A Review of Data Reuse Research at Home and Abroad

Authors: Lin Xinyue, Yang Jianlin

Date: 2022-09-19T00:00:00+00:00

Abstract

Abstract: [Purpose/Significance] This study systematically reviews the current state of data reuse research both domestically and internationally, summarizes and analyzes its characteristics and deficiencies, and provides references for future data reuse research. [Method/Process] We employed the literature survey method to obtain research literature on data reuse from both domestic and international sources, classified them based on content analysis, summarized the characteristics and deficiencies of current data reuse research, and proposed recommendations for future research. [Results/Conclusions] Existing data reuse research has played a certain role in avoiding duplicate data collection, improving data utilization efficiency, and promoting data sharing among researchers; it has also gradually focused on broader user groups, expanded and extended the disciplines and fields of research, and diversified the types of data reuse research being conducted. However, the overall research direction remains relatively narrow, research methods are relatively limited, and the quantity of research is relatively scarce. In future data reuse research, we should further broaden the user groups studied, pay greater attention to new problems arising in the process of socio-economic development, and focus more on the new requirements for data reuse research posed by the big data era, thereby further promoting more effective and reliable data reuse research to contribute to scientific and technological innovation, social progress, and national development.

Full Text

Review of Research on Data Reuse at Home and Abroad

Lin Xinyue[1],[2], Yang Jianlin[1],[2]

[1]School of Information Management, Nanjing University, Nanjing 210023
[2]Jiangsu Key Laboratory of Data Engineering and Knowledge Service, Nanjing 210023

Abstract

[**Purpose/Significance**] This study systematically reviews the current state of data reuse research both domestically and internationally, summarizes and analyzes its characteristics and shortcomings, and provides references for future research in this area. [**Method/Process**] We employed literature survey methods to collect relevant research on data reuse and applied content analysis to classify these studies. We then summarized the characteristics and limitations of current data reuse research and proposed recommendations for future studies. [**Result/Conclusion**] Existing data reuse research has played a role in avoiding redundant data collection, improving data utilization efficiency, and promoting data sharing among researchers. It has gradually expanded its attention to broader user groups, extended its scope across disciplines and fields, and diversified its research types. However, the overall research direction remains relatively narrow, research methods are somewhat limited, and the volume of research is comparatively small. Future data reuse research should further broaden its user groups, pay greater attention to emerging issues in socioeconomic development, and address new demands posed by the big data era, thereby promoting more effective and reliable data reuse research that contributes to scientific and technological innovation, social progress, and national development.

Keywords: Data Reuse; Data Management; Data Reusability; Reuse Dilemma

---

Scientific data constitutes a crucial foundational strategic resource for national scientific and technological innovation as well as socioeconomic development, and serves as the basic resource for research activities. Most scientific research is conducted based on data collection and analysis. In 1957, to improve the management of scientific and technical data and thereby enhance data utilization to promote scientific development, the International Council for Science (ICSU) successively established the World Data Center (WDC) and the Committee on Data for Science and Technology (CODATA). In recent years, China has continuously emphasized scientific and technological development with substantial resource investment, leading to a significant increase in research activities and researchers' innovative capabilities. Consequently, scientific data has experienced "explosive" growth. Despite considerable efforts in scientific data management and open sharing, numerous deficiencies persist, such as low efficiency and limited scope of scientific data sharing, scattered and even lost data, and failure to maximize data value [1]. This situation results from multiple factors, including the lack of national-level regulatory guarantees, absence of professional data storage systems, researchers' indifference to the potential value of collected and analyzed data, insufficient awareness of preservation, and lack of professional data preservation guidance. Therefore, to improve scientific data utilization, researchers need a deeper understanding of data reuse.

Data reuse, also known as data recycling, secondary data use, or repurposing, currently lacks a unified definition. Zhang Xiaoyue et al. categorized definitions

of research data reuse into four types based on their focus: the "intention"school, which emphasizes defining reuse from the perspective of usage intent beyond the original purpose; the "context" school, which enumerates specific scenarios of data reuse; the "content" school, which elaborates on the specific forms and presentation requirements of reused data, focusing on new research outputs formed after data reuse; and the "process" school, which views data open sharing and reuse as a complete process, emphasizing that the data reuse process must be traceable [2]. Reusing scientific data can effectively avoid redundant data collection, save research project costs, and conserve researchers' time and energy.

