Postprint: A Study on University Researchers' Willingness for Open Data Sharing and Its Influencing Factors

Authors: Yang Lijuan, Zhang Yuzhen, Hu Qiandai, Sun Qian, Li Han

Date: 2025-06-16T15:42:52+00:00

Abstract

In the era of Big Data, the trend toward open sharing of research data is irreversible. As China's emphasis on research data openness gradually deepens, investigating the data sharing willingness and its influencing factors among researchers in universities—the "primary arena"—holds significant practical importance for enhancing the willingness of university researchers and society at large to openly share research data. This study conducts semi-structured interviews with researchers from both natural and social sciences, employs procedural grounded theory to analyze the interview data, and constructs a model of university researchers' willingness to openly share research data and its influencing factors. The findings reveal that university researchers' willingness to openly share research data is primarily affected by factors such as cost, risk, and demand; furthermore, university researchers in the natural sciences demonstrate higher willingness to openly share research data compared to their counterparts in the social sciences. The study proposes corresponding countermeasures and recommendations, providing intellectual support for various stakeholders to promote research data open sharing in a targeted manner.

Full Text

Abstract

In the era of big data, the trend toward open sharing of scientific research data is unstoppable. As China gradually increases its emphasis on research data openness, investigating the willingness and influencing factors of data sharing among university researchers—who constitute the "main front" of research—holds important practical significance for enhancing data open sharing willingness among university researchers and society at large. This study conducted

semi-structured interviews with researchers from both natural and social sciences, analyzed the interview data using procedural grounded theory, and constructed a model of university researchers' willingness and influencing factors regarding research data open sharing. The findings reveal that university researchers' willingness to openly share research data is primarily influenced by factors such as cost, risk, and demand; moreover, researchers in natural sciences demonstrate higher willingness than their counterparts in social sciences. The study proposes corresponding countermeasures and recommendations to provide intellectual support for all sectors of society to promote research data open sharing in a targeted manner.

---

1. Literature Review

This study reviews existing literature on university researchers' attitudes toward research data open sharing and related influencing factors, focusing on two aspects: influencing factor research and countermeasure research, to provide theoretical support for the current investigation.

Previous research has employed theories from sociology, psychology, and other disciplines to establish or further develop models. Theories such as Planned Behavior Theory, Peer Pressure Theory, Social Cognitive Theory, Social Exchange Theory, New Institutional Theory, and the Technology Acceptance Model have been utilized. Studies by Hu et al. [6], He et al. [7], Liu et al. [2, 8], Sheng et al. [9], Zhang et al. [10], Liu [11], Liu and Gui [12], Sun et al. [13], Bao et al. [14], Zheng [15], Pu [16], Yu [17], Sun and Yan [18], Tenopir et al. [19], Kim and Zhang [20], Kim and Stanton [21], Mason et al. [22], and Zuiderwijk et al. [23] have explored various influencing factors.

The classification of influencing factors on research data open sharing primarily follows three approaches. The first perspective examines the five elements involved in the data sharing process. For instance, Kim [5] discussed how institutional and resource factors affect researchers. Additionally, a few studies mentioned platform and funding influences. The second approach divides factors into direct and indirect influences. He et al. [7] proposed direct factors such as benefits and sharing channels, and indirect factors like researchers' age and discipline. Liu et al. [2] categorized factors into surface-level, middle-level, and deep-level influences. The third approach classifies factors into different hierarchical levels.

Existing literature has thoroughly investigated influencing factors, but most studies rely on quantitative research and lack in-depth qualitative analysis. Few studies have analyzed disciplinary differences in research data open sharing among researchers from different fields, and rarely have studies incorporated

disciplinary domains as control variables when constructing willingness models. This study addresses these gaps by dividing university researchers into natural sciences and social sciences, conducting semi-structured interviews separately, and innovatively introducing disciplinary differences as a control variable to construct an influencing factor model for university researchers' data open sharing willingness.

Countermeasure research has primarily proposed strategies from single stakeholder perspectives. For example, Bao et al. [14] suggested that funding agencies, journals, and other entities should establish policies specifying open content, data formats, quality standards, usage objects, and usage rights. Pu [16] recommended accelerating infrastructure construction such as data open sharing platforms. Other scholars have proposed that researchers should actively improve their data management capabilities, establish correct perceptions of data sharing, and that universities should create incentive mechanisms and issue supportive initiatives [22].

