Survey Report on Chinese Researchers' Awareness and Utilization of Preprint Platforms

Authors: ChinaXiv Project Team

Date: 2023-05-21T00:00:00+00:00

Abstract

ChinaXiv (Chinese Academy of Sciences Preprint Platform, https://chinaxiv.org) is oriented towards researchers nationwide, constructing a reliable and standardized open repository for Chinese scientific research papers in the field of natural sciences, receiving preprint submissions and open archiving of published scientific papers in both Chinese and English. It builds an academic exchange ecosystem that supports rapid communication and publication under the Chinese Academy of Sciences standards, supports rapid pre-publication of high-level Chinese scientific research papers, and effectively supports the research priority rights of Chinese scientists. To further promote open academic exchange, facilitate the development of preprint platforms, and improve preprint platform construction, the ChinaXiv project team launched a "Preprint Platform Awareness and Usage Survey" questionnaire survey nationwide for researchers (including graduate students) on April 20, 2022. By 24:00 on May 20, a total of 3,028 questionnaires were collected, of which 2,912 were valid, with an effective rate of 96.17%. Survey findings: 1. The majority of researchers consider preprints "very important" (34.38%) or "somewhat important" (33.31%) in academic exchange. 2. Among the preprint platforms used by researchers, arXiv usage accounts for the highest proportion (66.64%). 3. "Rapid publication and sharing of research results" (60.78%) and "Obtaining network priority rights for research results" (60.69%) are the two main reasons attracting users. 4. "Results published on preprint platforms cannot be used for personal academic evaluation" (47.25%) and "Concern that research results cannot be formally published in journals" (42.82%) are the two main reasons hindering users from using preprint platforms. 5. Researchers generally believe that formulating preprint policies is of significant importance for promoting preprint exchange and preprint platform development in China. Preprint policies at the "national level" and "institutional level" are most needed at the current stage. 6. The preprint policy demands strongest among researchers are "journals allowing papers to be submitted after being posted on preprint

platforms" and "having clear intellectual property ownership identification and protection."

Full Text

Preamble

ChinaXiv (the Chinese Academy of Sciences Preprint Platform, https://chinaxiv.org) is a reliable and standardized open repository for Chinese scientific research papers in natural science fields, serving researchers nationwide. It accepts both preprint submissions and open archiving of published scientific papers in Chinese and English. The platform aims to build a standardized academic communication ecosystem within the Chinese Academy of Sciences that supports rapid dissemination, enabling swift pre-publication of high-quality Chinese research and effectively safeguarding Chinese scientists' right of first disclosure.

To further promote open academic exchange and advance preprint platform development, the ChinaXiv project team launched a nationwide "Preprint Platform Awareness and Usage Survey" on April 20, 2022, targeting researchers (including graduate students). By May 20, 2022, at 24:00, a total of 3,028 questionnaires were collected, of which 2,912 were valid, yielding an effective response rate of 96.17%.

The survey revealed several key findings. First, the majority of researchers considered preprints "very important" (34.38%) or "somewhat important" (33.31%) for academic communication. Second, among the preprint platforms used by researchers, arXiv had the highest usage rate at 66.64%. Third, the two main attractions for users were "rapid publication and sharing of research results" (60.78%) and "securing the right of first disclosure online" (60.69%). Fourth, the two primary barriers to using preprint platforms were "preprint platform publications cannot be used for personal academic evaluation" (47.25%) and "concerns that research results cannot be formally published in journals" (42.82%). Fifth, researchers generally agreed that establishing preprint policies would be significant for promoting preprint exchange and platform development in China, with policies at the "national level" and "institutional level" being most urgently needed. Sixth, the strongest policy demands were for "journals allowing submission after preprint platform publication" and "clear intellectual property rights attribution and protection."

1.1 Data Screening

The questionnaire was distributed nationwide to researchers (including graduate students) from April 20, 2022, to May 20, 2022, for a period of one month. A total of 3,028 questionnaires were collected, and 116 invalid questionnaires (3.83% of the total) were eliminated based on response time, consecutive identical answers, and open-ended question content, resulting in 2,912 valid questionnaires

with an effective rate of 96.17%. The specific screening process was as follows:

Response time-based elimination (71 questionnaires removed): For respondents who had used preprint platforms (project team members required 127 seconds for rapid test completion), 35 questionnaires with response times under 100 seconds were eliminated. For those who had not used preprint platforms (project team members required 58 seconds for rapid test completion), 36 questionnaires with response times under 50 seconds were eliminated.