Some scholars have previously reviewed data reuse research, but these reviews have primarily focused on solutions to reuse dilemmas rather than providing comprehensive and systematic analysis. Therefore, this paper analyzes domestic and international data reuse research, including scholars' focal topics, user groups, and research methods, reveals emerging characteristics, and evaluates the strengths and weaknesses of current research to provide references for future studies.

2.1 Analysis Sample

We selected CNKI and Web of Science as data sources. For Chinese literature, we used "data reuse," "data recycling," "secondary data use," and "repeated data use" as subject terms, limiting the search to the discipline of "Library and Information Science and Digital Libraries." For English literature, we used the search query: Title=(data reuse) or (data reusing) or (data re-use) or (dataset reuse) or (secondary data reuse) or (data reusability), limiting the search to the research direction of "Information Science Library Science." The search was conducted up to March 14, 2022, yielding 56 Chinese articles and 1,944 English articles. Through manual screening and snowballing—tracking citations and references supplemented by Google Scholar—we obtained a final sample of 98 articles to form the basis of this study. The distribution of literature types is shown in Figure 1 [Figure 1: see original paper]. As illustrated, empirical survey papers account for 57% of the total, representing the dominant publication format in data reuse research.

2.2 Coding Analysis

We employed content analysis to code the sample literature and identify fundamental characteristics of data reuse research, 梳理 ing and summarizing research progress and existing problems. The coding framework included research themes, research objects, research methods, theoretical models used or referenced, and research findings or results, as shown in Table 1 .

Table 1 Literature Coding Framework

Dimension	Description/Coding
Research Theme	Different directions in data reuse research, including: data reuse behavior studies (attitudes, willingness, behavioral characteristics); benefits and drawbacks of data reuse; foundational data reuse research (concepts, definitions, processes, classification); influencing factors of data reuse (external environmental factors like policies and infrastructure, and internal factors related to reusers themselves); data reusability assessment studies (criteria such as accessibility, quality, understandability, trustworthiness); and comprehensive data reuse studies (general discussions that cannot be categorized into specific types)
Research Object	Groups involved in the study sample, such as social science researchers, archaeologists, etc.; studies without differentiation by occupation, age, or identity are collectively referred to as general populations
Research Method	Methods employed in research, divided into data collection methods (e.g., interviews) and data analysis methods (e.g., structural equation modeling)
Theory/Model	Theories or models used or referenced to support research, such as Theory of Planned Behavior (TPB), User Satisfaction Theory
Research Findings/Results	Main research results and findings, as well as innovations presented in the studies

3 Characteristics of Data Reuse Research

As shown in Figure 2 [Figure 2: see original paper], "data reuse behavior studies" account for 37% of the research, "data reuse influencing factors studies" for 22%, "benefits and drawbacks of data reuse studies" for 12%, and "foundational data reuse studies" for 18%. These themes constitute the main areas of data reuse research, representing 89% of the total. Research objects are primarily researchers from major fields including social sciences, STEM disciplines, and biomedicine, with additional attention to education and teachers, food science and nutrition, astrophysics, archaeology, and other disciplines.

Table 2 shows the distribution of data collection and analysis methods used in the 61 research papers. Scholars predominantly employed questionnaires (31.15%), interviews (18.03%), retrieval methods (19.67%), and secondary data

(8.20%) for data collection. Structural equation modeling (18.03%) and statistical methods (9.84%) were commonly used to analyze questionnaire data; interview data were primarily analyzed through coding analysis (16.39%); and retrieval data were typically studied using content analysis and statistical methods.

