Trends in Scholarly Publishing

OPEN ACCESS

Bridging the Metadata Gap: Open Science Readiness in Pakistan's Scholarly Publishing Landscape

[1]Shafia Arshad, [2]Mujtaba Ellahi Mahar and [3]Sabeen Saher

[1]The Government Sadiq College Women University, Bahawalpur, Pakistan

[2]Department of Chemistry, Faculty of Arts and Basic Sciences, Balochistan University of Information Technology and Management Sciences, Quetta, Pakistan

[3]ACE College for Women, Faisalabad, Pakistan

Corresponding Author: Shafia Arshad (shafia.arshad@iub.edu.pk)

ABSTRACT

The global movement toward open science has raised expectations for journals to implement persistent identifiers, structured metadata, and preservation systems that enhance interoperability, discoverability, and research impact. However, the extent to which journals in developing regions meet these standards remains unclear. This study provides the first systematic assessment of metadata infrastructure and open science readiness across 728 scholarly journals published in Pakistan. Using a descriptive, cross-sectional design, journals were evaluated against five core indicators: DOI assignment, ORCID integration, JATS/XML metadata, archiving services, and online submission systems, along with open access policies, APC transparency, and geographic orientation. The assessment reveals substantial infrastructural limitations. While over 95% of journals provide online submission systems and 96% operate as open access, fewer than 7% require ORCID identifiers and under 3% supply JATS/XML metadata. DOI adoption is inconsistent (about 60%), and only one-third of journals with DOIs report an archiving service, raising concerns about long-term preservation. A composite readiness score shows that only 0.4% of journals meet all five indicators, whereas most journals (64%) demonstrate only moderate compliance. Additionally, authorship and readership patterns indicate a predominantly national focus, reflecting limited international visibility. These findings underscore the structural and policy constraints affecting Pakistan's scholarly publishing ecosystem. Targeted measures such as mandatory DOI and ORCID policies, incentives for metadata standardization, national preservation infrastructure, and greater editorial internationalization are essential to bridge existing gaps. Strengthening these foundational components will improve the global visibility, interoperability, and credibility of Pakistan's scholarly output and facilitate its integration into the global research ecosystem.

KEYWORDS

Open science, metadata standards, DOI, ORCID, JATS/XML, scholarly publishing, Pakistan, journal infrastructure, persistent identifiers, research visibility

Copyright © 2026 Arshad et al. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

ISSN: 2789-9446 (Online)
ISSN: 2789-9438 (Print)
https://doi.org/10.21124/tsp.2026.90.100

Received: 10 Dec. 2025
Accepted: 30 Apr. 2026
Published: 30 May, 2026

INTRODUCTION

The pursuit of open science has led to mounting expectations that scholarly publications adopt infrastructure standards that enhance interoperability, discoverability, and long-term preservation[1]. At the core of this transformation are persistent identifiers (such as DOIs and ORCID) and structured metadata formats like JATS/XML, which together form the backbone of modern academic publishing[2,3]. DOIs ensure persistent access links and cross-platform citation, strengthening research credibility and integration into global knowledge networks[4]. ORCID identifiers, meanwhile, address author name ambiguity and streamline attribution workflows, with adoption steadily increasing across disciplines[5]. JATS/XML has become the de facto standard for scholarly content exchange due to its machine-readability and compatibility with indexing systems[6].

Despite these advancements, adoption remains uneven, especially within smaller or regional publishing ecosystems. Many journals in developing contexts still rely on legacy formats like PDF or basic HTML without embedding rich metadata, limiting their discoverability and integration with indexing services[7]. Such infrastructural gaps reduce visibility, citation potential, and the broader impact of research outputs. Recent analyses of XML and metadata adoption trends show that while leading publishers have fully transitioned to machine-readable workflows, many regional journals continue to lag, creating a widening visibility divide[8].

Pakistan offers a compelling case for examining these dynamics. With more than 700 active peer-reviewed journals across diverse fields, the country's scholarly publishing landscape is growing rapidly. However, systematic evidence about how many of these journals comply with international standards for metadata, identifiers, and archiving is lacking. This study addresses that gap by assessing Pakistani journals against five core dimensions of open science infrastructure: DOI assignment, ORCID requirement, JATS/XML availability, archiving service use, and online submission systems. Additionally, we examine open access practices, APC transparency, and geographic composition to contextualize these trends.

By combining descriptive statistics, comparative analysis, and a readiness scoring approach, this study presents the first comprehensive overview of metadata and identifier maturity in Pakistan's journal ecosystem. The results aim to inform editors, policymakers, and funding agencies about existing gaps and actionable strategies for improving visibility, interoperability, and global integration.

MATERIALS AND METHODS

Study design and data sources: This study adopted a descriptive and cross-sectional research design to evaluate the metadata practices and infrastructural readiness of scholarly journals published in Pakistan. Data were compiled from publicly available journal websites, official publisher pages, and indexing platforms between January and June 2025. The dataset consisted of 726 journals spanning a broad range of subject disciplines, including agricultural and biological sciences, physical sciences, engineering, medicine, and social sciences.

