and colleagues2 investigated the demand for non-literature resources, such as scientific data and software tools, among frontline researchers at the Chinese Academy of Sciences. The survey employed in-depth face-to-face interviews to explore the acquisition of various non-literature resources. The results revealed that, concerning information acquisition channels, the most crucial source for survey participants was "library-subscribed resources," accounting for over 40%. The next significant source was "search engines," constituting approximately 30%. Other frequently used channels included scientific data portals, peer research institution websites, and more. In terms of the objectives of information acquisition, 80% of respondents aimed to "search for professional scientific datasets, find solutions to temporary general issues, and browse information from professional associations and peer websites." Additionally, 63% of participants aimed to "attend conferences, publish papers, search for professional software tools for writing papers or analyzing data, and understand project funding and application information." Furthermore, different academic disciplines exhibited significant variations in the purposes of acquiring various non-literature information. Regarding the demand for specific non-literature resources, scientific data showed the highest demand, with 96% of respondents frequently using it in their research and learning activities. Following closely is the demand for software tools, with 62% of participants frequently using them. The most common channels for acquiring software tools are from peers or purchasing from professional companies. Some researchers also mentioned obtaining information about software tools from literature. Difficulties and problems encountered in software acquisition include: 1) limitations in free public software trials and the inability to upgrade promptly; and 2) challenges and high costs associated with obtaining professional software tools.

Guo Weineng3, in his paper analyzing the survey of datasets from members of the American Academic Library Alliance, pointed out that researchers primarily access datasets through four main avenues: 1) proprietary data generated by the school in the course of work, teaching, and scientific research; 2) usersubmitted data, where library users personally or research teams upload and share data; 3) third-party organization-authorized data, referring to non-governmental third-party organizational data beyond schools and users, which can be categorized into authorized data under a payment mechanism and openly shared data; 4) government statistical data and administrative source data.

Wang Xue et al. 4, through bibliometric methods, conducted a study on the citation behavior and impact of scientific data in the field of bioinformatics literature. The results indicated a high reliance on datasets in the field of bioinformatics, with 61.5% of Chinese literature exhibiting citation behavior towards scientific data. In terms of citation trends, since 1998, the frequency of Chinese literature citing datasets has increased annually, with a growing trend. Additionally, the paper demonstrated a significant correlation between dataset quality and paper quality.

Based on existing research results, it can be observed that the demand for non-literature resources, particularly scientific data, has become a common trend among domestic researchers. The specific degree and characteristics of the demand for scientific data vary across different disciplines, indicating potential differences in the level of demand and acquisition purposes for scientific data resources among researchers in various academic fields.

2. Survey Design

The main objective of the survey is to understand the current utilization status and characteristics of scientific data resources among customers from various member units of NSTL. Additionally, the survey aims to identify commonalities and differences in the demand characteristics for scientific data resources across different research fields.

2.1 Survey Content

Investigation of the Current Demand for Scientific Data:

Evaluation of the importance of scientific data by users of NSTL.

Identification of urgently needed scientific data.

Investigation of the Current Utilization of Scientific Data:

Names of scientific data used by users of NSTL.

Channels for obtaining scientific data.

Difficulties and problems encountered during the process of obtaining and utilizing scientific data.

Instances of citing scientific data in research outcomes.

2.2 Survey Methods

Based on the aforementioned survey content, both survey questionnaires and interview outlines were designed for user demand research:

Questionnaire Survey:

Online survey questionnaires were distributed to customers of various member units of NSTL. The member units include nine Librarys or organizations: National Science Library,Chinese Academy of Sciences(hereinafter referred to as"NSLC"), China National Chemical Information Center Co., Ltd(hereinafter referred to as"CNCIC") , Institute of Medical Information/Medical Library,Chinese Academy of Medical Sciences(hereinafter referred to as"IMICAMS"), National Agricultural Library(hereinafter referred to as"NAL"), China Metallurgical Information and Standardization Institute(hereinafter referred to as "CMISI"), National Institute of Metrology,China(hereinafter referred to as"NIM"), Institute of Scientific and Technical Information of China (hereinafter referred to as"ISTIC"), Machinery Industry Information Institute，(hereinafter referred to as"MIII") , and China Institute of Standardization(hereinafter referred to as"CNIS") . The selection of terminal research subjects within each member unit was conducted separately using a random sampling method.

