International Open Science Research Advances Postprint

Authors: Huang Ruhua, Zhao Yang, Huang Yuting

Date: 2023-04-01T16:02:44+00:00

Abstract

[Purpose/Significance] Open science can render scientific research more collaborative, transparent, and efficient, and has garnered widespread global attention. From an international perspective, systematically reviewing and analyzing current research topics in open science facilitates the advancement of this field and provides references and insights for subsequent research. [Method/Process] Employing a combined methodology of literature survey and information visualization, this study selects papers on open science research indexed in the Web of Science platform, and integrates them with relevant government documents, research reports, news articles, and other literature to reveal the thematic distribution of international open science research. [Result/Conclusion] International open science research topics encompass: open access research, data sharing research, research output reuse research, knowledge innovation research, and infrastructure construction research. Future recommendations advocate for systematic and in-depth research from three dimensions: policy system, infrastructure, and stakeholder participation.

Full Text

International Research Progress in Open Science

Huang Ruhua, Zhao Yang, Huang Yuting School of Information Management, Wuhan University, Wuhan 430072

Abstract: [Purpose/Significance] Open science can make scientific research more collaborative, transparent, and efficient, attracting widespread global attention. Identifying and analyzing current research themes in open science from an international perspective can promote development in this field and provide reference for future research. [Method/Process] Using a mixed method combining literature survey and information visualization, this paper analyzed articles on open science included in the Web of Science platform, together with relevant policy documents, research reports, and news literature, to reveal the thematic distribution of international open science research. [Result/Conclusion] International open science research themes include: open access research, data sharing research, research output reuse, knowledge innovation research, and infrastructure construction research. Future research should conduct systematic and in-depth studies from three aspects: policy system, infrastructure, and stakeholder participation.

Keywords: open science; open data; open access; data sharing

Open science will bring a major revolution to scientific research and has attracted widespread attention from the global scientific community. In March 2020, to discuss the importance of international scientific cooperation in the context of the COVID-19 pandemic, UNESCO convened representatives from scientific departments of multiple countries for an online meeting, calling for strengthened "open science and collaboration" [1]. Previously, in September 2019, UNESCO had already initiated the development of the "UNESCO Recommendation on Open Science" [2]. From an academic research perspective, foreign scholars have conducted comprehensive studies on the concept [3], framework [4], and practice [5-6] of open science, while research in China is still in the initial stages of concept analysis [7-8] and experience borrowing [9], and a complete open science research system has not yet been established.

Based on this, this paper adopts a mixed method of quantitative and qualitative analysis to review international open science research, aiming to 梳理 the progress in this field and provide reference for future research.

1 Research Design

1.1 Concept of Open Science

As an emerging topic, consensus on the basic issues such as the connotation and extension of open science has not yet been reached. Therefore, before conducting literature analysis, it is necessary to explain the concept of open science and determine the retrieval strategy accordingly.

Currently, there is no unified view on the concept of open science in academia and industry. The five major schools of open science research include the infrastructure school (involving technical architecture), the public school (involving accessibility of knowledge creation), the metrology school (involving altmetrics), the democratic school (involving knowledge access), and the pragmatic school (involving collaborative research) [10]. This paper mainly refers to the European Commission' s definition of open science, which is the practice of implementing, disseminating, and transforming scientific research through digital tools and networks. Open science depends on the joint influence of technological development and cultural change on research collaboration and openness. By providing unlimited, barrier-free, and open access to research outputs, open science aims to make the scientific research process more efficient, transparent, and effective

[11]. The FOSTER portal, funded by the EU, has constructed an open science taxonomy that divides open science into thematic areas including open access, open data, open reproducible research, open science evaluation, open science guidelines, open science policies, open science projects, and open science tools [12].

1.2 Data Sources and Research Methods

The data for this paper mainly comes from the Web of Science Core Collection database. Considering that research on different components of open science (such as open access, open data, open resources) is extensive, if the search terms covered all components, the retrieval results would be very large and could not study open science from a macro perspective. Therefore, this paper set the search formula as TS=("open science" OR "opening science"), selected the three citation indexes SCI-E, SSCI, and A&HCI, limited the document type to article, and the time range to 1900-2019, retrieving a total of 731 records. After manually removing irrelevant data, 386 valid records were obtained, including full records and cited references. From the publication year perspective, open science research began in 1985, and the number of papers has grown rapidly since 2014, which is closely related to policy guidance from various countries and international organizations in recent years. To more comprehensively understand the latest developments in international open science, this study focused on 386 documents, supplemented by relevant policy documents, research reports, research plans, news reports, and other literature.

This paper selected VOSviewer visualization analysis software, which can construct co-occurrence networks based on keywords and provides good clustering analysis effects. Keywords condense the essence of an article and can reflect its core ideas to a certain extent. This paper set the keyword frequency threshold to 2. After preprocessing the keywords, including unifying singular/plural forms, uppercase/lowercase, and abbreviations/full names, merging synonyms and near-synonyms into common terms, and deleting words without practical meaning or unrelated to the theme, 332 keywords were finally selected to draw a keyword co-occurrence clustering map of open science research.