Variables and data extraction: For each journal, metadata were recorded across 15 key variables reflecting international best-practice criteria: Journal title, ISSN (print and electronic), publisher, subject category and sub-category, editorial leadership, current issue availability, online submission system, DOI assignment, ORCID requirement, geographic composition, open-access status, APC transparency, archiving service, and availability of JATS/XML metadata. These variables align with open-science infrastructure recommendations outlined by NISO and related metadata standards[9,10]. Duplicate records and incomplete entries were excluded, and inconsistent field names (such as variations in "XML metadata" terminology) were standardized before analysis.

Data analysis: Descriptive statistics were generated to summarize the prevalence of individual features such as DOI assignment, ORCID implementation, XML metadata, archiving practices, and submission platforms. To explore relationships between variables, comparative analyses were conducted, for example, examining correlations between online submission systems and DOI assignment, or between open access and APC transparency. Such two-dimensional comparisons are widely used in journal-level metadata assessments to reveal patterns across disciplines and publishing models[11,12].

Readiness score construction: A Readiness Score (0-5) was developed to assess each journal's alignment with global metadata and open-science benchmarks. Journals received one point each for the presence of a DOI assignment, ORCID requirement, JATS/XML metadata, archiving service, and an online submission system. This composite indicator approach has been used in previous studies to evaluate journal infrastructure maturity and interoperability[13,14]. The aggregated scores were analyzed to identify the proportion of journals meeting none, some, or all of these criteria.

RESULTS

A total of 726 journals were included in this study, representing a diverse range of subject areas and publishing models. The findings reveal substantial variation in the adoption of open science infrastructure and metadata standards across the national publishing landscape.

Subject distribution: The dataset covered journals from across major disciplines, with the largest representation in medicine, social sciences, and agricultural sciences. This disciplinary spread provides a broad view of national publishing practices shown in Table 1.

Adoption of core scholarly infrastructure: Infrastructure maturity varied widely across journals (Fig. 1). While online submission systems were nearly universal, appearing in over 95% of journals, other infrastructure elements lagged. DOIs were assigned by just over 60% of journals, and fewer than 7% required authors to provide ORCID identifiers. XML/JATS metadata, essential for machine readability and interoperability, was present in less than 3% of titles.

Table 1: Distribution of journals by subject area

Subject category	No. of Journals	Total of (%)
Social Sciences	174	23.90
(No subject listed)	141	19.37
Arts and Humanities	137	18.82
Business, Management and Accounting	51	7.01
Medicine	47	6.46
Health Professions	36	4.95
Agricultural and Biological Sciences	27	3.71
Biochemistry, Genetics and Molecular Biology	22	3.02
Computer Science	20	2.75
Multidisciplinary	18	2.47
Economics, Econometrics and Finance	11	1.51
Engineering	11	1.51
Psychology	7	0.96
Pharmacology, Toxicology and Pharmaceutics	4	0.55
Immunology and Microbiology	3	0.41
Materials Science	3	0.41
Mathematics	3	0.41
Nursing	3	0.41
Chemistry	3	0.41
Earth and Planetary Sciences	2	0.27
Dentistry	2	0.27
Physics and Astronomy	1	0.14
Energy	1	0.14
Veterinary	1	0.14