Consecutive identical answer-based elimination (36 questionnaires removed): Among respondents who selected that they had used preprint platforms, 8 questionnaires with more than 4 consecutive "ABCD" selections were eliminated. Among those who had not used preprint platforms, 28 questionnaires with all three required questions answered as "AB" and overly short response times were eliminated.

Open-ended question content-based elimination (9 questionnaires removed): Six questionnaires with obvious nonsense content such as "啊啊啊" were eliminated. Three questionnaires with unidentifiable or non-extractable information such as "东海县" (Donghai County), "高中毕业" (high school graduate), and "0250" were eliminated.

1.2 Data Standardization

Based on respondents' answers, the author roles and disciplinary fields were standardized as follows:

2 Overall Analysis

The overall analysis yielded the following main conclusions. Researchers who had used preprint platforms demonstrated higher levels of understanding of preprints and greater recognition of their importance in academic communication. Most researchers learned about preprint platforms through "journal articles/books" (60.23%) and "peer recommendations" (55.55%), indicating that promotion through "academic conferences" (25.52%) and "policy documents" (5.68%) needs strengthening. Among researchers who had used preprint platforms (1,109 individuals), arXiv was the most widely used at 68.26%, followed by ChinaXiv at 34.08%. The preferred platform for publishing research was arXiv, with 62.22% having published there, followed by ChinaXiv at 55.56%.

Only 179 respondents had published papers through ChinaXiv. Their main reasons for using ChinaXiv centered on the option of "being able to publish my research results in Chinese," along with "rapid (first-time) sharing with peers" and "securing the right of first disclosure online." Researchers identified major shortcomings of ChinaXiv as "no help in personal academic evaluation processes such as graduation, promotion, or project completion," "lack of organic integration between preprint publication and journal publication," and "absence of reliable peer review."

No significant differences were observed in preprint platform usage behavior between researchers who had published on preprint platforms and those who had not. Usage frequency concentrated on "weekly" or "monthly" visits, with usage duration primarily "less than 1 hour on average" and "1-2 hours on average." Among researchers who had published research results, publication frequency was concentrated at "1-2 papers per year." Based on researchers' reasons for using preprint platforms, a push-pull diagram of researcher usage motivations was drawn.

2.1 Personal Basic Information

The basic information of respondents in the valid questionnaires is as follows: Age distribution was dominated by young and middle-aged researchers. Gender distribution was predominantly male at 67.17%, with females accounting for 32.83%. Educational background showed that those with doctoral degrees comprised more than half at 58.79%. Roles were primarily researchers, master's and doctoral students. Institutions were mainly "higher education institutions" and "research institutes" as frontline research units. Specific details are shown in the table below.

The age distribution was as follows: 25 years and under (19.20%), 26-35 years (30.43%), 36-45 years (28.50%), 46-59 years (19.20%), and 60 years and above (2.68%). Gender distribution was 67.17% male and 32.83% female. Educational qualifications were: bachelor's degree or below (10.71%), master's degree (30.49%), and doctoral degree (58.79%). Roles included: researchers (59.99%), master's or doctoral students (28.81%), administrative staff (3.67%), librarians (2.51%), journal editors (2.34%), and others (1.65%). Institution types were: higher education institutions (63.02%), research institutes (25.24%), enterprises (4.84%), government agencies (3.43%), hospitals (1.44%), and others (0.82%).

Discipline distribution is shown in the table below (listing the top 5 disciplines in the valid questionnaires): Computer Science and Technology, Library and Information Science and Archives Management. The number of respondents and percentage are provided.

In the valid questionnaires, the situation regarding the six disciplinary fields that ChinaXiv aims to prioritize for preprint development is shown in the table below: Computer Science and Technology, Library and Information Science and Archives Management. The number of respondents and percentage are provided.

2.2 Preprint Awareness and Usage

In comparison, researchers who had used preprint platforms demonstrated higher levels of understanding of preprints and greater recognition of their importance in academic communication. Researchers primarily learned about preprint platforms through "journal articles/books" (60.23%) and "peer recommendations" (55.55%), indicating that promotion through "academic

conferences" (25.52%) and "policy documents" (5.68%) requires strengthening.

Among researchers who had used preprint platforms (1,109 individuals), arXiv usage was highest at 68.26%, with ChinaXiv ranking second at 34.08%. The preferred platform for publishing research results was arXiv, with 62.22% having published there, followed by ChinaXiv at 55.56%. Only 179 respondents had published papers through ChinaXiv. Their main reasons for using ChinaXiv focused on "being able to publish my research results in Chinese," along with "rapid (first-time) sharing with peers" and "securing the right of first disclosure online." Researchers identified major shortcomings of ChinaXiv as "no help in personal academic evaluation processes such as graduation, promotion, or project completion," "lack of organic integration between preprint publication and journal publication," and "absence of reliable peer review."