Interviews:

One-on-one interviews were conducted with representative individual users to gain in-depth insights into the utilization scenarios and demand characteristics of researchers regarding scientific data.

3. Analysis of Survey Results

3.1 Basic Survey Information

Questionnaire Source and Quantity:

A total of 477 questionnaires were collected from member units of NSTL. Detailed information on the questionnaire sources is provided in Figure 1.

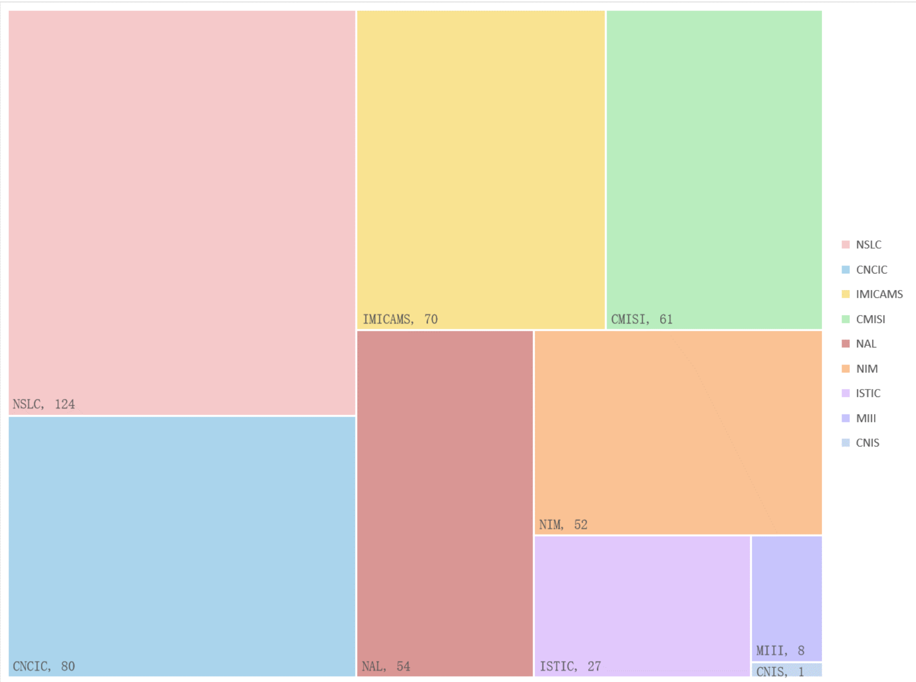

Fig.1. Source of Questionnaire

Fields of the Respondents

In this survey, the respondents were engaged in various fields, including engineering, medicine, science, agriculture, and management. Among them, 155 respondents were in the field of engineering, representing the highest proportion at 32%, followed by 122 respondents in the medical field, accounting for 26%. Additionally, there were 80 respondents in the field of science, comprising 17%, and 59 respondents in agriculture, making up 12%. Furthermore, 39 respondents, or 8%, were engaged in the field of management. The "other" category in the classification includes economics, literature, history, philosophy, law, and more. For detailed information, refer to Figure 2.