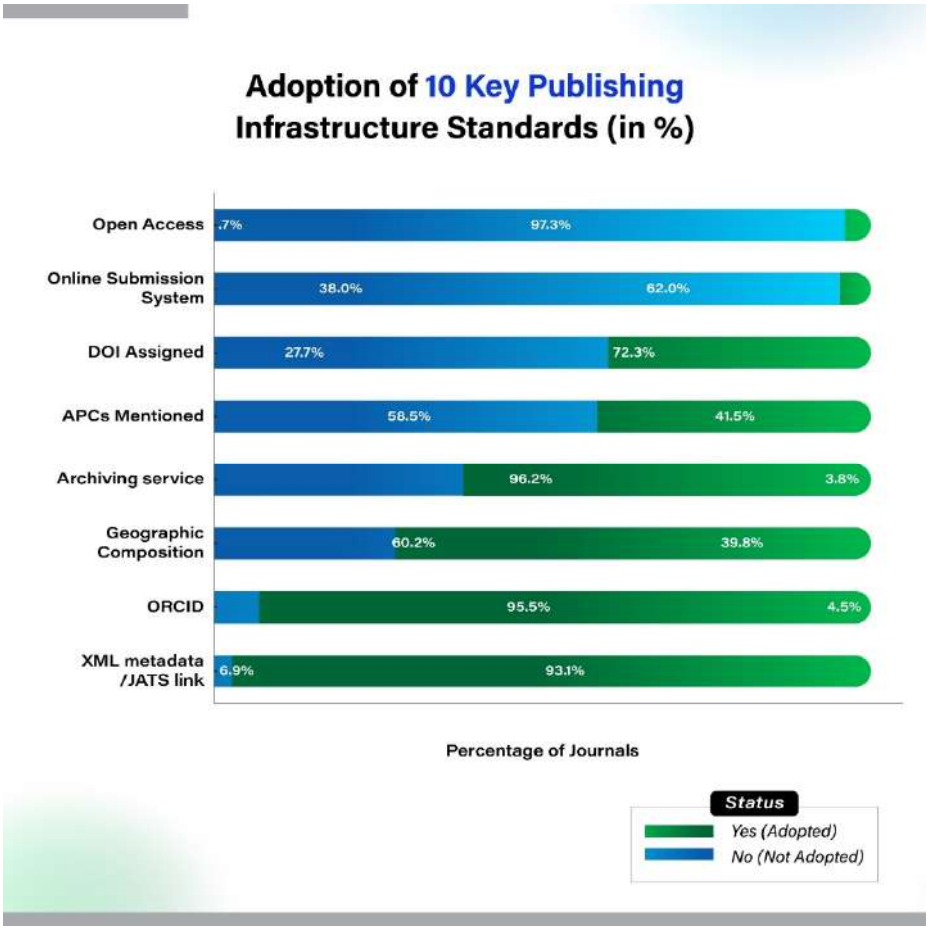


Fig. 1: Percentage of journals adopting core practices (DOI, ORCID, XML/JATS, Archiving, OA, etc.)