Table 2 Distribution of Data Collection and Analysis Methods in Research Papers

Data Collection Method	Count (%)	Data Analysis Method	Count (%)
Questionnaire	19 (31.15%)	Structural Equation Modeling	11 (18.03%)
Interview	11 (18.03%)	Coding Analysis	10 (16.39%)
Retrieval	12 (19.67%)	Statistical Methods	6 (9.84%)
Secondary Data (Other methods with <3% each)	5 (8.20%)	Content Analysis (Other methods with <3% each)	4 (6.56%)

To conduct their research, scholars utilized and referenced various theoretical models. The inclusion criteria for "theory/model use and reference" were: detailed introduction of relevant theories/models in the literature, and research design based on these theories/models (e.g., proposing research hypotheses based on a theory). Studies that merely cited theories without basing their research design on them were excluded from our statistics [3]. The terminology for theories and models in Table 3 originates from the analyzed literature.

As shown in Table 3, the most frequently used or referenced theories in data reuse research are the Theory of Planned Behavior (TPB) and the Theory of Reasoned Action (TRA).

Table 3 Theories and Models Used or Referenced in Data Reuse Research and Their Frequency

Theory/Model	Frequency of Use/Reference
Theory of Planned Behavior (TPB)	11
Technology Acceptance Model (TAM)	6
User Satisfaction Theory	4
Continued Use Intention Theory	3

Theory/Model	Frequency of Use/Reference
Walsh and Downe Model, Citation Motivation Theory (Garfield, 1979), Institutional Theory, Incentive Theory/Motivation Theory, Unified Theory of Acceptance and Use of Technology (UTAUT), Item Response Theory, Rasch Model, Ecosystem Model, Data Curation Lifecycle Theory, MOA Theory, LDA Topic Model, Information Use Environment Theory, Expectation Confirmation Theory, Perceived Risk Theory, Community Capability Model, Self-Determination Theory, Technology Adoption and Utilization Integration Theory, Theory of Reasoned Action (TRA), IS Success Model	1 each

In summary, domestic and international data reuse research exhibits the following characteristics:

(1) **Broad attention to user groups.** While researchers constitute the primary group in data reuse studies, other populations that do not extensively reuse data—such as archaeologists, teachers, and cultural center and museum staff—have also received attention. Among researchers' fields, in addition to social sciences, biomedicine, and STEM disciplines with long traditions of data sharing and reuse, scholars have continuously expanded the disciplinary scope to include nutrition, food science, agricultural chemistry, astrophysics, and more. However, despite the growing number of data reuse studies, most focus on data reuse practices within specific disciplines, lacking cross-disciplinary research [4].

(2) **Data reuse behavior studies receive the most extensive attention.**

This area can be further subdivided into more detailed research domains including data reuse behavioral intention and influencing factors, data reuse dilemmas, and data reuse promotion strategies. Domestic data reuse research almost exclusively focuses on behavior studies, while international research covers broader themes, though behavior studies still dominate.

(3) **Overall research volume is relatively small, with domestic research lagging behind international research.** The search results alone show that among data reuse-related literature, domestic papers number 23 while international papers number 75. Moreover, domestic research themes concentrate primarily on behavior studies, with a relatively narrow scope compared to international research, and few studies address definitions, processes, or frameworks of data reuse.

(4) **Data reuse research is closely linked to data sharing.** Data sharing is the prerequisite for data reuse, and data reuse is the purpose of data sharing; thus, the two research areas often intersect. From the perspective of the data lifecycle, data sharing and data reuse occupy different stages —data sharing involves data processing and preservation, which are prerequisites for data retrieval and evaluation in subsequent reuse processes. Multiple studies indicate that data sharing is a key factor influencing data reuse, while data reuse experience significantly affects researchers' perceptions of and norms for data sharing.

(5) **Some data reuse-related issues, despite receiving considerable attention and discussion, have yet to reach consistent conclusions due to their complexity.** For instance, data reusability assessment, as a core stage in the data reuse process, presents a highly complex problem when understood from the perspective of reusers' perceptions. Although numerous scholars have explored data reusability across multiple dimensions—including data itself (e.g., accessibility, quality, repository stability and security), data producers and users (e.g., reliability), and data contextual information (e.g., relevance)—no definitive conclusions have been reached.