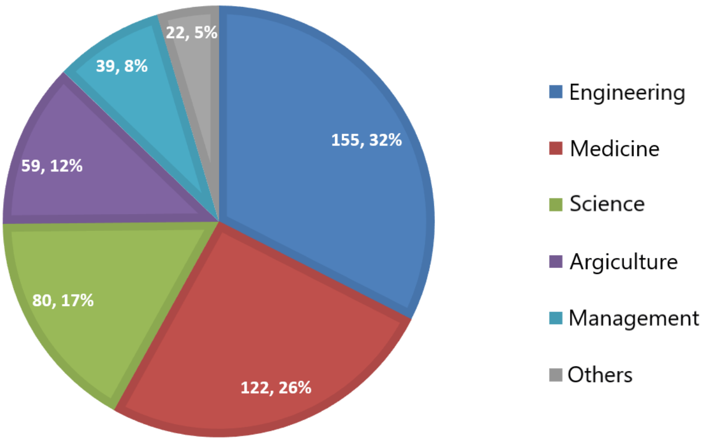

Fig.2. Composition of Respondents' Fields of Expertise

Educational Background and Professional Status of Respondents

Among all respondents, nearly half held a master's degree (255, 47%), while over one-third had a doctoral degree or higher (163, 34%). Additionally, 56 respondents held a bachelor's degree.

The majority of respondents were either students or researchers, with students accounting for over 40% (209, 44%) and researchers representing nearly 40% (188, 39%). There were also 38 research managers, 7 clinical healthcare professionals, and 35 individuals in other professions.

More than half of the respondents held junior professional titles or equivalent positions (251, 53%), while 101 respondents held intermediate professional titles (21%), and 125 respondents held senior professional titles or higher (25%).

Table 1. Composition of Respondents' Educational Background and Professional Status

Characteristic	n	Proportion
Education		
Doctorate and above	163	34%
Master's	225	47%
Bachelor's	56	12%
Other	33	7%

Occupation		
Researchers	188	39%
Science and Technology Managers	38	8%
Clinical Healthcare Professionals	7	1%
Students	209	44%
Other	35	7%
Professional Title		
Junior and other	251	53%
Intermediate	101	21%
Senior and above	125	26%

3.2 Analysis of Researcher's Demand and Usage Characteristics of Scientific Data

Through statistical analysis of the data collected from the questionnaires and the organization of user interview results, the following characteristics of researchers' demand for and utilization of scientific data were summarized.

Overall High Demand with Variations in Demand Intensity Across Disciplines

Among all surveyed respondents across various fields, over 60% (292) expressed that scientific data is highly important for their research work. Only 16 respondents (3%) considered scientific data to be not important for their research work. The detailed distribution can be found in Figure 3. Additionally, 236 respondents (49%) mentioned citing scientific data in their past research work. The frequently cited data included medical databases, economic data, demographic data, meteorological data, and more.

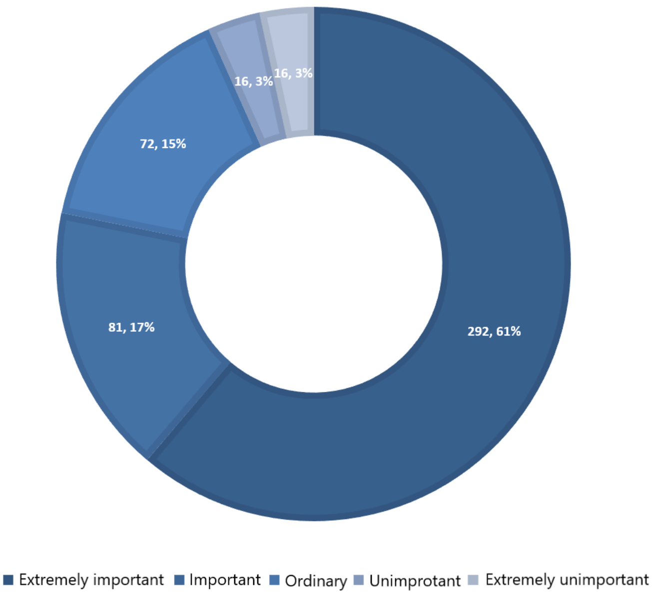

Fig.3. Distribution Chart of Respondents' Perception of the Importance of Scientific Data

The importance of scientific data for research showed significant differences between different disciplines (P=0.012), with more than 60% of respondents in the fields of agriculture, engineering, and medicine believing that scientific data are very important, only 40% in management, and only 20% in other fields (economics, literature, history, philosophy, law, etc.). There was also a significant difference in the citation of previous scientific data across different fields (P<0.001), with the percentage of previous citations in medicine significantly smaller than in other fields, possibly related to the fact that research in medicine relies more on clinical and community-collected data (Table 2).