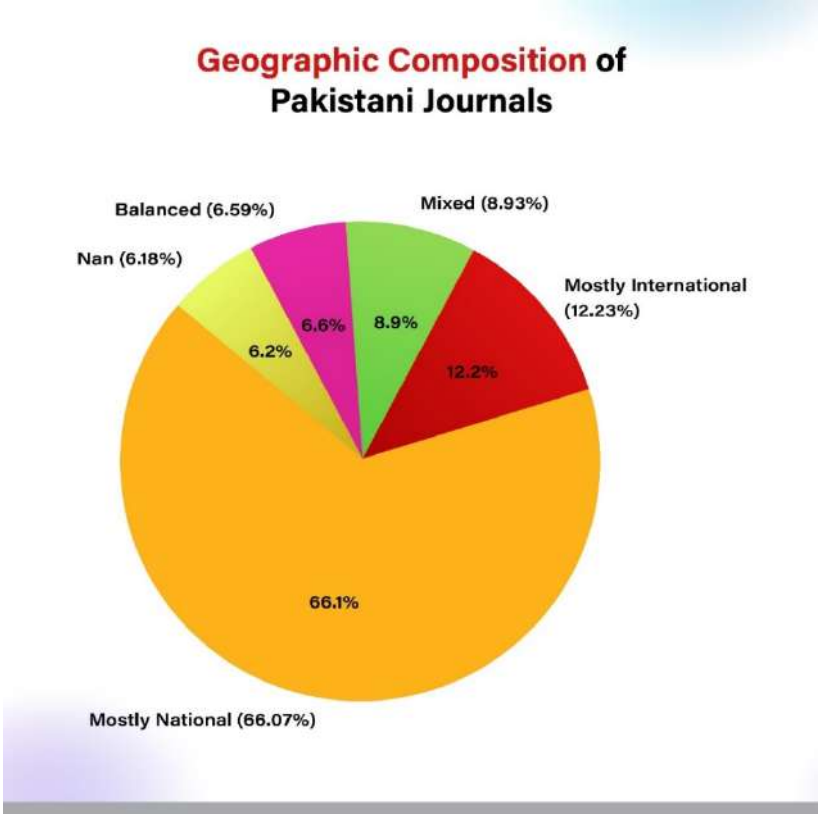


Fig. 2: Geographic composition of journals

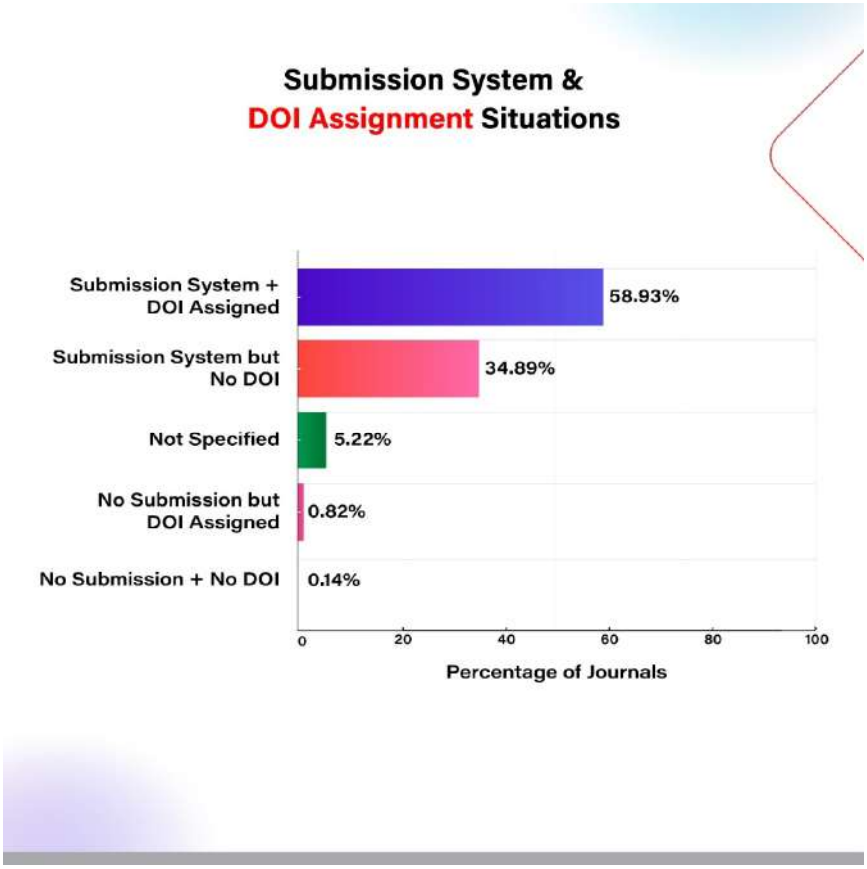


Fig. 3: DOI assignment vs. submission system availability

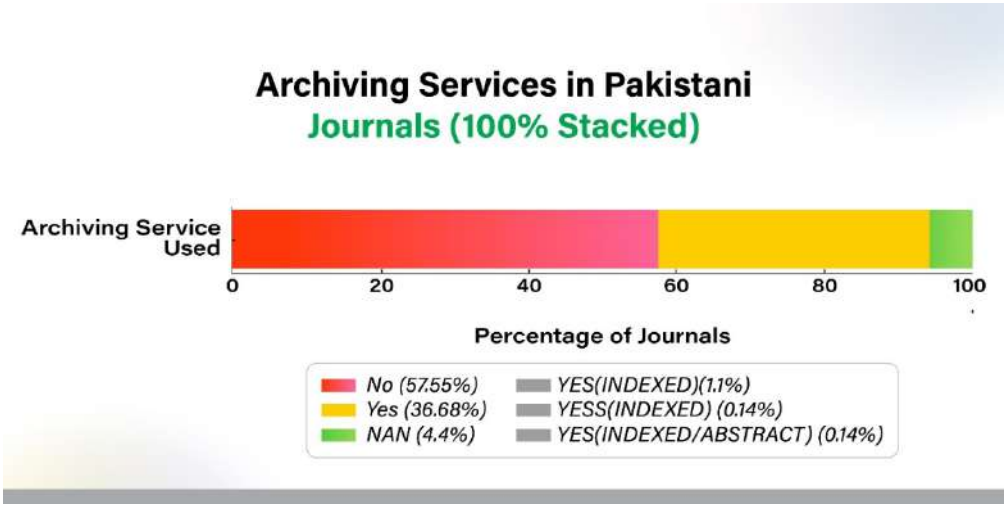


Fig. 4: DOI assignment vs. archiving service usage

Geographic orientation of journals: A majority (~66%) of journals had a predominantly national authorship and readership, while only about 12% described themselves as mostly international. A small fraction was balanced or explicitly mixed in their geographic composition in Fig. 2.

Metadata and preservation gaps: Cross-sectional comparisons further illustrate structural disconnects between different components of the publishing workflow. Among journals with online submission systems, more than a third had not implemented DOI assignment, suggesting a disconnect between digital workflows and persistent identification. Similarly, only about 30% of journals combined DOI assignment with a declared archiving solution, leaving the majority without guaranteed long-term preservation in Fig. 3-4.