4 Research Progress in Different Fields

Due to space limitations, this paper selects several key areas for detailed discussion: data reuse behavior research; research on data reuse concepts, definitions, and frameworks; data reusability assessment and analysis; and studies on the benefits and drawbacks of data reuse.

4.1 Data Reuse Behavior Research

Domestic data reuse research almost exclusively focuses on this area, while a substantial portion of international research also addresses it, making data reuse behavior studies the largest component of data reuse research. This field emphasizes exploring researchers' attitudes, willingness, behavioral practices,

and related influencing factors regarding data reuse. Domestic scholars have studied data reuse behaviors among different user groups and disciplines, primarily including biomedical researchers and social science researchers. International scholars have conducted broader investigations of user groups, mainly covering STEM (Science, Technology, Engineering, Mathematics) researchers, biomedical researchers, and social science researchers. Data reuse behaviors of researchers in other disciplines (astrophysics, zoology, food science, nutrition, agricultural chemistry, education) and among teachers, archaeologists, and health scientists have also received attention. Additionally, some scholars have specifically studied novice data reusers [5] and experienced data reusers [6].

Scholars have extensively discussed data reuse behavioral intention and influencing factors, data reuse dilemmas, and data reuse promotion strategies. Data reuse dilemmas primarily originate from three dimensions: legal and regulatory issues, technical feasibility, and cognitive acceptance. Research indicates that current laws and policies cannot adequately address the challenges posed by rapidly growing data reuse demands. Helena Ursic et al. noted that European intellectual property rights restrict reusers from fully utilizing datasets, and data localization requirements in data retention laws hinder international data transmission, limiting global data exchange and reuse [7]. There are no guidelines for publicly releasing datasets for reuse; for example, whether and how sensitive data can be reused remains unclear [8]. Kathrin Dentler et al. pointed out that although data reuse offers tremendous benefits for patients and society, patient data involves privacy issues requiring protection. The EU Data Protection Directive states that personal data should not be disclosed, provided, or used for purposes other than those specified unless consent is obtained from the data subject or authorized by law—a principle known as the purpose limitation principle. Furthermore, data itself constitutes a major cause of data reuse dilemmas, including factors such as data quality, data access, and data portability. Ayoung Yoon found that some participants abandoned data reuse due to difficulties in data access [9]; James W. McAllister noted that information about data provenance is often missing, making it impossible to ensure data accuracy or understand the data, thus causing difficulties and limitations in reusing historical empirical data.

Behaviors are often products of attitudes, which are in turn influenced by personal experience. Research shows that researchers generally hold positive attitudes toward data reuse but still harbor many concerns. Many researchers are unwilling to share their collected data with others, viewing it as a valuable competitive advantage. Data reuse is considered a mechanism for reducing data collection costs and shortening research processes—a feasible and economical option for researchers with limited time and resources. However, the mismatch between the original data collection intent and secondary use intent often requires additional time for data integration, offsetting the benefits of reuse. Wei Yinzhen et al. found that researchers' primary concerns during data reuse involve potential copyright infringement, insufficient understanding of the data, and obstacles to publishing research results [10]. Some scientists may believe

their data has no value to others simply because they cannot imagine what others might do with it. If data were shared more widely and more scientists recognized the potential for unexpected uses of their data, they would be more likely to share.

Given the potential benefits of data reuse, scholars have studied strategies to promote it. Li Jialu et al. found that data literacy significantly impacts both researchers' data reuse behaviors and data reusability. Higher data literacy enables researchers to perceive greater accessibility of data repositories and scientific data reusability, while researchers with higher data literacy are more likely to engage in actual scientific data reuse behaviors. Therefore, widely implementing data literacy education and improving researchers' data literacy can help promote scientific data reuse [11]. Zhang Xiaoyue et al. noted that optimizing open scientific data infrastructure environments, establishing data policy environments oriented toward balanced interests, providing open scientific data services through research support institutions, and attending to researchers' subjective factors affecting data reuse behaviors all facilitate data reuse promotion [2]. Data sharing presents new opportunities for the academic reward system. When scientists share data, they make important scholarly contributions, yet no widely recognized method currently exists to measure and acknowledge this contribution.