As shown in Figure 1 [Figure 1: see original paper], there are 5 main clusters that intersect and associate with each other, showing close connections. The measurement indicators for nodes in Figure 1 include occurrences and total link strength, representing keyword frequency and total co-occurrence times with other keywords, respectively, which can be used to identify core keywords. Table 1 shows the top 20 core keywords ranked by frequency. The following sections will sequentially analyze the five research themes according to cluster size, combined with core keywords and literature content.

2 Research on Open Access

Cluster 1 is the largest and most core cluster, with densely distributed keyword nodes forming a research theme centered on open access (OA). Other core keywords include models, impact, network, information, and systems. Since its proposal in the 1990s, open access has been a hot topic in academia and industry. The "Budapest Open Access Initiative" (BOAI) released in 2002 first clearly defined open access [13]. Subsequently, the open access movement entered a vigorous development stage and is considered an important practice of open science [14].

2.1 Open Access Models

The academic community has extensively discussed various open access models to improve effectiveness, including multiple publishing models divided by publishing carrier and availability, and review models focusing on open peer review.

2.1.1 Open Access Publishing Models Open access publishing models are shown in Table 2 , among which Green OA and Gold OA are the two main publishing models. M. Laakso et al. divided Gold OA into Direct OA, Delayed OA, and Hybrid OA based on journal content availability [15]. Some scholars also consider Hybrid OA alongside Green OA and Gold OA as the main types of open access publishing models [16]. In addition, less discussed but common publishing models include Bronze OA [16] and Black OA [17].

Many scholars have investigated and compared different open access models. J.C. Valderrama-Zur [18] and L. Zhang [19] compared different open access models from education, physics, and other research fields. H. Piwowar et al., through the open online service tool oaDOI, found that the proportion of open access articles is increasing, driven by Gold OA and Hybrid OA [16]. By studying the citation impact of open access articles, H. Piwowar et al. also confirmed that under the influence of Green OA and Hybrid OA, open access articles receive 18% more citations than the average level [16].

2.1.2 Open Access Review Models The current open access model has led to the emergence of predatory journals, which conduct peer review in a very superficial manner and accept manuscripts within days to collect publication fees [20]. Open peer review can not only curb predatory journals but also form more rigorous (through communication with academic communities) and relevant (through communication with other stakeholders) research [20]. T. Ross-Hellauer conducted a systematic review of open peer review, summarizing it as disclosing reviewer and author identities, publishing review reports, and involving more experts in the review process [21]. L. Dobusch and M. Heimstädt, by analyzing predatory publishing in management research, called for open peer review, believing it helps achieve transparency in the review process, distinguish high-quality journals from predatory journals, and establish a sustainable review culture [20].

2.2 Impact of Open Access

As a new publishing model, open access facilitates the free dissemination of research results through the network, promotes academic information exchange and utilization, and has generated significant academic, economic, and social impacts.

2.2.1 Academic Impact of Open Access The academic impact of open access is mainly reflected in academic citations and academic communication. In terms of citations, open access literature typically has a citation advantage [22-23]. Although this advantage varies across disciplines and research fields, the overall trend indicates that open access literature helps improve academic impact [24]. In terms of academic communication, open access can increase scholars' access to academic resources and promote the dissemination of their research results [25]. X. Wang et al. believe that open access literature also has advantages in altmetrics [26], is more likely to receive social media attention, and thus generates broader academic impact in informal academic communication.

2.2.2 Economic Impact of Open Access The economic impact of open access is reflected in both funding and costs. In terms of funding, open access can not only save financial expenditures for publishers and research funders but also improve the economic benefits of traditional subscription-based journals [27]. R.P. Holley pointed out that large commercial publishers can benefit from open access by charging Article Processing Charges (APCs) and acquiring small publishing houses [25]. In terms of costs, open access to research results and data can save access costs, labor costs, and transaction costs. M.J. Fell believes that if open access allows higher accessibility and result utilization, then R&D returns will be higher [28]. Compared with paywalled journals, open access journals are more marketable, practical, and attractive, providing greater access, visibility, and loyalty [6].

2.2.3 Social Impact of Open Access Open access not only benefits scholars but also has a broad social impact. J.P. Tennant et al. proposed that open access literature supports research by any user who can access the internet and retrieve and read materials, providing an equitable environment for promoting citizen science and lifelong learning [27]. S.E. Hampton et al. believe that open access can support academic resource acquisition for scholars, institutions, and countries with limited resources, further promoting scientific democratization and knowledge diversification [29].

3 Research on Data Sharing

Cluster 2 forms a research theme centered on data sharing, with core keywords including data sharing, transparency, open data, policy, and challenges. Cluster 2 is relatively dispersed and partially overlaps with Cluster 1, reflecting the close connection between openness and sharing. Data sharing can not only improve research transparency and reproducibility, prompting researchers to improve research quality, but also stimulate secondary innovation and promote field development when other researchers use shared data [30-31].

3.1 Data Sharing Policies

Governments, international organizations, research funding agencies, publishers, academic journals, universities, libraries, and other entities have issued research data policies including data sharing to promote open science development. At the national level, the United States, Australia, and the United Kingdom are among the countries that have issued the most policies. The EU announced in its new research and innovation framework program "Horizon 2020" [32] that it would fully implement an open research data policy to further promote the "open science" strategy. A. Zuiderwijk and M. Janssen established a comparative framework for open data policies (including environment and context, policy content, performance indicators, and public value factors) to compare seven policies at different government levels in the Netherlands [33].