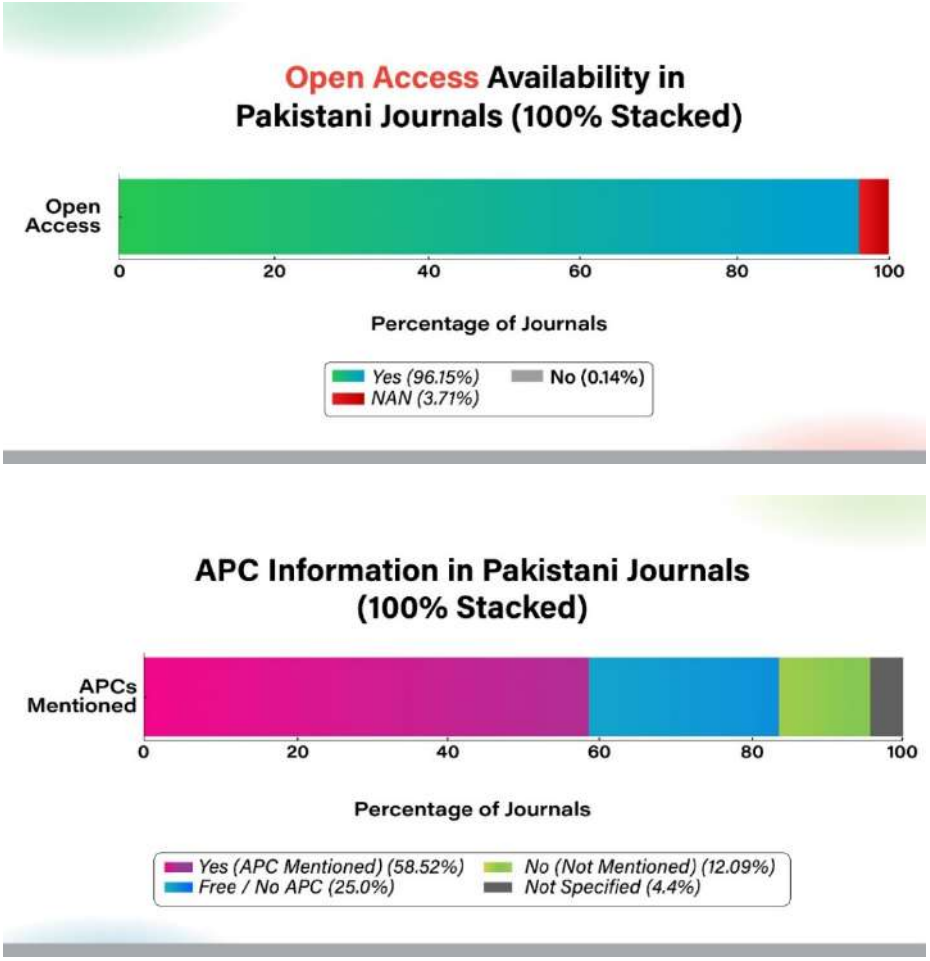


Fig. 5: Open Access status vs. APC disclosure

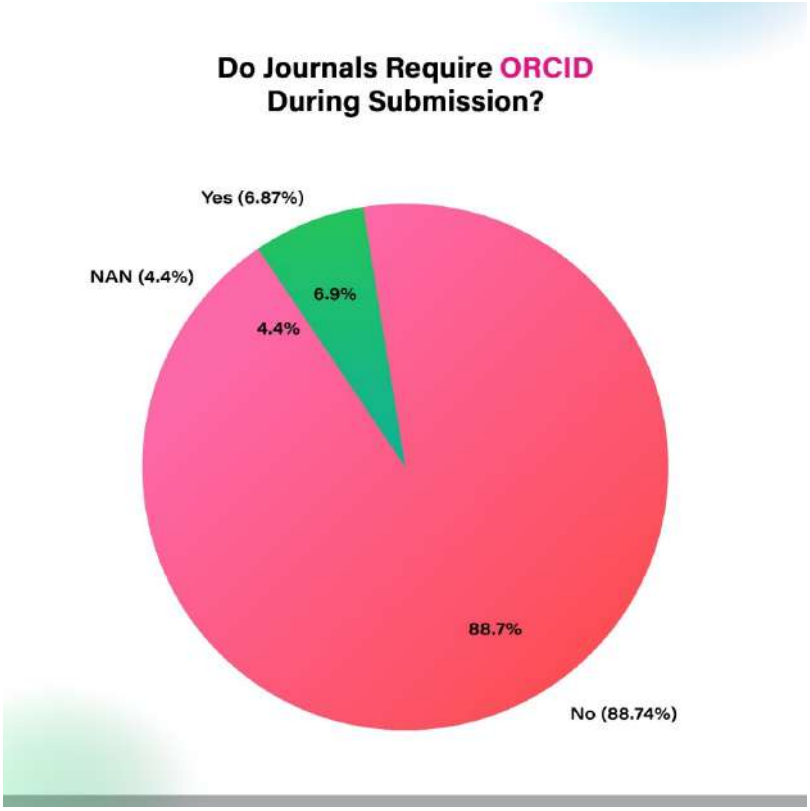


Fig. 6: ORCID requirement by subject area

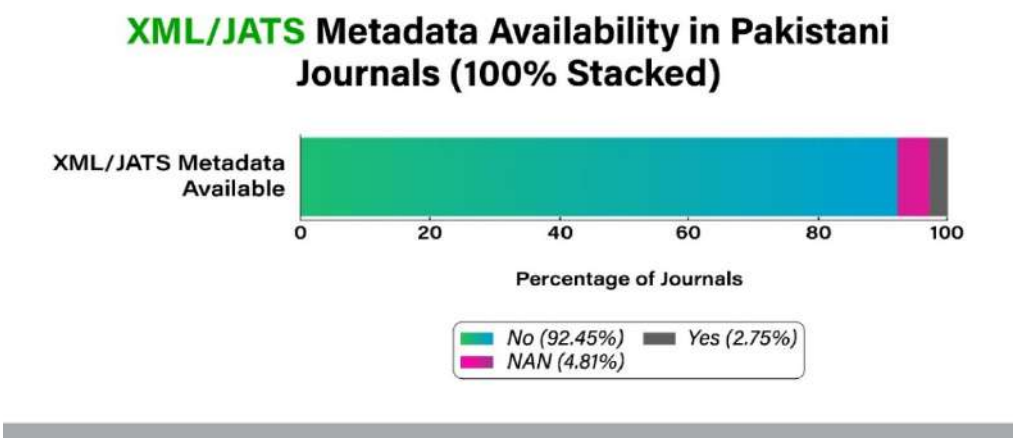


Fig. 7: XML/JATS metadata availability by subject area

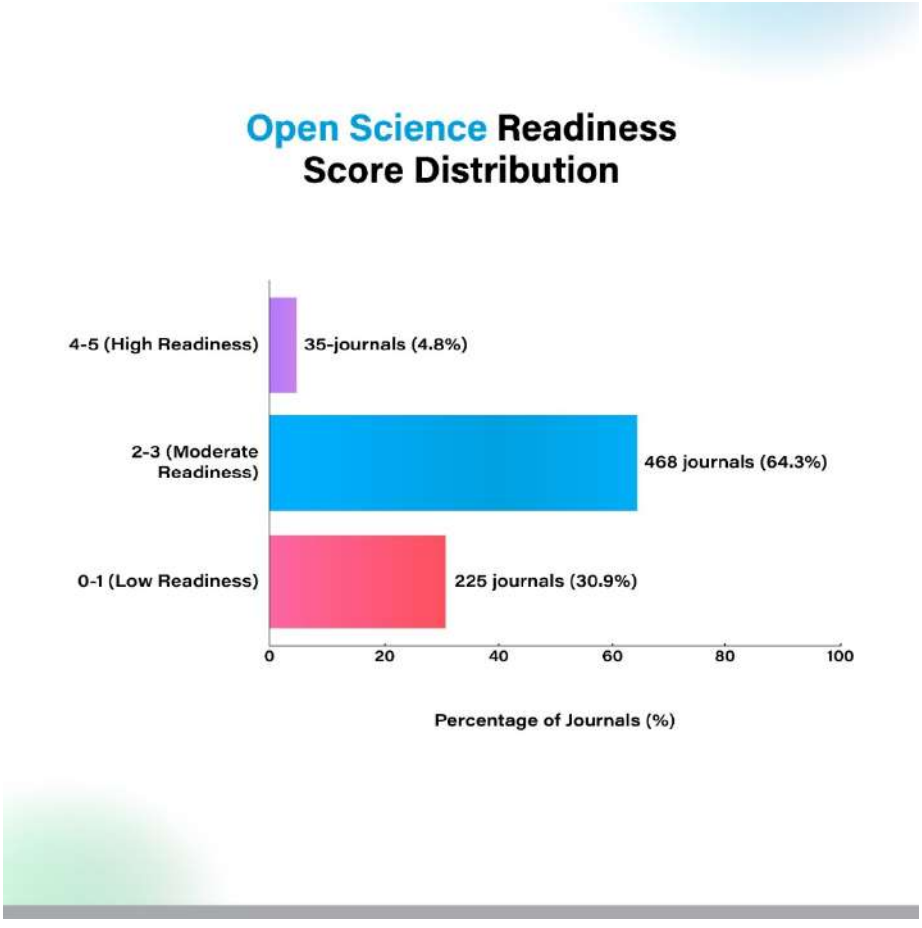


Fig. 8: Distribution of journals by open science readiness score

Open access and cost transparency: Open access was the dominant model, with over 96% of journals providing free access to content. However, cost transparency varied. Roughly 61% of OA journals disclosed article processing charges (APCs), while about 26% stated that no charges were levied. Notably, 12% provided no information on publication costs, reflecting a lack of transparency in financial policies shown in Fig. 5.

Metadata practices across disciplines: Figure 6 illustrates the proportion of Pakistani journals that require authors to provide an ORCID iD during submission. The chart shows that an overwhelming majority of journals (about 88.7%) do not require ORCID, while only 6.9% mandate it. A small fraction (4.4%) has missing or unclear information. This distribution demonstrates that ORCID adoption remains extremely limited across subject areas, indicating a major gap in author identification practices and alignment with global publishing standards.

Figure 7 presents a 100% stacked bar chart showing the availability of JATS/XML metadata across journals. The results reveal that 92.45% of journals do not provide XML metadata, while only 2.75% support it. Another 4.81% have unclear or unreported data. This indicates that structured, machine-readable metadata, the backbone of discoverability and interoperability, is almost absent across subject areas, highlighting one of the most critical infrastructural weaknesses in Pakistan's scholarly publishing system.

Open science readiness: A composite Open Science Readiness Score (0-5) was calculated for each journal based on the presence of DOI, ORCID, XML/JATS metadata, archiving service, and an online submission system. The results revealed that the vast majority of journals are not yet fully aligned with open science infrastructure:

- Low readiness (0-1): ~31%
- Moderate readiness (2-3): ~64%
- High readiness (4-5): ~5%

Only 0.4% of journals achieved the maximum score of 5, indicating comprehensive alignment with best practices shown in Fig. 8.