Table 2. Importance of scientific data and previous citations in different fields of research

	Engineering	Medicine	Science	Agriculture	Management	Others	P-value
Importance of scientific data for research							**0.012**
very important	99 (64%)	78 (64%)	47 (59%)	46 (78%)	17 (44%)	5 (23%)	
more important	23 (15%)	20 (16%)	14 (17%)	6 (10%)	11 (28%)	7 (32%)	
general	20 (13%)	18 (15%)	14 (17%)	5 (9%)	9 (23%)	6 (27%)	
unimportant	6 (4%)	3 (2%)	2(3%)	0 (0%)	2 (5%)	3 (14%)	
Very unimportant.	7 (4%)	3 (2%)	3(4%)	2(3%)	0 (0%)	1(4%)	
Has scientific data been cited in previous research							**<0.001**
quote	88 (57%)	37 (30%)	45 (56%)	36 (61%)	19 (49%)	11 (50%)	
Uncited	67 (43%)	85 (70%)	35 (44%)	23 (39%)	20 (51%)	11 (50%)	

Scientific data frequently acquired is free of charge, with differences in the distribution of sources across disciplines

The four main ways of acquiring scientific data are open access, purchase, cooperative acquisition, and organizational self-construction. Statistics on the ways of obtaining scientific data of 477 respondents (Table 3) showed that more than half of the respondents had used open and free databases (246, 52%), a quarter of the respondents had used paid-for databases (117, 24%), and 10% of the respondents had used cooperatively-acquired (59, 12%) and self-constructed by their own organization (64, 13%) databases. The cross-tabulation results show that there are differences in the access to scientific data by people in different research fields.

Table 3. Sources of scientific data mentioned by respondents from different research fields

	Open Access	Purchase	Cooperative Acquisition	Organizational Self-Construction
NSLC	66 (53%)	25 (20%)	7 (6%)	13 (10%)
CNCIC	27 (34%)	19 (24%)	8 (10%)	8 (10%)
IMICAMS	47 (67%)	10 (14%)	8 (11%)	-
CMISI	29 (47%)	21 (34%)	7 (11%)	12 (20%)
NAL	28 (52%)	12 (22%)	7 (13%)	6 (11%)
NIM	29 (56%)	19 (37%)	13 (25%)	15 (29%)
ISTIC	16 (59%)	9 (33%)	8 (30%)	10 (37%)
MIII	4 (50%)	2 (25%)	1 (12%)	-
CNIS	*	*	*	*
Total	246 (52%)	117 (24%)	59 (12%)	64 (13%)

Note: - indicates that the source is not mentioned

*Indicates that there are too few data to count

Open access scientific data

Open access scientific data sources which are freely available to the respondents are shown in Table 4 (statistical frequency > 1 only), and among all the open access scientific data, TCGA was mentioned most frequently, followed by GEO, NCBI, NIST, Genbank, and CNKI.