Research data policies of funding agencies, publishers, and journals are key factors in promoting researchers to share research data, and research on these three types of policies is the current focus. C. Neylon analyzed multiple national and international research funding agencies to examine the purposes and models of research data sharing (RDS) policies and provide policy recommendations for funding agencies [34]. L. Jones et al. analyzed multi-level data sharing policies provided by publishers such as Taylor & Francis, Springer Nature, Elsevier, and Wiley [35]. Since most disciplines disseminate new research results through journal publications, journal data sharing policies have a significant impact on when and where researchers provide research data. Multiple scholars have analyzed the progress and models of journal data sharing policies from different fields and disciplines [36-37]. D.M. Gorman, by analyzing research data policies of 250 journals, pointed out that current journal research data policies urgently need standardization and unification [38]. To develop a universal framework for research data policies, the Research Data Alliance (RDA), jointly initiated by the EU, U.S. National Science Foundation, U.S. National Institute of Standards and Technology, and Australian Department of Innovation, released the "Journal and Publisher Research Data Policy Master Framework" in 2018 [39] to ensure that journals and publishers continue to support and encourage data sharing across disciplines.

3.2 Challenges of Data Sharing

Although most researchers hold positive attitudes toward data sharing [40-41], they generally believe that data sharing still faces many obstacles and challenges, mainly including: Legal, normative, or ethical issues such as privacy and ethics involved in the data sharing process [42-44]. For example, in clinical medicine, data usually involves patient privacy, and unrestricted data sharing would trigger a series of problems; the sharing of confidential data is even more related to national security. Convenience and technical issues of data sharing [30,45]. If the convenience and technical support of data sharing do not meet expectations, it may cause time constraints, resource constraints, and funding constraints. Competition of interests caused by data sharing. F. Rockhold et al. pointed out that some researchers may worry about commercial or academic competitors benefiting from analyses based on shared data [46]. Some scholars have also revealed challenges facing data sharing from a global perspective, including insufficient time and data misuse at the individual level; insufficient data sharing training, lack of incentive mechanisms, and lack of internal policy guarantees at the institutional level; and weak policies, ethical and legal norms, lack of data infrastructure, and interoperability issues at the international level [47].

3.3 Data Sharing Practices

Various disciplines have developed numerous data sharing projects and platforms, providing references for data sharing practices. The Yale University Open Data Access (YODA) Project developed a data sharing platform to improve public health and promote data sharing, providing researchers with convenient access to clinical trial data [48]. The Ontario Brain Institute (OBI) relies on its large-scale information platform Brain-CODE to promote research data sharing in neuroscience, and its robust and scalable data governance structure is worthy of reference for various data sharing platforms [49]. Chemoinformaticians also publicly provide their models to the academic community, enabling researchers from different fields to manage, visualize, analyze, and model chemical genomics data through the open-access web portal ChemBench [50].

4 Research on Output Reuse

The core keywords of Cluster 3 include reproducibility, replication, replicability, incentives, psychology, and preregistration, forming a cluster centered on reproducibility and replicability, with research themes focusing on open science practices aimed at output reuse.

4.1 Reproducibility

Reproducibility refers to obtaining consistent research results using the same data, computational steps, methods, code, and analysis conditions as the original research [51]. When data are available, discoverable, reproducible, and well-described, scientists can avoid duplicate research and innovate based on existing research [21]. Currently, fields such as ecology, astronomy, climatology, neuroscience, and oceanography attach great importance to research result reproducibility. How to improve research reproducibility has become a focus of scholarly attention. B. Marwick summarized the basic principles of reproducibility research [52]: publicly provide data and code and store them in appropriate locations, use programming languages to write scripts for data analysis and visualization, use version control to manage multiple versions of files and collaborator contributions, and document and share the computational environment for data analysis. J.J. Brito et al. proposed improving reproducibility research in biomedicine by teaching computational skills, developing data and software, implementing reproducible research, and implementing incentive mechanisms [53]. In addition, establishing scientific workflow-based management systems, using reproducibility tools, and auditing publication reproducibility are also important methods to improve reproducibility research [51]. In the library field, the ReproHackNL organized by the Leiden University Library Digital Scholarship Center provided suggestions and references on how to improve research reproducibility through community engagement [54].

4.2 Replicability

Replicability is stricter than reproducibility, referring to obtaining consistent research results using new data or new computational methods when addressing the same scientific problem [51]. Compared with reproducibility, replicability requires experiments to be conducted and data to be collected again. The "Transparency and Openness Promotion Guidelines" (TOP Guidelines) formulated by the U.S. Center for Open Science (COS) includes replicability as one of the transparency and openness modular standards, using three hierarchical levels to classify the degree of support journals provide for replicability research submissions [55], reflecting the importance of replicability in open science practice.

In recent years, psychology, neuroscience, and related fields have frequently experienced a "replication crisis," due to insufficient research capacity, publication bias, imprecise theories, imperfect statistical procedures, insufficient sample sizes, and inappropriate analysis methods [56-57]. To improve research replicability, M.J. Larson recommended reducing researchers' flexibility in data analysis, reporting sample size information, increasing sample sizes through collaboration, and improving reporting standards (following established reporting data guidelines and increasing the adoption of preregistration and registered reports) [57]. Other recommendations include establishing open science guidelines, awarding open science badges and rewards, checking errors and anomalous results before publication, encouraging publication of all research results, encouraging replicable research, and formulating effective funder policies [51].