No significant differences were observed in preprint platform usage behavior between researchers who had published on preprint platforms and those who had not. Usage frequency concentrated on "weekly" or "monthly" visits, with usage duration primarily "less than 1 hour on average" and "1-2 hours on average." Among researchers who had published research results, publication frequency was concentrated at "1-2 papers per year." Based on researchers' reasons for using preprint platforms, a push-pull diagram of researcher usage motivations was drawn.

2.3 Functional Expectations and Needs for Preprint Platforms

Comparing researchers' current usage patterns with their expectations for preprint platforms, researchers expressed greater desire to publish, browse, cite, and comment on research results on preprint platforms, while their expectation for "recommending platforms or papers to others" was lower than current usage.

Researchers' strongest functional demands for preprint platforms were basic functions such as "browsing papers," "downloading papers," and "publishing papers." Next were academic exchange functions like "open peer review," "academic community communication," and "preserving different versions of research results." Finally, extended functions such as "plagiarism detection and similarity checking" were also mentioned.

2.4 Preprint Policy Awareness and Needs

Researchers generally agreed that establishing preprint policies is important for promoting preprint exchange and platform development in China, with over 70% considering it "very important" or "somewhat important." Among those who considered preprint policies "very important," "somewhat important," or "moderately important," all prioritized "national-level" and "institutional-level" preprint policies as the most important.

Researchers' strongest policy demands were for "journals allowing submission after preprint platform publication" and "clear intellectual property rights attribution and protection," followed by "preprints being recognized as academic achievements for promotion, graduation, and project completion" and "recognition and protection of the right of first disclosure online."

3 Multi-Perspective Analysis (Age, Role, Disciplinary Field)

Analysis of valid questionnaires by age, role, and disciplinary field yielded the following main conclusions. Researchers across different ages, roles, and disciplines showed varying levels of understanding of preprints, yet almost all affirmed the important value of preprints for academic communication. Researchers' preferences for preprint platforms were highly correlated with disciplinary fields, with library and information science being the only field in this survey that showed some usage of various preprint platforms beyond arXiv and ChinaXiv. Researchers aged 26-45 were "high-engagement" users of preprint platforms, while high-volume publishers were mostly senior researchers aged 46-59 and 60 and above.

Different roles showed different motivational structures for using preprint platforms, yet researchers encountered basically the same obstacles. Except for librarians, researchers in other roles had low expectations for "recommending platforms or papers to others." Senior researchers aged 60 and above attached particular importance to preprint policies, with over 50% considering them "very important." Researchers across different ages, roles, and disciplines all prioritized "national-level" and "institutional-level" preprint policies. The demand for "preprints being recognized as academic achievements for promotion, graduation, and project completion" decreased with age, while the demand for "recognition and protection of the right of first disclosure online" increased with age.

3.1 Preprint Awareness

Researchers aged 36-45 showed a "healthy distribution" in their understanding of preprints (defined as a group showing a sequentially decreasing pattern from "very important" to "not important at all"), indicating that young and middle-aged researchers have a good cognitive foundation for the preprint concept. Most researchers across age groups considered preprints "very important" or "somewhat important" for academic communication, showing high recognition of preprints' academic value.

Researchers showed a "healthy distribution" in their understanding of preprints, indicating a good cognitive foundation for the concept. Frontline researchers such as research staff, master's and doctoral students, and librarians all considered preprints "very important" or "somewhat important" for academic communication, demonstrating high recognition of preprints' academic exchange value.

The fields of "Computer Science and Technology" and "Physics" showed a "healthy distribution" in preprint awareness, indicating a good cognitive foundation for the concept. Most researchers across different disciplines considered preprints "very important" or "somewhat important" for academic communication, showing high recognition of preprints' academic value.

3.2 Preprint Platform Usage Preferences

According to survey results, researchers rarely published on preprint platforms other than arXiv and ChinaXiv. Therefore, this section focuses on detailed comparative analysis of research output publication on arXiv and ChinaXiv.

Researchers across age groups primarily used arXiv and ChinaXiv, with generally higher usage of arXiv, except that researchers aged 60 and above used ChinaXiv more than arXiv. Research output publication volume on arXiv exceeded that on ChinaXiv across all age groups, though researchers aged 25 and under and those aged 60 and above had relatively similar publication volumes on both platforms.