4.2 Data Reuse Concepts, Definitions, Frameworks, and Processes

Data reuse still lacks an official, definitive definition. Broadly, the term often refers to any use after initial utilization. In the narrow sense, data collected by an individual for a specific research project is first used by that individual to address specific research questions. When the dataset is submitted to a repository, retrieved by others, and deployed in another project, it is typically considered reuse.

Some scholars have distinguished types of data reuse. Bart Custers et al. differentiated three categories: (1) data recycling—using data multiple times for the same purpose in the same context; (2) data repurposing—using data for purposes different from the original collection purpose but still within the same context as the original purpose; and (3) data recontextualization—using data in a context different from where it was originally collected [12].

Xiaoguang Wang et al. analyzed the data reuse process. Data reuse is a dynamic process. The initial stage, stimulated by data needs, involves the decision of whether to reuse existing data. The second stage is exploration and collection, during which researchers must discover, acquire, understand, and select needed data from various sources through various means, focusing on data entities and context. If sufficient relevant, valid, and usable data are found, determination is enhanced or abandoned. When a favorable decision is made, researchers transition between discovering and acquiring data, understanding and selecting data, correcting inappropriate data selections, or searching for additional data to

achieve optimal fit before selecting and obtaining final data. After data selection, collected data are matched with research purposes and secondary processing—repurposing—begins. This stage adapts research data, employs multiple data processing operations, and conducts research on secondary processed data [13].

4.3 Data Reusability Assessment and Analysis

Meloda is a metric developed by Alberto Abella et al. to assess the degree of data reusability, first introduced in 2011 [14]. Versions 1.0 to 2.5 of Meloda considered three dimensions: technical standards of datasets, mechanisms for accessing data, and legal licensing of data. Through application in datasets from Spain' s national open data portal, it was noted that a fourth dimension was necessary to consider the data model to be published, reflecting the importance of data structure for processing information (machine processability). The improved Meloda 3 includes four dimensions: technical standards of datasets, mechanisms for accessing data, legal licensing of data, and data models. Meloda 4 added two more dimensions: geographic information and data update frequency. During five years of use, Meloda 4 revealed some limitations. To gain deeper understanding, an international expert panel was surveyed on two aspects: what other factors should be considered to determine the reusability of published datasets, and whether the internal structure (i.e., levels within each dimension of the metric) should be increased, merged, deleted, or divided. The panel ultimately considered two new dimensions—dissemination and reputation —proposed new internal structures, determined levels for each dimension, and produced the latest version, Meloda 5 [14].

Table 4 Meloda 5

Dimension	Scoring Criteria
Legal Licensing (max 6 points)	1. Non-commercial reuse2. Commercial reuse or unrestricted
Information Access (max 6 points)	1. Web access to dataset or unique URL parameter2. Web access with unique parameters for individual data3. API or query language
Technical Standards (max 6 points)	1. Closed standards reusable and open non-reusable2. Open standards reusable3. Open standards with independent metadata
Standardization (max 10 points)	1. Own data model standardization2. Published own special data model standardization3. Local standardization4. Global standardization
Geographic Content (max 6 points)	1. No geographic information2. Simple or complex text fields3. Coordinates or complete geographic information

Dimension	Scoring Criteria
Data Update Frequency (max 15 points)	1. More than 1 month2. Between 1 month and 1 day3. Between 1 day and 1 hour4. Between 1 hour and 1 minute5. Seconds (update cycle less than 1 minute)
Dissemination (max 6 points)	1. Unsystematic communication/dissemination2. Updated available resources (i.e., RSS feeds)3. Proactive dissemination/push communication (automatic and immediate information)
Reputation (max 6 points)	1. No source for reputation data2. Statistics or reports on user opinions3. Indicators or rankings of data source reputation

Jihyun Kim et al. explored how earthquake engineering researchers assess the reusability of colleagues' experimental data [15]. When evaluating data reusability, earthquake engineering researchers primarily consider three questions: (1) Is the data relevant? (2) Is the data understandable? (3) Is the data trustworthy? Assessment approaches mainly include journal articles, personal networks, documentation, and conversations with colleagues who generated the data. They often employ multiple approaches simultaneously, as each offers different capabilities to support their evaluation of colleagues' data reusability.