5 Research on Knowledge Innovation

Compared with other clusters, Cluster 4 has relatively high-frequency core keywords including knowledge, innovation, intellectual property, technology, and patents, reflecting a research theme centered on knowledge innovation, specifically manifested in research on the patent system and open licenses. How to establish a balance between intellectual property protection and open science and use intellectual property to promote knowledge innovation has become a focus of academic attention.

5.1 Patent System

How the patent system adapts to open science development is a current research focus. Many empirical studies have shown that the patent system hinders open science development to some extent [58]. Scientists delay publishing research results they wish to patent and are unwilling to share research materials and results with other scientists. Contrary to its intended purpose, the patent system does not necessarily provide incentives for patent applicants to disclose information about inventions. In addition, J.S. Gans et al. pointed out that patent quality may be negatively correlated with the scientific impact of related papers [59].

Notably, ownership itself is not a barrier to openness and sharing. The patent system aims to disclose information about inventions, and clarifying ownership is a necessary prerequisite for sharing [4]. Instead, licensing methods may have an important impact on open science. Some research institutions, such as the Montreal Neurological Institute, are exploring "patent-free" and open-use approaches to promote open sharing of research results [60], believing this is a fundamental requirement of open science—users can utilize publicly funded research results without restriction except for standard citation requirements. To resolve the conflict between the patent system and open sharing of scientific results, M. Bentwich proposed mandatory Provisional Patented Paper Applications (PPPAs) [58] based on internationally recognized Provisional Patent Applications (PPAs), requiring patent applicants to grant simple licenses to others conducting research on their inventions while protecting patent and intellectual property integrity, making patent applicants obligated to share scientific research results with others through scientific publications.

5.2 Open Licenses

Open licenses refer to licenses that impose minimal restrictions on the use of materials or products by users [61], with research objects mainly involving open access publications, data, and open-source software and hardware.

Open access publications typically use Creative Commons (CC) licenses. CC licenses provide six basic model clauses, among which CC BY and CC BY-SA are free licenses; another widely used content open license is the GNU Free Documentation License (GFDL) for manuals, textbooks, and other reference materials. For data, in addition to CC licenses, the open data sharing website (www.opendatacommons.org) has also created a set of license agreements specifically for databases. Meanwhile, the UK has launched the "Open Government License," facilitating data reuse by government and other public sector organizations [62]. Open-source software licensing issues have received considerable attention [63], with open licenses allowing free use, modification, and sharing of software. For open-source hardware, the Open Source Hardware Association recommends seven licenses ranging from general licenses such as CC BY-SA to more specific licenses like the CERN Open Hardware License and TAPR Open Hardware License [64]. R.M. Luis Felipe et al. described the EU' s latest open license agreements, especially provisions for open-source hardware, and elaborated on the prospects and challenges of extending open-source software license terms to open-source hardware [65].

6 Research on Infrastructure Construction

Cluster 5 has fewer keyword nodes, with core keywords including collaboration, database, and evolution, reflecting a research theme centered on infrastructure construction research. Infrastructure in open science practice is mainly manifested in: distributed computing, which utilizes the computing power of multiple users for research; and social and collaborative networks for scientists to support researcher interaction and collaboration [10]. Research on e-infrastructures, open science tools, and data repositories constitutes the current research theme.

6.1 E-infrastructure

Enabling e-infrastructure ensures the exchange and analysis of scientific information and data provided through data repositories. It includes advanced services embedded in some repositories, supercomputing facilities, and distributed computing networks [10]. For example, the Open Science Grid supports large-scale data-intensive research projects by connecting multiple computers to high-performance computer networks. M. Altunay et al. pointed out that e-infrastructure provides scientists and researchers with a collaborative research environment to jointly solve distributed computing problems [66]. However, e-infrastructure is affected by factors such as national security, privacy and confidentiality, commercial sensitivity, and intellectual property; the cost of data collection, provision, and storage; and challenges to international data flows are also important issues facing e-infrastructure [4].

6.2 Open Science Tools

Open science tool research mainly includes three categories. The first category is tools for the entire research process, such as the social website myExperiment introduced by D. De Roure et al. for discovering, sharing, and curating scientific workflows and experimental plans [67]. The second category is platforms that integrate multiple needs, such as the U.S. Open Science Framework (OSF).

The third category is digital tools for various research stages, such as open lab notebooks, collaborative writing tools, open workflows, collaborative bibliographies, and text and data mining sharing tools. With the development of various tools, researchers need to spend more time learning and updating skills. How to improve the interactivity and usability of open science tools has become an urgent research problem to be solved.