2. Research Design

This study employed interview methods to collect data. Through consultation with relevant experts and open discussions with researchers, the interview outline was designed around six core questions: (1) What do you understand research data to be? (2) What research data do you typically encounter, use, or generate? (3) Are you willing to openly share your research data? (4) Do you have experience with open sharing of research data? (5) What factors do you believe influence your or other researchers' willingness and behavior regarding open sharing? (6) If research data must be openly shared, what protection and support would you need?

Participant selection followed theoretical saturation standards and theoretical sampling methods based on category generation and theoretical development requirements, combined with the research questions. This study selected 23 participants, all faculty members and doctoral students from different universities who fall within the category of university researchers and have faced behavioral choices regarding research data open sharing. The participants were evenly distributed between natural and social sciences fields, each with over three years of research experience. Through one-on-one in-depth interviews lasting approximately 60 minutes each, and with participants' consent, interviews were recorded and transcribed to form original analysis texts, yielding more than 30,000 words of textual records. Participants' basic information was randomly selected to ensure representativeness.

3. Data Coding and Analysis

This study primarily used NVivo software for coding. The research team established coding rules through back-to-back coding principles and group discussions to achieve coding consensus. One interview text was randomly reserved

for theoretical saturation testing. Drawing on Strauss's procedural grounded theory, this study employed three levels of coding: open coding, axial coding, and selective coding.

In the open coding stage, the research team conducted line-by-line analysis of interview texts, ultimately obtaining 87 initial concepts such as data collection cost, achievement preemption risk, and reputation risk. Concepts appearing fewer than 3 times were deleted or merged, and duplicate concepts were consolidated. After open coding, 15 basic categories were formed, including data collection cost, achievement preemption risk, and data ownership.

In the axial coding stage, the 15 basic categories were analyzed to identify logical relationships and form main categories. Five main categories emerged: cost, risk, demand, data open sharing platform, and institutional mechanisms. Cost includes data collection cost and data processing cost. Risk comprises achievement preemption risk, privacy leakage risk, and reputation risk. Demand encompasses economic benefit demand, academic exchange demand, and research advancement demand. The data open sharing platform includes platform availability and platform management level. Institutional mechanisms include publication norms, incentive mechanisms, and data protection mechanisms.

Selective coding involved analyzing the main categories to identify core concepts and construct a theoretical model. The story line reveals that university researchers' willingness to openly share research data is influenced by internal drivers (demand), direct influences (cost, risk, data characteristics), indirect influences (awareness), prerequisite conditions (platform), and external safeguards (institutional mechanisms).

Theoretical saturation testing is crucial in grounded theory. After preliminary model construction, the reserved interview text was used for validation. Results showed no new concepts or categories emerged, indicating theoretical saturation.

4. Influencing Factors Model for University Researchers' Research Data Open Sharing Willingness

Based on coding results, this study constructed an influencing factor model consisting of independent variables, control variables, and dependent variables. The independent variables include cost, data open sharing platform, data characteristics, institutional mechanisms, demand, and risk. The control variable is disciplinary difference (natural sciences vs. social sciences). The dependent variable is university researchers' willingness to openly share research data.

Cost Factors: Cost directly influences willingness. Researchers consider collection costs, processing costs, and time/effort costs when deciding whether to share. As one researcher noted: "Data I purchased at high financial cost, or data requiring significant time and effort to obtain, I'm unwilling to share."

Data Open Sharing Platform Factors: Platform availability is a prerequisite for sharing. The existence and management level of platforms determine

whether researchers can share data smoothly and receive technical support. As one interviewee stated: "Having an authoritative institution or platform for data sharing would be ideal, allowing those who need it to access data through a unified website."

Data Characteristic Factors: Data characteristics directly affect willingness, including confidentiality requirements, ownership, importance, quality, and legal utilization. Sensitive data involving enterprise secrets or national interests cannot be shared.

Institutional Mechanism Factors: Institutional mechanisms provide external safeguards, including publication norms, incentive mechanisms, funding support, data protection, and management mechanisms. Incomplete mechanisms constrain sharing willingness. One researcher mentioned: "Sharing my data doesn't count as an achievement in performance evaluations, nor does it bring rewards, so people aren't motivated."

Demand Factors: Demand serves as internal motivation, including emotional satisfaction, economic benefits, research advancement needs, and achievement verification needs. Some researchers pursue shared benefits, while others expect material rewards.

Risk Factors: Risk directly impacts willingness, including privacy leakage, reputation damage, achievement preemption, and data tampering. These risks significantly reduce sharing willingness.