Table 4. Open Access scientific data sources mentioned by the respondents

Respondents' group	Scientific database	Frequent (frequency)
All respondents	TCGA	36 (7.5%)
	GEO	34 (7.1%)
	NCBI	26 (5.4%)
	NIST	8 (1.7%)
	Genbank	7 (1.5%)
	CKNI	6 (1.3%)
	Pdb protein structure database	4 (0.8%)
	PubMed	4 (0.8%)
	World steel	4 (0.8%)
	(China) National Bureau of Statistics (NBS)	4 (0.8%)
	Literature data	4 (0.8%)
	sci-hub	3 (0.6%)
	AlphaFold database	2 (0.4%)
	ENA	2 (0.4%)
	FAO	2 (0.4%)
	Global biodiversity	2 (0.4%)
	go	2 (0.4%)
	NASA	2 (0.4%)
	Swiss-Prot	2 (0.4%)
	TCGO	2 (0.4%)
	Uniprot	2 (0.4%)
	web of science	2 (0.4%)
	Molecular Information for Aladdin Maclean	2 (0.4%)
	Scientific and technical report data	2 (0.4%)
	USGS	2 (0.4%)
NSLC	TCGA	9 (7.2%)
	NCBI	5 (4.0%)
	Genbank	4 (3.2%)
	Pdb protein structure database	4 (3.2%)
	Literature data	4 (3.2%)
	(China) National Bureau of Statistics (NBS)	4 (3.2%)
	NIST	3 (2.4%)
	GEO	2 (1.6%)
	go	2 (1.6%)
	NASA	2 (1.6%)
CNCIC	CNKI	4 (5.0%)
	web of science	2 (2.5%)
	PubMed	2 (2.5%)

	Molecular Information for Aladdin Maclean	2 (2.5%)
IMICAMS	GEO	17 (24.3%)
	TCGA	15 (21.4%)
	NCBI	4 (5.7%)
	Genbank	3 (4.2%)
	sci-hub	3 (4.2%)
CMISI	TCGA	5 (8.2%)
	World steel	4 (6.6%)
	GEO	3 (4.9%)
	TCGO	2 (3.3%)
	USGS	2 (3.3%)
	CNKI	2 (3.3%)
NAL	NCBI	13 (24.1%)
	FAO	2 (3.7%)
	ENA	2 (3.7%)
	Global biodiversity	2 (3.7%)
	Uniprot	2 (3.7%)
	PubMed	2 (3.7%)
NIM	GEO	6 (11.6%)
	NIST	5 (9.6%)
	NCBI	4 (7.7%)
	TCGA	4 (7.7%)
	AlphaFold database	2 (3.8%)
	Swiss-Prot	2 (3.8%)
ISTIC	GEO	6 (22.2%)
	TCGA	3 (11.1%)
	Scientific and technical report data	2 (7.4%)

Most of the open access scientific data sources mentioned by the respondents in the survey are established by government agencies or scientific research institutions for collecting, storing and distributing scientific data or datasets generated by large-scale scientific research projects, and most of the data are rigorously organized according to different data units or hierarchies, and some of them may be processed and sorted out by official administrators. Most of the basic data in open scientific datasets are uploaded by researchers around the world and made available for free download by others, with high data quality and easy access, which is the main source of authoritative scientific data for researchers at present.

Paid scientific data

Paid scientific data mentioned by respondents from different groups is shown in Table 5 (only statistical frequency >1), among all respondents, the most used paid database is CKNI, followed by Wanfang, TAIR, web of science and GEO.

Table 5. Paid scientific data mentioned by the respondents

Respondents' Group	Scientific Database	Frequency (Frequency)