6.3 Data Repositories

Data repositories play a key role in supporting open science, connecting with other electronic infrastructures, and integrating digital platforms for research data management [4]. Some online repositories, such as runmycode.org, also provide public access to code and data for research publications. Existing research mainly focuses on data sharing and openness of data repositories, such as K. Nishikawa' s investigation from a knowledge-sharing perspective on how Japanese data repositories manage research data, constructing an analytical framework based on resource openness, community openness, and infrastructure openness [68]. V. Xafis and M.K. Labude outlined strategies for data sharing through data repositories and discussed a series of ethical issues arising from health-related data sharing [69]. Current problems facing data repositories mainly include non-uniform international sharing standards, non-standardized transnational network governance [4], and how to achieve sustainable business models [70].

7 Summary and Outlook

7.1 Summary

International open science research covers a wide range of fields, including psychology, education, information science, biology, economics, and other disciplines, showing a highly interdisciplinary characteristic. Research countries are mainly in Europe and North America, including the United States, United Kingdom, Germany, Canada, France, Netherlands, Italy, Spain, etc. Core research institutions include University College London, McGill University, Oxford University, University of North Carolina at Chapel Hill, University of Wisconsin, and other universities. Research themes focus on open access research, data sharing research, research output reuse, knowledge innovation research, and infrastructure construction research.

Overall, international open science research is in a developmental stage, with many studies on sub-themes such as open access, open data, and data sharing, but fewer studies on the macro-level of open science. Research limitations include: Lack of systematic research on policy content systems. Current research on open science policies focuses on open access and data sharing policies and strategies, being relatively specific, but insufficient on macro-level open science policy systems. Lack of research on infrastructure throughout the entire research process. Although current research involves multiple open science platforms and tools, the research points are relatively single and scattered, located at the edge of thematic clusters, and the research is not yet mature. Insufficient research depth on open science stakeholders. Although current research involves researchers, libraries, publishers, and other stakeholders, research on how each stakeholder participates in open science practice is insufficient, and further exploration is needed on how to improve each stakeholder' s awareness and skills.

7.2 Future Outlook

Based on the above analysis and combined with domestic and international policies and practices, future systematic and in-depth research on open science is recommended from the following three aspects:

7.2.1 Emphasize Systematic Policy System Research Open science policies can provide guidance and basis for practice. Related research can focus on: (1) National-level policy content systems. As open science becomes increasingly globalized and open science policies at different levels increase, how to construct a national-level open science policy content system has become a focus of attention. In addition to the currently widely discussed open access and open data policies, research should also focus on other components of open science, thinking from the perspective of the entire research process, especially focusing on topic selection, ideas, and methods before publication, as well as academic communication after publication. (2) Multi-stakeholder policy implementation mechanisms. Since open science development involves numerous stakeholders such as research funding organizations, research performing institutions, researchers, publishers, and libraries, both macro policy guidance and micro policy implementation details are needed [72]. Research should clarify the rights and obligations of each stakeholder, providing theoretical basis for formulating specialized policies for each stakeholder, such as scientific research funding policies for funding agencies, scientific research evaluation system policies for research performing institutions, research output sharing policies for publishers, and output integration and utilization policies for libraries. (3) Coordinated development strategies with related policies. The "Preliminary Study of the Technical, Financial and Legal Aspects of the Desirability of a UNESCO Recommendation on Open Science" emphasizes that developing open science requires solving issues such as intellectual property, data protection, and data privacy [73]. Therefore, while developing open science policies, coordination with policies on intellectual property, data protection, and data privacy is needed.

7.2.2 Improve Full-Process Infrastructure Research Building and improving infrastructure throughout the entire research process is the foundation for ensuring smooth open science practice. Throughout current infrastructure construction, although numerous in quantity, there is a lack of integrated infrastructure throughout the entire research process. It is recommended to establish a research project management platform throughout the entire process of research design, data collection, data cleaning, data analysis, output publication, dissemination and communication, and impact assessment. Infrastructures such as the U.S. Open Science Framework (OSF) and the European Open Science Cloud (EOSC) can provide references for research.

In the process of building and improving infrastructure, relevant standards must first be considered. The FAIR (Findable, Accessible, Interoperable, Reusable) principles, as an international methodology, particularly emphasize enhancing the ability of machines to automatically find and use data and support individuals in reusing data [74], and can be used to guide infrastructure construction. Second, attention should be paid to infrastructure performance, including how to establish easy-to-use metadata management systems, how to evaluate data quality and provenance, and how to develop data integration and unified analysis and visualization functions. Issues such as website dynamics, long-term data preservation, and participation, customization, and scalability based on different needs are also worth exploring. Finally, infrastructure interactivity needs to be considered, including platform access and interaction methods, convenience of online scientific communication, security, and sustainability.

7.2.3 Strengthen Multi-Dimensional Stakeholder Participation Research

(1) Establish and improve incentive mechanisms. The EU's "Open Science Policy Platform Recommendations" [75] places "rewards and incentives" first among eight priority development recommendations for open science. Effective incentive mechanisms help enhance researchers' open science awareness and promote open sharing behaviors in science and research. Related research needs to consider three aspects: how to provide financial support for open science work, such as how government departments, funding agencies, and management institutions provide economic support for open science guidance and training; recognizing and rewarding researchers participating in open science practice, such as establishing more complete citation standards for data and code, and reforming professional assessment and compensation systems; and considering broader evaluation indicators for open science and its impact, such as altmetrics indicators.