Researchers in different roles primarily used arXiv and ChinaXiv. Frontline researchers such as research staff and graduate students used arXiv more, while librarians and administrative staff used ChinaXiv more than arXiv. Research staff and graduate students preferred publishing on arXiv, with graduate students having relatively similar publication volumes on both platforms, while other roles published less frequently on preprint platforms.

Researchers' platform preferences were highly correlated with disciplinary fields. For example, the biology field used bioRxiv far more than other platforms, while library and information science was the only field showing some usage of various preprint platforms beyond arXiv and ChinaXiv. Library and information science and biology fields had slightly more publications on ChinaXiv than arXiv, while other fields published more frequently on arXiv.

3.3 Preprint Platform Usage Behavior

Relatively speaking, researchers aged 26-45 were "high-engagement" users of preprint platforms, with usage frequency concentrated at "weekly" or "monthly" intervals and usage duration primarily "less than 1 hour on average" and "1-2 hours on average." Over 50% of researchers in this age group published "1-2 papers per year" on preprint platforms. High-volume publishers on preprint platforms were mostly senior researchers aged 46-59 and 60 and above.

No significant differences were observed in preprint platform usage behavior across different roles. Except for librarians, researchers in other roles used preprint platforms with frequency concentrated at "weekly" or "monthly" intervals, and usage duration primarily "less than 1 hour on average" and "1-2 hours on average." Research staff and graduate students predominantly published "1-2 papers per year" on preprint platforms, while most researchers in other roles

never published on preprint platforms.

Disciplinary fields showed no obvious differences in preprint platform usage behavior characteristics. Usage frequency was concentrated at "weekly" or "monthly" intervals across fields, with usage duration primarily "less than 1 hour on average" and "1-2 hours on average." Except for library and information science and biology fields, researchers in other disciplines predominantly published "1-2 papers per year" on preprint platforms. Additionally, "Computer Science and Technology" and "Physics" fields had high-volume users publishing "6-10 papers per year" or "11 or more papers per year" on preprint platforms.

3.4 Motivations and Obstacles for Preprint Platform Usage

Age had minimal impact on researchers' reasons for using preprint platforms, which mainly centered on "peers using preprint platforms," "securing the right of first disclosure online," "rapid publication and sharing of research results," and "expanding research dissemination." Age also had minimal impact on reasons for not using preprint platforms. Except for researchers aged 60 and above who had greater concerns about "inability to obtain reliable peer review," other age groups' reasons for non-use concentrated on "concerns about research being plagiarized," "preprint publications cannot be used for personal academic evaluation," and "concerns about inability to formally publish in journals."

Different roles showed different motivational structures for using preprint platforms. Researchers and graduate students primarily used preprint platforms due to "peers using preprint platforms," "securing the right of first disclosure online," and "rapid publication and sharing of research results." Other roles, influenced by their specific positions in the academic communication system, had more dispersed reasons for use. Role had minimal impact on reasons for not using preprint platforms. Except for administrative staff and journal editors who had greater concerns about "inability to obtain reliable peer review," other roles' reasons for non-use concentrated on "concerns about research being plagiarized," "preprint publications cannot be used for personal academic evaluation," and "concerns about inability to formally publish in journals."

Disciplinary fields showed no significant impact on reasons for using preprint platforms, with options mainly concentrating on "peers using preprint platforms," "securing the right of first disclosure online," "rapid publication and sharing of research results," and "expanding research dissemination." Disciplinary fields also had minimal impact on reasons for not using preprint platforms, with researchers across fields citing "concerns about research being plagiarized," "preprint publications cannot be used for personal academic evaluation," and "concerns about inability to formally publish in journals."

Based on these results, a motivation/obstacle diagram for different researcher roles was created.

3.5 Expectations for Preprint Platform Usage

Researchers across age groups had high expectations for "publishing personal research results" on preprint platforms, particularly early-career researchers aged 25 and under. Senior researchers aged 60 and above had high expectations for "citing research results" but lower expectations for "commenting on research results." Researchers across all age groups had low expectations for "recommending platforms or papers to others."

Researchers in different roles all had high expectations for "publishing personal research results" on preprint platforms, with librarians being particularly notable. Additionally, librarians also had high expectations for citing and commenting on research results on preprint platforms. Except for librarians, researchers in other roles had low expectations for "recommending platforms or papers to others."

Researchers across disciplinary fields had high expectations for "publishing personal research results" on preprint platforms, though expectations for other usage methods varied considerably. Overall, expectations for "recommending platforms or papers to others" were low.