Disciplinary Differences: Disciplinary differences affect researchers' institutional environments, data characteristics, and platform usage. Natural sciences researchers show stronger willingness and more experience in data sharing than social sciences researchers. Natural sciences data (equipment data, experimental process data) are often reusable and subject to discipline norms requiring material sharing for result verification. Social sciences data (qualitative data, survey data) involve higher collection costs, shorter value cycles, and more privacy concerns, resulting in lower sharing willingness.

5. Strategies for Enhancing Research Data Open Sharing Willingness

Based on the six influencing factors, this study proposes targeted recommendations from national, institutional, and individual perspectives.

National Level: The most crucial measure is to formulate specific, enforceable policies and legal implementation 细则. Although China has issued the Data Security Law and Scientific Data Management Measures, they remain at the national strategic level with poor operability. Detailed regulations for different business scenarios are needed to ensure researchers can share data legally. The state should also supervise stakeholders and promote platform construction.

Institutional Level: Universities and research institutions should improve

institutional mechanisms based on national laws, including revising publication norms, establishing incentive mechanisms (material and spiritual), providing funding support, and creating data protection and management mechanisms. Institutions should align internal systems with public platforms and conduct training to enhance researchers' platform usage.

Individual Level: Researchers should increase their awareness of data sharing, correctly understand its value, and actively improve data management skills. They should balance cost considerations, demand satisfaction, and risk prevention, while establishing proper sharing concepts.

6. Conclusion

This study systematically investigated the influencing factors of university researchers' research data open sharing willingness through interviews and grounded theory analysis. The findings indicate that willingness is primarily affected by cost, risk, and demand. The study constructed a comprehensive influencing factor model and proposed targeted strategies. Disciplinary comparison revealed that natural sciences researchers demonstrate higher sharing willingness and more experience than social sciences researchers.

Theoretically, this study verifies previous research conclusions and innovatively incorporates disciplinary differences as a control variable. Practically, it provides intellectual support for promoting research data open sharing. However, limitations exist: the study employed convenience sampling with a small sample size, and the model's applicability requires further validation. Future research should expand sample size and conduct quantitative verification of the model.

References

[1] Tian X. Research on the Protection Mechanism of Government Data Open Sharing in China: Issues and Countermeasures[J]. Electronic Intellectual Property, 2021(4): 65-77.

[2] Liu L, Zhang C. Analysis of Influencing Factors on Research Data Sharing Based on Interpretive Structural Modeling[J]. Library and Information Service, 2020(5): 27-33.

[3] Zhang B. Review of Foreign Research on Data Sharing Behavior Influencing Factors[J]. Library and Information Service, 2014(4): 53-60.

[4] Zhang X. Investigation on Data Policies of Foreign Research Funding Agencies: Taking British and American Research Councils as Examples[J]. Library and Information Service, 2015(6): 53-60.

[5] Kim S. Institutional, motivational, and resource factors influencing health scientists' data-sharing behaviours[J]. Journal of Scholarly Publishing, 2015(4): 366-389.

[6] Hu Y, Zhang J. Research on Influencing Factors and Mechanisms of Scientific Data Sharing Among Local University Faculty in China[J]. Journal of Agricultural Library and Information Science, 2020(10): 31-42.

[7] He L, Sun L. Research on Researchers' Data Sharing Willingness[J]. Library and Information, 2014(5): 125-131.

[8] Liu L, Zhang C. Research on Influencing Factors of University Researchers' Research Data Open Willingness[J]. Library Science Research, 2019(18): 54-62.

[9] Sheng X, Liu G. Review of Research on Influencing Factors of Scientific Data Open Sharing at Home and Abroad[J]. Information Studies: Theory & Application, 2021(8): 102, 173-179.

[10] Zhang H, Liu Y. Research on Main Influencing Factors of University Scientific Data Sharing[J]. Library and Information Service, 2020(11): 57-64.

[11] Liu Y. Research on Influencing Factors of Scientific Data Sharing Among University Researchers[D]. Nanjing University of Aeronautics and Astronautics, 2019.

[12] Liu G, Pu Y. Research on Influencing Factors and Mechanisms of University Researchers' Research Data Opening[J]. Information Studies: Theory & Application, 2019(22): 73-82.

[13] Sun L, Li Y. Research on Key Influencing Factors of Scientific Data Sharing Based on Meta-ethnography[J]. New Century Library, 2020(3): 52, 91-97.