(2) Conduct open science education. The FOSTER portal [76] funded by the EU and the U.S. Center for Open Science [77] both provide various materials and training courses for open science, providing practical experience for related research. Research should clarify the subjects, objects, content, and models of open science education. Education subjects are mainly universities and research institutes, with libraries also playing important roles, needing to design appropriate open science training courses and materials. Education objects include all researchers, librarians, teachers, students, and other stakeholders participating in open science practice. Education content includes but is not limited to basic theories, practical methods,

and ethical norms of open science. Education models can adopt diversified training models combining online and offline, small classes and large lectures.

(3) Promote citizen science development. UNESCO' s 2019 "World Science Day for Peace and Development" took open science as its theme, emphasizing that open science is a sustainable global movement that makes scientific research and data accessible to all [78]. Citizen science, as a new open innovation model, helps promote open science development, harnessing collective wisdom to drive scientific and social progress [79]. However, citizen science also faces some problems, such as large capability gaps among participants, insufficient participation motivation, and inadequate funding support. Future research can focus on how to conduct citizen science projects and enhance public participation awareness, including studying the guarantee systems and operation models of citizen science projects, quality strategies for data provided by the public, public information literacy education, technical support represented by Information and Communications Technology (ICT), and the role and positioning of libraries.

References

[1] UNESCO. UNESCO mobilizes 122 countries to promote open science and reinforced cooperation in the face of COVID-19[EB/OL].[2020-06-10]. https://en.unesco.org/news/unesco-mobilizes-122-countries-promote-open-science-and-reinforced-cooperation-face-covid-19. [2] UNESCO. UNESCO takes the lead in developing a new global standard-setting instrument on open science[EB/OL].[2020-06-20]. https://en.unesco.org/news/unesco-takes-lead-developing-new-global-standard-setting-instrument-open-science. [3] Vicente-Saez R, Martinez-Fuentes C. Open Science now: a systematic literature review for an integrated definition[J]. Journal of business research, 2018, 88: 428-436. [4] Dai Q, Shine E, Smith C. Open and inclusive collaboration in science: a framework[EB/OL].[2019-12-20]. https://doi.org/10.1787/2dbff73f-en. [5] Cook BG, Lloyd JW, Therrien WJ. Open science in the field of emotional and behavioral disorders[J]. Education and treatment of children, 2019, 42(4): 579-600. [6] Banks GC, Field JG, Oswald FL, et al. Answers to 18 questions about open science practices[J]. Journal of business and psychology, 2019, 34(3): 257-270. [7] Sheng Xiaoping, Yang Zhiyong. Analysis of the relationship among open science, open sharing, and open data[J]. Library and Information Service, 2019, 63(17): 15-22. [8] Liu Guifeng, Qian Jinlin, Tian Lili. Open science: concept analysis, system analysis, and concept exploration[J]. Library Tribune, 2018, 38(11): 1-9. [9] Huang Yuting, Lai Tong. U.S. Center for Open Science: practice progress, characteristics, and implications[J]. Library and Information, 2019(3): 105-113. [10] Fecher B, Friesike S. Open science: one term, five schools of thought[M]//Opening science. New York: Springer, 2014: 17-47. [11] European Commission. Open