4.4 Benefits and Drawbacks of Data Reuse

Data reuse can benefit both data producers and data reusers. For data producers, having their data reused can enhance the impact of their corresponding papers. Heather A. Piwowar et al. found that after accounting for other factors affecting citation rates, data reuse still produced strong citation benefits, and the direct impact of third-party data reuse persisted for years after researchers published numerous papers reusing their own data. For data reusers, data reuse can reduce unnecessary duplicate experiments, shorten research cycles, lower research costs, and accelerate research progress.

However, all things have two sides. While data reuse offers many benefits, it inevitably has some drawbacks. Stefan Collini wrote in the preface to *The Slow Professor*: "One of the real obstacles to intellectual productivity in contemporary academia is that most scholars publish too many articles." Although the Chinese Communist Party Central Committee and the State Council issued the "Overall Plan for Deepening the Reform of Education Evaluation in the New Era," emphasizing that "quantitative research indicators such as paper counts, project numbers, and research funding should not be linked to performance-based salary distribution and rewards," the number of publications remains the most important performance indicator in science. Erik M. van Raaij noted that using the same dataset across multiple publications may constitute self-

plagiarism and result in duplicate, redundant, and overlapping publications. A 2007 Wiley Blackwell survey showed that redundant publication was the most serious and common type of the 16 research misconduct behaviors, followed by plagiarism, duplicate submission, and non-disclosure of author conflicts of interest. Data reuse itself is not problematic, but analysis of actual cases using the same data across multiple publications demonstrates that excessive and improper data reuse can lead to research misconduct [16].

5 Conclusion

Research data is not only a direct product of research activities but also a strategic resource supporting national scientific research and technological innovation. Data reuse of research data has attracted high-level attention from international organizations, government departments, and research institutions. Summarizing the past to guide the future is important work. By analyzing 98 data reuse-related research papers, this study has 梳理 ed progress in domestic and international data reuse research and summarized its characteristics. Over the past two decades, data reuse research has played a role in reducing redundant data collection, improving data utilization efficiency, and promoting researcher data sharing. It has gradually expanded its attention to broader user groups, extended its scope across disciplines and fields, and diversified its research types. Multiple scholars have reviewed key areas of data reuse research, including data reuse behavior, influencing factors, and benefits and drawbacks. Simultaneously, this paper has elaborated on current problems and shortcomings, such as relatively narrow research directions, limited research methods, and sparse research quantity.

Therefore, future data reuse research should seek breakthroughs in the following aspects:

(1) **Domestic data reuse research should emphasize greater diversification.** Previous research has focused on relatively homogeneous groups, often concentrating on universities, scholars, and researchers, with research directions primarily limited to behavior studies. To break these limitations, research perspectives should broaden to wider populations and more macroscopic social levels, 挖掘 ing deeper into data reuse research to focus on overall social development and genuinely assist different types of groups, enabling data reuse research to serve society and the public.

(2) **Attach importance to opportunities and challenges brought by big data for data reuse research.** The big data revolution has impacted all fields, and the accumulation of massive data provides an important foundation for realizing data reuse and thereby maximizing data value. However, how to develop and utilize these data to serve socioeconomic and scientific-technological development has become a tremendous challenge for data reuse research. These challenges arise not only from data storage technology and infrastructure needing improvement but also

from restrictions in data protection laws. Therefore, we must not only focus on the value of data reuse activities but also attend to data quality and address technical, ethical, and legal issues during the data reuse process to standardize data reuse.

(3) **Strengthen theoretical connections to data ecosystems.** The data ecosystem is a co-evolutionary whole, and its application theories do not form independently. For example, data openness and sharing are prerequisites for data reuse, while data management and data literacy are important influencing factors for data reuse. Deepening internal integration within the data ecosystem system can provide stronger support for further development of data reuse.