All respondents	CNKI	21 (4.4%)
	Wan Fang (1916-), PRC politician	13 (2.7%)
	TAIR	8 (1.7%)
	web of science	6 (1.3%)
	GEO	3 (0.6%)
	ACS	2 (0.4%)
	biobank	2 (0.4%)
	Elsevier	2 (0.4%)
	ICSD Crystal Information Platform	2 (0.4%)
	Nature	2 (0.4%)
	PubMed	2 (0.4%)
	scifinder	2 (0.4%)
	Wipro	2 (0.4%)
NSLC	CNKI	7 (5.6%)
	GEO	3 (2.4%)
	ACS	2 (1.6%)
	Nature	2 (1.6%)
	PubMed	2 (1.6%)
	Wan Fang (1916-), PRC politician	2 (1.6%)
CNCIC	Wan Fang (1916-), PRC politician	4 (5.0%)
	CNKI	3 (3.7%)
	web of science	3 (3.7%)
	biobank	2 (2.5%)
	ICSD Crystal Information Platform	2 (2.5%)
	scifinder	2 (2.5%)
IMICAMS	-	-
CMISI	CNKI	6 (9.8%)
	Wan Fang (1916-), PRC politician	5 (8.2%)
	Elsevier	2 (3.3%)
	Wipro	2 (3.3%)
NAL	TAIR	8 (14.8%)
NIM	CNKI	3 (5.8%)
	web of science	3 (5.8%)
	Wan Fang (1916-), PRC politician	2 (3.8%)
ISTIC	CNKI	2 (7.4%)

-Indicates that there is no data with a frequency > 1 appearing here

The paid scientific database mainly come from various of global information consulting companies, such as Clarivate, which provides specialized data and information analysis solutions, and Elsevier, whose database products are relevant. Based on the data from various sources, commercial companies manually screen, cite and organize the data to form professional database products or services in a certain subject area. The scientific data obtained through the paid route are often of high quality and can provide professional online search, analysis and other functions, so this route can be used as an effective supplement to free and open scientific data sources.

Collaboratively acquired scientific data

The collaboratively acquired data mentioned by the respondents are shown in Table 6 (statistical frequency > 1 only), and among that mentioned by all respondents, the main collaborative data source was GEO, followed by data sharing between laboratories and Knowledge Network .

Table 6. Collaborative scientific data sources mentioned by respondents

Respondents' Group	Scientific Database	Frequent (Frequency)
All respondents	GEO	5 (1.0%)
	Data sharing between laboratories	3 (0.6%)
	CNKI	2 (0.4%)
NSLC	GEO	2 (1.6%)
CNCIC	-	-
IMICAMS	GEO	3 (4.2%)
CMISI	-	-
NAL	-	-
NIM	Data sharing between laboratories	3 (5.8%)
	CNKI	2 (3.8%)
ISTIC	-	-

--Indicates that there is no data with a frequency > 1 appearing here

Currently, the data obtained through cooperation mainly comes from public data platforms and data sharing within laboratories, which is a major form of free data acquisition. The mode of collaborative data acquisition is that an individual registers for a scientific data platform account, and after successfully uploading his/her own experimental data, the account will be authorized to download the data from the platform. Among the public data platforms, the Gene Expression Omnibus (GEO) database structured by the National Center for Biotechnology Information (NCBI) of the United States provides high-throughput gene expression data with the widest scope of use, and the GEO database can be traced back to specific studies, mostly based on group data, with high data quality. In addition, the China Knowledge Network (CNN) has summarized some statistical data, mostly based on population surveys, with authoritative data sources. Intra-laboratory shared data, on the other hand, are mainly provided by staff within the laboratory, mostly based on individual data, and data quality cannot be assessed.

Self-constructed scientific data of the respondents' organizations

Scientific database of their own institutions mentioned by respondents is shown in Table 7 (only statistical frequency >1). The self-built scientific data are mainly semi-public data constructed by official organizations within different industries based on the characteristics of industry needs, and the data are basically summarized through the results of previous studies, and the quality of authoritative data is relatively high.