science[EB/OL].[2020-06-15]. https://ec.europa.eu/digital-single-market/open-science. [12] FOSTER. Open science definition[EB/OL].[2020-06-02]. https://www.fosteropenscience.eu/taxonomy/term/100. [13] Chan L, Cuplinskas D, Eisen M, et al. Read the budapest open access initiative[EB/OL].[2020-06-05]. https://www.budapestopenaccessinitiative.org/read. [14] UNESCO. Open science movement[EB/OL].[2020-06-05]. http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/open-science-movement/. [15] Laakso M, Welling P, Bukvova H, et al. The development of open access journal publishing from 1993 to 2009[J]. PloS one, 2011, 6(6): e20961. [16] Piwowar H, Priem J, Lariviere V, et al. The state of OA: a large-scale analysis of the prevalence and impact of open access articles[J]. PeerJ, 2018, 6: e4375. [17] Björk BC. Gold, green, and black open access[J]. Learned publishing, 2017, 30(2): 173-175. [18] Valderrama-Zurián JC, Aguilar-Moya R, Gorraiz J. On the bibliometric nature of a foreseeable relationship: open access and education[J]. Scientometrics, 2019, 120(3): 1031-1057. [19] Zhang L, Watson E. The prevalence of green and grey open access: where do physical science researchers archive their publications?[J]. Scientometrics, 2018, 117(3): 2021-2035. [20] Dobusch L, Heimstädt M. Predatory publishing in management research: a call for open peer review[J]. Management learning, 2019, 50(5): 607-619. [21] Ross-Hellauer T. What is open peer review? A systematic review[J]. F1000research, 2017, 6: 1-38. [22] Lawrence S. Free online availability substantially increases a paper' s impact[J]. Nature, 2001, 411: 521. [23] Hajjem C, Harnad S, Gingras Y. Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact[J]. IEEE data engineering bulletin, 2005, 28(4): 39-47. [24] McKiernan EC, Bourne PE, Brown CT, et al. How open science helps researchers succeed[J]. ELife, 2016, 5: e16800. [25] Holley RP. Open access: current overview and future prospects[J]. Library trends, 2018, 67(2): 214-240. [26] Wang X, Liu C, Mao W, et al. The open access advantage considering citation, article usage and social media attention[J]. Scientometrics, 2015, 103(2): 555-564. [27] Tennant JP, Waldner F, Jacques DC, et al. The academic, economic and societal impacts of open access: an evidence-based review[J]. F1000research, 2016, 5: 1-49. [28] Fell MJ. The economic impacts of open science: a rapid evidence assessment[J]. Publications, 2019, 7(3): 1-30. [29] Hampton SE, Anderson SS, Bagby SC, et al. The Tao of open science for ecology[J]. Ecosphere, 2015, 6(7): 1-13. [30] Roudere JN. The what, why, and how of born-open data[J]. Behavior research methods, 2016, 48(3): 1062-1069. [31] Hardwicke TE, Mathur MB, MacDonald K, et al. Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal cognition[J]. Royal Society open science, 2018, 5(8): 180448. [32] European Commission. Open science (Open access)[EB/OL].[2020-07-20]. https://ec.europa.eu/programmes/horizon2020/en/h2020-section/open-science-open-access. [33] Zuiderwijk A, Janssen M. Open data policies, their implementation and impact: a framework for comparison[J]. Government information quarterly, 2014, 31(1): 17-29. [34] Neylon C. Compliance culture or culture change? The role of funders in improving data management and sharing practice amongst researchers[J]. Research ideas and outcomes, 2017, 3: e14673. [35] Jones L, Grant R, Hrynaszkiewicz I. Implementing publisher policies that inform, support and encourage authors to share data: two case studies[J]. Insights, 2019, 32(1): 1-11. [36] Rousi M, Laakso M. Journal research data sharing policies: a study of highly-cited journals in neuroscience, physics, and operations research[J]. Scientometrics, 2020, 124: 131-152. [37] Naughton L, Kernohan D. Making sense of journal research data sharing policies[J]. Science and engineering ethics, 2020, 26(3): 1625-1632. [38] Gorman DM. Availability of research data in high-impact addiction journals[J]. Insights, 2016, 29(1): 84-89. [39] Data Policy and Standardisation and Implementation Interest Group. Journal and publisher research data policy master framework DRAFT v1.2 Feb 2018[EB/OL].[2020-06-20]. https://docs.google.com/document/d/1DTAfOKkE1a2n2f_1hGcrXlrw-5Tq_AL5tk-ju8B82_E/edit. [40] Yancy CW, Harrington RA, Bonow RO. Data sharing—the time has (not yet?) come[J]. JAMA cardiology, 2018, 3(9): 797-798. [41] Abele-Brehm AE, Gollwitzer M, Steinberg U, et al. Attitudes toward open science and public data sharing: a survey among members of the German Psychological Society[J]. Social psychology, 2019, 50(4): 252-260. [42] Thorogood A. Canada: will privacy rules continue to favour open science?[J]. Human genetics, 2018, 137(8): 595-602. [43] Hoeyer K, Tupasela A, Rasmussen MB. Ethics policies and ethics work in cross-national genetic research and data sharing: flow, non-flows, and overflows[J]. Science, technology & human values, 2017, 42(3): 381-404. [44] Chawinga WD, Zinn S. Global perspectives of research data sharing: a systematic literature review[J]. Library & information science research, 2019, 41(2): 109-122. [45] Zuiderwijk A, Spiers H. Sharing and re-using open data: a case study of motivations in astrophysics[J]. International journal of information management, 2019, 49: 228-241. [46] Rockhold F, Bromley C, Wagner EK, et al. Open science: the open clinical trials data journey[J]. Clinical trials, 2019, 16(5): 539-546. [47] Chawinga WD, Zinn S. Global perspectives of research data sharing: a systematic literature review[J]. Library & information science research, 2019, 41(2): 109-122. [48] Ross JS, Waldstreicher J, Bamford S, et al. Overview and experience of the YODA Project with clinical trial data sharing after 5 years[J]. Scientific data, 2018, 5: 180268. [49] Lefaivre SC, Behan B, Vaccarino AL, et al. Big data needs big governance: best practices from Brain-CODE, the Ontario Brain Institute' s neuroinformatics platform[J]. Frontiers in genetics, 2019, 10: 191. [50] Capuzzi SJ, Kim ISJ, Lam WI, et al. ChemBench: a publicly accessible, integrated cheminformatics portal[J]. Journal of chemical information and modeling, 2017, 57(2): 105-108. [51] National Academies of Sciences, Engineering, and Medicine. Reproducibility and replicability in science[M]. Washington: The National Academies Press, 2019. [52] Marwick B. Computational reproducibility in archaeological research: basic principles and a case study of their implementation[J]. Journal of archaeological method and theory, 2017, 24(2): 424-450. [53] Brito JJ, Li J, Moore JH, et al. Recommendations to enhance rigor and reproducibility in biomedical research[J]. Gigascience, 2020, 9(6): giaa103. [54] Hettne K, Propper T, Nab L, et al. ReproHackNL