In conclusion, future data reuse research should further broaden its user groups, pay greater attention to new problems arising during scientific and technological development, and emphasize new requirements posed by the big data era for data reuse research, thereby contributing to scientific and technological innovation, social progress, and national development. This study has certain limitations: the analysis sample includes only literature retrieved from CNKI and Web of Science and manually identified, and despite exhaustive search terms, the structured search strategy inevitably results in relatively low retrieval recall, which we compensated for primarily through snowballing. Additionally, the U.S. National Science Board (NSB) defines data as "any information···including text, numbers, images, video or film, audio, software, algorithms, equations, animations, models, simulations, etc." Searching only with "Data" as the core term without considering that other forms of information may represent data in different disciplines could introduce bias in the results. Future work will continue to track developments in data reuse research to address these limitations and more comprehensively reveal characteristics of domestic and international data reuse research to better promote development in this field.

References

[1] Xing Wenming, Hong Cheng. "Openness as the Norm, Non-Openness as the Exception" : Interpreting Scientific Data Sharing and Utilization in the "Administrative Measures for Scientific Data" [J]. *Library Forum*, 2019, 39(1): 8.

[2] Zhang Xiaoyue, Gu Liping, Hu Lianglin. Review of Solutions to Open Scientific Data Reuse Dilemmas at Home and Abroad [J]. *Library*, 2021, 000(003): 80-89.

[3] Li Yuelin, Zhang Jianwei, Wang Shanshan, et al. Looking Back at the "13th Five-Year Plan" : Characteristics, Shortcomings, and Prospects of Domestic Information Behavior Research [J]. *Journal of Information Resources Management*, 2022, 12(1): 14.

[4] Kim Y, Yoon A. Scientists' data reuse behaviors: A multilevel analysis [J].

Journal of the Association for Information Science & Technology, 2017, 68(12).

[5] Faniel I M, Kriesberg A, Yakel E. Data reuse and sensemaking among novice social scientists [J]. *Proceedings of the American Society for Information Science and Technology*, 2012, 49(1).

[6] Yoon, Ayoung. Data reusers'trust development [J]. *Journal of the Association for Information Science and Technology*, 2016.

[7] Ursic H, Custers B. Legal Barriers and Enablers to Big Data Reuse [J]. *Lexxion Publisher*, 2016(2).

[8] Saxena S. Drivers and barriers towards re-using open government data (OGD): A case study of open data initiative in Oman [J]. *Foresight*, 2018, 20(2): 00-00.

[9] Yoon, Ayoung. Red flags in data: Learning from failed data reuse experiences [C]// 2016: 126.

[10] Wei Yinzhen, Deng Zhonghua, Yang Gaizhen. Research on Influencing Factors of Researchers' Data Reuse Intention [J]. *Library Theory and Practice*, 2020(3): 6.

[11] Li Jialu. Research on Influencing Factors and Promotion Strategies of Researchers' Data Reuse Behavior [D]. Northeast Normal University, 2019.

[12] Custers B, Uršič H. Big data and data reuse: A taxonomy of data reuse for balancing big data benefits and personal data protection [J]. *International Data Privacy Law*, 2016, 6(1): 4-15.

[13] Wang X, Duan Q, Liang M. Understanding the process of data reuse: An extensive review [J]. *Journal of the Association for Information Science and Technology*, 2021, 72(4).

[14] Abella A, Ortiz-de-Urbina-Criado M, De-Pablos-Heredero C. Meloda 5: A metric to assess open data reusability [J]. *El profesional de la información (EPI)*, 2019, 28(6).

[15] Faniel I M, Jacobsen T E. Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data [J]. *Kluwer Academic Publishers*, 2010, 19(3-4): 355-374.

[16] Erik, M, van, et al. Déjà lu: On the limits of data reuse across multiple publications [J]. *Journal of Purchasing & Supply Management*, 2018.

Author Contributions

Lin Xinyue: Paper writing and revision
Yang Jianlin: Paper guidance

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