DISCUSSION

The analysis of 726 Pakistan-based scholarly journals offers a revealing snapshot of a publishing ecosystem that remains in the early stages of alignment with global open science and metadata standards. Despite substantial growth in journal output over the past two decades, key infrastructure elements essential to scholarly communication including persistent identifiers, structured metadata, archiving practices, and submission workflows remain inconsistently implemented across most titles. This fragmented landscape mirrors broader structural weaknesses in national research systems, where policy support for journals has historically lagged behind investment in research production itself[15,16].

A striking finding is the limited uptake of author identifiers: Only a small fraction of journals requires ORCID during submission, a figure far below adoption rates reported in leading publishing nations[17]. Because ORCID integration facilitates unambiguous author attribution, enhances discoverability, and supports interoperable metadata exchange[18], its absence significantly constrains the visibility of Pakistani research in global indexes. Similarly, DOI assignment, a foundational layer for citation linking, persistent access, and scholarly credit, remains uneven. Although some journals assign DOIs, their use is far from universal, echoing patterns seen in other low- and middle-income publishing contexts[19,20].

Metadata maturity is another area of concern. Only a small proportion of journals provide XML or JATS-formatted metadata, which is required by major indexing services and crucial for machine readability and text mining[21]. The absence of structured metadata not only impedes inclusion in services like PubMed Central or Scopus but also limits integration with emerging open science infrastructures. Likewise, less than half of journals report any archiving strategy, leaving long-term preservation uncertain and threatening the durability of the national scholarly record[22].

Although most journals identify as open access, transparency around article processing charges (APCs) remains inconsistent, with many failings to disclose fee policies clearly. This opacity undermines trust and complicates compliance with funder mandates[23]. The finding aligns with broader concerns about OA implementation quality in Global South contexts, where OA is often adopted as a label rather than a fully realized publishing model[24].

Geographic composition patterns further underscore structural imbalances: Most journals primarily serve national audiences, with limited international authorship or editorial participation. This insularity may contribute to low citation visibility and indexing challenges, reinforcing the marginalization of Pakistan's scholarship in global knowledge flows[25,26]. Yet, journals that exhibit more international participation tend to show stronger metadata practices and higher readiness scores, suggesting that globalization of editorial workflows may act as a catalyst for technical and procedural improvements.

Taken together, these findings highlight the need for coordinated national strategies that go beyond journal proliferation to emphasize quality, interoperability, and sustainability. Policy interventions, such as mandating DOI registration, incentivizing ORCID integration, supporting shared metadata infrastructure, and establishing national preservation services, could significantly accelerate Pakistan's alignment with international standards[27]. Without such structural reforms, many of the country's scholarly journals risk remaining invisible in global indexing systems, undermining both their credibility and the impact of the research they publish.

CONCLUSION

The analysis highlights a publishing ecosystem at a crossroads. Pakistan's journals have made significant strides in accessibility and digitization, with widespread adoption of online submission systems and open access models. However, deeper layers of infrastructure, including persistent identifiers, metadata interoperability, archiving, and author identification, remain underdeveloped. These gaps limit the visibility, discoverability, and long-term preservation of the nation's scholarly output and constrain its integration into the global research landscape.

Bridging these gaps requires coordinated action from journal editors, publishers, funding agencies, and regulatory bodies. Capacity-building initiatives focused on metadata standards such as JATS XML, persistent identifiers like DOIs and ORCID, and long-term preservation practices must be prioritized. Policymakers and national indexing agencies could also play a catalytic role by linking accreditation or funding eligibility to the adoption of these standards. Additionally, fostering international collaborations and diversifying editorial boards could help journals expand their geographic reach and align more closely with global best practices.

By addressing these structural and technical gaps, Pakistan's journals can significantly enhance their visibility, improve their chances of inclusion in global databases, and contribute more effectively to the international scientific conversation. The data presented here provide a baseline for monitoring progress over time and can inform national strategies aimed at strengthening the country's scholarly communication infrastructure.

SIGNIFICANCE STATEMENT

This study provides the first comprehensive assessment of the metadata infrastructure and open-science readiness of Pakistan's scholarly journals. By evaluating 726 journals across key indicators such as DOI adoption, ORCID integration, JATS/XML metadata, archiving practices, and submission systems, the study reveals critical infrastructural gaps that limit visibility, discoverability, and long-term preservation. The

findings offer an evidence-based foundation for national policy reforms, capacity-building initiatives, and editorial improvements aimed at aligning Pakistan's journals with global standards. Strengthening these foundational systems will enhance international interoperability, increase global research impact, and support the country's integration into the global scientific ecosystem.