3.6 Functional Requirements for Preprint Platforms

Age had minimal impact on researchers' functional requirements for preprint platforms. All age groups strongly demanded basic functions such as "browsing papers," "downloading papers," and "publishing papers." Next in priority were academic exchange functions like "open peer review," "academic community communication," and "preserving different versions of research results." Additionally, researchers aged 25 and under and those aged 60 and above paid more attention to functions like "plagiarism detection and similarity checking" and "helping to push manuscripts to relevant journals."

Different roles showed some variation in functional requirements. Beyond the universally strong demand for basic functions like "browsing papers," "downloading papers," and "publishing papers," librarians, administrative staff, and journal editors placed greater emphasis on academic exchange functions, showing high demand for "open peer review," "academic community communication," and "preserving different versions of research results." They also showed some demand for extended functions like "platform statistics (research output distribution, paper publication flow, etc.)" and "real-time citation display."

Disciplinary fields had a more noticeable impact on functional requirements. Beyond strong demand for basic and academic exchange functions across fields, library and information science researchers showed relatively balanced demand for various platform functions. Astronomy researchers also paid attention to open comments from the public, while mathematics, computer science and technology, and physics fields placed greater emphasis on "plagiarism detection and similarity checking functions."

3.7 Awareness of Preprint Policy Importance

Researchers across age groups generally agreed that establishing preprint policies is important for promoting preprint exchange and platform development in China. Senior researchers aged 60 and above attached particular importance to preprint policies, with over 50% considering them "very important." Researchers across all age groups prioritized "national-level" and "institutional-level" preprint policies as most important, followed by "journal-level" policies.

Researchers in different roles generally agreed on the importance of preprint policies, with librarians being particularly notable as over 50% considered them "very important." Researchers across all roles prioritized "national-level" and "institutional-level" preprint policies as most important, followed by "journal-level" policies.

Researchers across disciplinary fields generally agreed on the importance of preprint policies, with computer science and technology, library and information science, and physics fields having over 50% considering them "very important." Researchers across all fields prioritized "national-level" and "institutional-level" preprint policies. Compared to "journal-level" policies, researchers in mathematics, computer science and technology, physics, and astronomy fields placed greater importance on "association-level" preprint policies.

3.8 Preprint Policy Requirements

Age had some influence on researchers' preprint policy requirements. All age groups prioritized "journals allowing submission after preprint platform publication" and "clear intellectual property rights attribution and protection." Additionally, the demand for "preprints being recognized as academic achievements for promotion, graduation, and project completion" decreased with age, while the demand for "recognition and protection of the right of first disclosure online" increased with age.

Role had minimal influence on preprint policy requirements. Researchers across different roles all prioritized "journals allowing submission after preprint platform publication" and "clear intellectual property rights attribution and protection." Additionally, librarians, administrative staff, and graduate students had stronger demand for "preprints being recognized as academic achievements for promotion, graduation, and project completion."

Disciplinary fields had minimal influence on preprint policy requirements. Researchers across fields all prioritized "journals allowing submission after preprint platform publication" and "clear intellectual property rights attribution and protection." Additionally, library and information science and biology fields had relatively stronger demand for "preprints being recognized as academic achievements for promotion, graduation, and project completion," while computer science and technology and physics fields placed greater importance on "recognition and protection of the right of first disclosure online."

3.9 Profiles of Early-Career and Senior Researchers' Preprint Awareness and Usage

Based on questionnaire results, profiles of early-career and senior researchers' preprint awareness and usage were drawn.

Based on data processing and analysis results, the following recommendations are proposed:

First, regarding concept promotion, priority could be given to disciplinary fields with better preprint awareness foundations, such as computer science and technology and physics, to expand and enhance preprint influence through key demonstration cases. Efforts should be made to deepen understanding of preprint platforms across more fields and strengthen promotional efforts through academic conferences and policy documents. Senior researchers and other high-volume publishers should be actively engaged to leverage their influence in attracting more early-career researchers to use and trust preprint platforms.

Second, regarding functional development, the research result publication process should be optimized to enhance platform usability, security, and overall friendliness to researchers. Review mechanisms should be strengthened to alleviate researchers' concerns about "inability to obtain effective peer review," thereby improving overall preprint quality. Communication communities should be enriched through active collaboration with academic communities to expand the user base of professional researchers.

Third, regarding policy priorities, efforts should be made to actively promote the introduction of national-level and institutional-level preprint policies to strengthen top-level design. Collaboration with academic associations should be enhanced, particularly in mathematics, computer science and technology, physics, and astronomy, to raise awareness at the association level and formulate relevant policies to provide a supportive environment for researchers using preprints. Coordination with journals should be emphasized to promote journal group support for preprints and foster organic connections between preprints and journals to advance academic communication.

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