[14] Bao D, Gu L, Zhang X. Research on Behavioral Influencing Factors of Open Research Data: Taking Earth Science as an Example[J]. Information Studies: Theory & Application, 2019(5): 38-45.

[15] Zheng L. Review of Research on Researchers' Data Sharing Willingness and Influencing Factors[J]. Library Science Research, 2018(9): 39-44, 78.

[16] Pu Y. Research on the Opening Mechanism of University Research Data[D]. Nanjing University, 2019.

[17] Yu L. Research on Influencing Factors of Researchers' Scientific Data Sharing Willingness[D]. Jilin University, 2016.

[18] Sun X, Yan Y. Investigation and Analysis of Factors Influencing Scientific Data Sharing Oriented Toward Researchers: Based on Planned Behavior Theory[J]. Library Science Research, 2019(5): 38-45.

[19] Tenopir C, Allard S, Douglass K, et al. Data sharing by scientists: practices and perceptions[J]. PLoS One, 2011(6): e21101.

[20] Kim Y, Zhang P. Understanding data sharing behaviors of researchers: the roles of attitudes, norms, and data repositories[J]. Library & Information Science Research, 2015(3): 189-200.

[21] Kim Y, Stanton M. Institutional and individual factors affecting scientists' data-sharing behaviors: a multilevel analysis[J]. Journal of the Association for Information Science and Technology, 2016(4): 776-799.

[22] Mason M, Box J, Burns M. Research data sharing at a national science agency: understanding the relative importance of organisational, disciplinary, and domain-specific influences[J]. PLoS One, 2020(8): e0238071.

[23] Zuiderwijk A, Shinde R, Jeng W. What drives and inhibits researchers to share research data? A systematic literature review to analyze factors influencing research data sharing adoption[J]. PLoS One, 2020(9): e0239283.

[24] Zhang J. Research on Scientific Data Sharing Willingness of University Researchers in China[J]. Library and Information Service, 2013(10): 25-30.

[25] Bi D, Wang S. Research on Influencing Factors of Humanities and Social Sciences Data Sharing Willingness: Based on Peer Pressure Perspective[J]. Information and Documentation Services, 2020(4): 64-71.

[26] Sayogo S, Pardo A. Exploring determinants of scientific data sharing: understanding motivation to publish research data[J]. Government Information Quarterly, 2013(Supplement 1): S19-S31.

[27] Kim S. Institutional, motivational, and resource factors influencing health scientists' data-sharing behaviours[J]. Journal of Scholarly Publishing, 2015(4): 366-389.

Figures
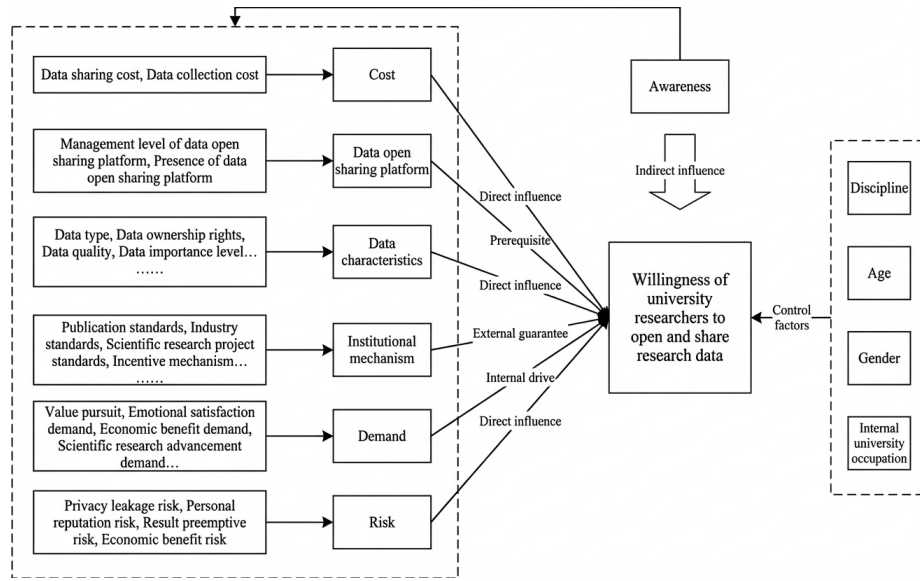
Source: ChinaXiv — Machine translation. Verify with original.

Figure 1: Figure 1