REFERENCES

1. NISO, 2024. The role of metadata and persistent identifiers (PIDs) in open science. NISO.
2. Hillmann, D.I., R. Marker and C. Brady, 2008. Metadata standards and applications. Serials Librarian, 54: 7-21.
3. Subaveerapandiyan A, Amreen Taj, A.R. Nair 2026. Advancing open science: A bibliometric study of scholarly metadata research (1995-2024). Sci. Technol. Lib., 45: 100-129.
4. Musap, L.J., 2023. Enhancing scientific publishing: Automatic conversion to JATS XML. Eur. Sci. Ed., Vol. 49. 10.3897/ese.2023.e114977.
5. Porter, S.R., P.D. Umbach and C. Willis, 2025. Understanding ORCID adoption among academic researchers. Scientometrics, 130: 2783-2797.
6. Feeney, P., 2023. Making data citations machine readable in article and other content metadata. Inf. Serv. Use, 43: 323-326.
7. Lor, P., 2023. Scholarly publishing and peer review in the Global South: The role of the reviewer. JLIS.it, 14: 10-29.
8. Gregg, W.J., C. Erdmann, L.A.D. Paglione, J. Schneider and C. Dean, 2019. A literature review of scholarly communications metadata. Res. Ideas Outcomes, Vol. 5. 10.3897/rio.5.e38698.
9. Meadows, A., L.L. Haak and J. Brown, 2019. Persistent identifiers: The building blocks of the research information infrastructure. Insights, Vol. 32. 10.1629/uksg.457.
10. Vaughn, M. anad R. Higgins, 2025. Leveraging LLMs in library publishing: JATS XML encoding with ChatGPT. J. Librarianship Scholarly Commun., Vol. 13. 10.31274/jlsc.18048.
11. Papin-Ramcharan, J.I. and R.A. Dawe, 2006. Open access publishing: A developing country view. First Monday, Vol. 11. 10.5210/fm.v11i6.1332.
12. Matheka, D.M., J. Nderitu, D. Mutonga, M.I. Otiti, K. Siegel and A.R. Demaio, 2014. Open access: Academic publishing and its implications for knowledge equity in Kenya. Globalization Health, Vol. 10. 10.1186/1744-8603-10-26.
13. Haak, L.L., M. Fenner, L. Paglione, E. Pentz and H. Ratner, 2012. ORCID: A system to uniquely identify researchers. Learned Publ., 25: 259-264.
14. Dappert, A., A. Farquhar, R. Kotarski and K. Hewlett, 2017. Connecting the persistent identifier ecosystem: Building the technical and human infrastructure for open research. Data Sci. J., Vol. 16. 10.5334/dsj-2017-028.
15. Turki, H., G. Fraumann, M.A.H. Taieb and Mohamed Ben Aouicha, 2023. Global visibility of publications through digital object identifiers. Front. Res. Metrics Anal., Vol. 8. 10.3389/frma.2023.1207980.
16. Newton, C.R., 2020. Research and open access from low- and middle-income countries. Dev. Med. Child Neurol., 62: 537-537.
17. Porter, S.J., 2022. Measuring research information citizenship across ORCID practice. Front. Res. Metrics Anal., Vol. 7. 10.3389/frma.2022.779097.
18. Haak, L.L., A. Meadows and J. Brown, 2018. Using ORCID, DOI, and other open identifiers in research evaluation. Front. Res. Metrics Anal., Vol. 3. 10.3389/frma.2018.00028.
19. Gorraiz, J., D. Melero-Fuentes, C. Gumpenberger and J.C. Valderrama-Zurián, 2016. Availability of digital object identifiers (DOIs) in Web of Science and Scopus. J. Informetrics, 10: 98-109.
20. Meghanandha and U. Naik, 2025. A comparative review of metadata, communication, content, and digital preservation standards in modern libraries. Am. J. Inf. Sci. Technol., 9: 24-33.
21. Shi, J., M. Nason, M. Tullney and J.P. Alperin, 2025. Identifying metadata quality issues across cultures. Coll. Res. Lib., 86: 101-134.

22. Mosha, N.F. and P. Ngulube, 2023. Metadata standard for continuous preservation, discovery, and reuse of research data in repositories by higher education institutions: A systematic review. Information, Vol. 14. 10.3390/info14080427.
23. Kendall, G., 2025. More transparency is required for article processing charges. J. Acad. Librarianship, Vol. 51. 10.1016/j.acalib.2025.103107.
24. Onaolapo, S., P. Ayeni and S. Mncube, 2025. Open access publishing in an African context: Notable improvements and recurring challenges. IFLA J., 10.1177/03400352251351113.
25. Khanna, S., J. Ball, J.P. Alperin and J. Willinsky, 2022. Recalibrating the scope of scholarly publishing: A modest step in a vast decolonization process. Quant. Sci. Stud., 3: 912-930.
26. Bol, J.A., A. Sheffel, N. Zia and A. Meghani, 2023. How to address the geographical bias in academic publishing. BMJ Global Health, Vol. 8. 10.1136/bmjgh-2023-013111.
27. Mazzega, P., D.M. Rugmini and A.F. Barros-Platiau, 2025. Where is the “Global South” located in scientific research? Earth Syst. Governance, Vol. 25. 10.1016/j.esg.2025.100269.