2019: how libraries can promote research reproducibility through community engagement[EB/OL].[2020-07-17]. https://doi.org/10.31235/osf.io/6f4zv. [55] Mellor DT. Guidelines for Transparency and Openness Promotion (TOP) in journal policies and practices "The TOP Guidelines" [EB/OL].[2020-07-17]. https://osf.io/9f6gx/wiki/Guidelines/. [56] Lewandowsky S, Oberauer K. Low replicability can support robust and efficient science[J]. Nature communications, 2020, 11: 358. [57] Larson MJ. Improving the rigor and replicability of applied psychophysiology research: sample size, standardization, transparency, and preregistration[J]. Biofeedback, 2020, 48(1): 2-6. [58] Bentwich M. Changing the rules of the game: addressing the conflict between free access to scientific discovery and intellectual property rights[J]. Nature biotechnology, 2010, 28(2): 137-139. [59] Gans JS, Murray FE, Stern S. Contracting over the disclosure of scientific knowledge: intellectual property and academic publication[J]. Research policy, 2017, 46(4): 820-835. [60] Owens B. Montreal institute going "open" to accelerate science[EB/OL].[2020-07-10]. http://www.sciencemag.org/news/2016/01/montreal-institute-going-open-accelerate-science. [61] Open Data Institute. Publisher' s guide to open data licensing[EB/OL].[2020-07-10]. https://theodi.org/article/publishers-guide-to-open-data-licensing/. [62] Korn N, Oppenheim C. Licensing open data: a practical guide[EB/OL].[2020-07-10]. http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide. [63] Rosen L. Open source licensing: software freedom and intellectual property law[M]. Upper Saddle River: Prentice Hall, 2004. [64] Open Source Hardware Association. Open-Source hardware FAQ[EB/OL].[2020-07-15]. http://www.oshwa.org/faq/. [65] Luis Felipe RM, Kau T, Laia PP, et al. Open hardware licenses: parallels and contrasts[EB/OL].[2020-08-31]. https://helda.helsinki.fi/bitstream/handle/10138/312235/ki011983eenn.pdf?sequence=1. [66] Altunay M, Avery P, Blackburn K, et al. A science driven production cyberinfrastructure—the open science grid[J]. Journal of grid computing, 2011, 9(2): 201-218. [67] De Roure D, Goble C, Aleksejevs S, et al. Towards open science: the myExperiment approach[J]. Concurrency and computation: practice and experience, 2010, 22(17): 2335-2353. [68] Nishikawa K. How are research data governed at Japanese repositories? A knowledge commons perspective[J]. Aslib journal of information management, 2020, 72(5): 837-852. [69] Xafis V, Labude MK. Openness in big data and data repositories: the application of an ethics framework for big data in health and research[J]. Asian bioethics review, 2019, 11(3): 255-273. [70] OECD. Business models for sustainable research data repositories[EB/OL].[2020-07-22]. https://doi.org/10.1787/302b12bb-en. [71] United Nations. Open science conference 2019[EB/OL].[2020-07-10]. http://research.un.org/conferences/OpenScienceUN. [72] Gu Liping. Data management services in the transformation of research models: approaches to achieving open access, open data, and open science[J]. Journal of Library Science in China, 2018, 44(6): 43-58. [73] UNESCO. Preliminary study of the technical, financial and legal aspects of the desirability of a UNESCO recommendation on open science[R/OL].[2020-07-05]. https://unesdoc.unesco.org/ark:/48223/pf0000370291. [74] Wilkinson M, Dumontier M, Aalbersberg I, et al. The FAIR guiding principles for scientific data management and stewardship[J]. Scientific data, 2016, 3: 160018. [75] European Commission. Open science policy platform recommendations[EB/OL].[2020-07-02]. https://ec.europa.eu/research/openscience/pdf/integrated_advice_opspp_recommendations.pdf#view= [76] FOSTER. About FOSTER[EB/OL].[2020-07-02]. https://www.fosteropenscience.eu/about. [77] Center for Open Science. Openness and reproducibility research practices training[EB/OL].[2020-07-02]. https://cos.io/our-services/training-services/. [78] UNESCO. World science day for peace and development[EB/OL].[2020-07-02]. https://zh.unesco.org/commemorations/worldscienceday#theme. [79] Jin Ying, Zhang Xiaolin, Hu Zhihui. Development and challenges of citizen science[J]. Library and Information Service, 2019, 63(13): 28-33.

Author Contributions:

Huang Ruhua: Topic selection, guidance in writing the paper; Zhao Yang: Data collection, writing and revising the paper; Huang Yuting: Providing paper revision suggestions, proofreading.

Research Progress in the International Research of Open Science

Huang Ruhua, Zhao Yang, Huang Yuting School of Information Management, Wuhan University, Wuhan 430072

Abstract: [Purpose/significance] Open science can make scientific research more collaborative, transparent and efficient, which has attracted widespread attention worldwide. Sorting out and analyzing the current research themes of open science from an international perspective is conducive to promoting the development of this field and providing reference and experience for subsequent research. [Method/process] Using a mixed method combining literature survey and information visualization, this paper analyzed articles on open science included in the Web of Science platform, combined with relevant government documents, research reports, news reports and other literature. [Result/conclusion] The research topics of international open science include: open access research, data sharing research, research reuse, knowledge innovation research, and infrastructure construction research. In the future, it is recommended to conduct systematic and in-depth research from three aspects: policy system, infrastructure, and subject participation.

Keywords: open science; open data; open access; data sharing

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