Table 7. Scientific database of their own institutions mentioned by respondents

Respondents' Group	Scientific Database	Frequent (Frequency)
All respondents	China Academy of Measurement Sciences Document Library	5 (1.0%)

	Metallurgical Information Network	4 (0.8%)
	GEO	2 (0.4%)
	Ionic Liquid Database	2 (0.4%)
	Metallurgical industry cost data	2 (0.4%)
	eRice	2 (0.4%)
	Library data	2 (0.4%)
	Fire Explosives Database	2 (0.4%)
	Research Management Data	2 (0.4%)
NSLC	GEO	2 (1.6%)
CNCIC	Ionic Liquid Database	2 (2.5%)
IMICAMS	-	-
CMISI	Metallurgical Information Network	4 (6.6%)
	Metallurgical industry cost data	2 (3.3%)
NAL	eRice	2 (3.7%)
NIM	China Academy of Measurement Sciences Document Library	5 (9.6%)
	Library data	2 (3.8%)
ISTIC	Fire Explosives Database	2 (7.4%)
	Research Management Data	2 (7.4%)

-Indicates that there is no data with a frequency > 1 appearing here

Application scenarios for scientific data

The results of the interviews show that the application scenarios of scientific data are distributed in all aspects of scientific research activities for the professional field, including experimental design and validation, paper writing, knowledge discovery, etc. The following table lists the scientific data application scenarios mentioned in the interviews.

Table 8. Listing of scientific data application scenarios mentioned in the interviews

Scientific DataType	Application Scenario	Applicable Field
Protein data and gene data	Machine Learning of Sequence-Structure-Function-Interaction Mapping and Essential Laws of Proteins; Analysis of Viral Phylogeny	microbiology
Route data	Research on disease transmission patterns	microbiology
Geographical and meteorological data	Study of virus propagation patterns	microbiology
Infrared spectral data, emission source inventory data	Experimental validation and paper writing	Environmental Optics
Experimental data, simulation data	Doing experiments, writing papers, doing designs, building system platforms, etc.	Computing and its cross-disciplines

Basic data on nuclear physics	Reference value reproduction and precision measurements	Electricity radiation metering

Difficulties and problems in accessing and using scientific data

After using R (version 3.6.3, https://www.r-project.org/) for natural language processing of the open questions, developing a dictionary suitable for this questionnaire, and using semi-supervised machine learning models to delineate words, count word frequencies and merge synonyms and near-synonyms, it was found that the biggest difficulty respondents mentioned in acquiring and using scientific datawas data access difficulties (107 mentions), followed by concerns about data quality (58 mentions) and funding difficulties (46 mentions). The details are shown in Figure 4, in which the difficulties in data acquisition are reflected in the difficulties in finding access channels and obtaining raw data; the worries about data quality are reflected in the worries about the authenticity, authority, timeliness, completeness and standardization of data; the financial difficulties are reflected in the high price of applying for databases and insufficient funding for the project; the difficulties in data processing are reflected in the integration of data, data cleansing, and data analysis; and the operational difficulties are reflected in the complexity of the application process and the large amount of work. Data processing difficulties are reflected in data integration, data cleaning and data analysis; operational difficulties are reflected in the complexity of the application process, workload, etc.; copyright difficulties are reflected in the lack of open-source data and the limited scope of sharing.

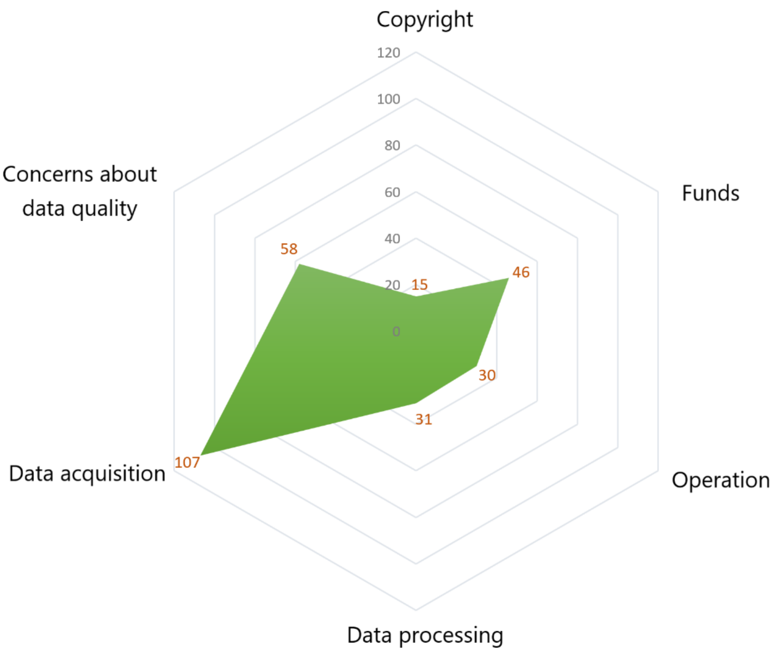

Figure 4. Difficulties in access to and use of scientific data

Concrete difficulties encountered in the process of acquiring scientific data mentioned in the interviews include: (1) the lack of knowledge about relevant information platforms and how to find scientific data in their own disciplines; (2)the lack of sharing platforms for scientific data in China. Concrete problems in the use of scientific data include: (1) the lack of data standards, and data characterization methods are not uniform; (2) the difficulty to ensure the comprehensiveness, authenticity, originality, uniqueness, security and standardization of data, and there is a lack of disclosure of negative data.

4 Research Conclusions and Implications for Literature Resource Construction and Service

4.1 Scientific data demand of researchers is remarkable, and it is of great significance to carry out the construction and service of scientific data resources.

The results of the user survey show that scientific data has become one of the most important resources in the scientific research process, and the importance of scientific data for scientific research has become a consensus for researchers in all disciplines, although there are slight differences in the degree of importance. For open scientific data, researchers generally need channels and platforms to access relevant information openly; for paid-for scientific data, researchers generally need financial support and shared use within the organization.

In the interviews, most of the respondents expressed their greatest concern about the quality of scientific data, including data accuracy, comprehensiveness, authenticity, uniqueness, traceability, security, and format standardization.

The results of the above research show that it is of great significance to carry out the construction and service of scientific data resources, with emphasis on establishing a selection and evaluation system for scientific data resources, exploring an effective guarantee mechanism for joint construction and sharing, establishing a platform for scientific data sharing and an information dissemination channel, and carrying out the management of and research on the rights and interests of scientific data, among other things.

4.2 The demand for scientific data is characterized by obvious disciplinary differences, requiring personalized resource construction and services.

The demand for scientific data has strong specialized characteristics, and the scientific data required by each specialized field depends entirely on the direction of research. Therefore, in the process of organizing and revealing such resources and providing services, the key issue to be considered is how to organize and provide personalized services according to the characteristics of the resources and the usage habits of each specialized field.

4.3 Some researchers have insufficient knowledge of scientific data, and publicity and training services are necessary

The results of the questionnaire show that although the questionnaire contains explanatory text about scientific data, a considerable part of the survey respondents still have insufficient knowledge or cognitive bias about the definition and scope of scientific data, and confuse scientific data with traditional literature resources. This also inspires us to focus on the publicity and training of users in the process of new literature resources service, to improve the data literacy of researchers and help them better utilize scientific data as an important scientific research resource.

Reference

1.Yang Congke. Research on the construction of data resources for agricultural science in China[D]. Chinese Academy of Agricultural Sciences, 2007.DOI:10.7666/d.Y1057337.

2. Ma Jianling, Wang Nan, Zhang Yanmin, et al. A study of research users' demand for non-documentary resources_Taking research users and graduate students of the Chinese Academy of

Sciences as an example[J]. Intelligence Theory and Practice, 2011, 34(2):5.
3. Weining Guo. A survey and analysis of the American consortium of academic libraries TRLN member research dataset[J]. Library and Intelligence Work, 2017, 61(8):7.
4. WANG Xue, MA Shengli, SHE Zengli, et al. Research on citation behavior of scientific data and its influence[J]. Journal of Intelligence, 2016, 35(11):8.