

Open Alex as Open Scholarly Infrastructure: Evolution for Academic Libraries

Harihararao Mojjada

Assistant Librarian, Central Library, MVGR College of Engineering (Autonomous), Vizianagaram,
Andhra Pradesh, India

ORCID: <https://orcid.org/0000-0003-1737-6146>

Abstract

Purpose: This paper examines OpenAlex as an open scholarly infrastructure. It traces the genesis of OpenAlex as the successor to Microsoft Academic Graph (MAG), describes its evolution into a widely used open research-information source, and identifies its applications and opportunities for academic libraries.

Design/methodology/approach: The study adopts a qualitative, descriptive and conceptual design. It is based on a documentary analysis of peer-reviewed articles, preprints, policy declarations and official OpenAlex documentation published between 2010 and 2026, supplemented by a thematic reading of an open-science infographic on OpenAlex. The findings are synthesised into a five-layer conceptual framework.

Findings: OpenAlex has moved from an emergency replacement for MAG, launched in January 2022, to a central component of the open research-information movement endorsed by the Barcelona Declaration. Its entity model of works, authors, sources, institutions, topics and funders offers libraries a free and programmable alternative to subscription databases. Recent evaluations show reference coverage comparable to Web of Science and Scopus for recent literature, but persistent gaps in affiliations, abstracts, document types, language metadata and retraction flags. Eight application areas for libraries are identified, ranging from literature discovery to institutional research intelligence and metadata stewardship.

Research limitations/implications: The paper is conceptual and does not report primary empirical data. Because OpenAlex changes rapidly, some technical details may date quickly.

Practical implications: Academic libraries, particularly those in India with uneven access to commercial citation indexes, can use OpenAlex for bibliometric services, open-access monitoring and research-profile management, provided they follow a six-step responsible-use protocol of define, retrieve, clean, validate, analyse and report.

Originality/value: The paper offers an LIS-focused synthesis that connects the technical and evaluative literature on OpenAlex with the service design of academic libraries, and proposes a conceptual framework to guide adoption.

Keywords: OpenAlex; Open scholarly infrastructure; Open research information; Microsoft Academic Graph; Bibliometrics; Academic libraries; Research intelligence; Open science

1. Introduction

For more than half a century, the infrastructure that records who published what, where, and with what influence has been dominated by a small number of proprietary citation indexes. The Science Citation Index, launched in 1963, and later Web of Science, followed by Elsevier's Scopus in 2004, became the default evidence base for research evaluation, university rankings and library collection decisions (Borrego & Urbano, 2025; Mongeon & Paul-Hus, 2016). These databases are carefully curated, but their metadata are available only to subscribing institutions and cannot be freely reused, audited or redistributed. The Barcelona Declaration on Open Research Information argues that decisions about researchers' careers and institutional futures increasingly depend on such closed, black-box indicators, and commits its signatories to making openness of research information the default (Kramer et al., 2024).

This shift towards open research information is part of the wider open science agenda set out in the UNESCO Recommendation on Open Science (UNESCO, 2021) and in the FAIR principles, which call for data that are findable, accessible, interoperable and reusable (Wilkinson et al., 2016). It also reflects the Principles of Open Scholarly Infrastructure (POSI), which stress community governance, financial sustainability and insurance against capture by any single commercial actor (Bilder et al., 2020). Academic libraries sit at the centre of this transition. They pay for bibliographic databases, support researchers in finding and evaluating literature, advise on publishing and open access, and increasingly prepare research-performance data for institutional leadership, accreditation and rankings.

OpenAlex, an open index of the global research system, has emerged as the most prominent candidate for an open alternative to proprietary citation databases. This paper examines OpenAlex from the perspective of academic libraries: where it came from, how it has evolved, what it contains, how libraries can use it, and what cautions they should observe.

1.1 About OpenAlex

OpenAlex is a free, open and comprehensive catalogue of the world's scholarly works and of the entities connected to them, including authors, sources (journals, conferences and repositories), institutions, topics, publishers and funders (Priem et al., 2022). It was developed by OurResearch, the non-profit organisation that created Unpaywall, and is named after the ancient Library of Alexandria (Borrego & Urbano, 2025). In September 2025 the organisation itself adopted the OpenAlex name, bringing Unpaywall and the journal-package analysis tool Unsub under a single brand (Borrego & Urbano, 2025). Three features distinguish OpenAlex from subscription databases. First, its data are openly licensed and can be downloaded, reused and redistributed without a subscription. Second, its technical infrastructure and data model are publicly documented, which allows users to understand and scrutinise how records are built (Priem et al., 2022). Third, it offers several access routes: a web search interface, a REST API and a complete downloadable data snapshot. Its scale is considerable. As of November 2025, OpenAlex indexed about 271.3 million works, with a further expansion set of around 192 million lower-quality records drawn from DataCite and institutional repositories (Borrego & Urbano, 2025).

OpenAlex is sustained through philanthropic grants, public support and premium subscriptions that provide more frequent updates and higher API limits. In 2024 the French Ministry of Higher Education and Research announced a partnership to improve OpenAlex data, and OurResearch received a grant of

US\$7.5 million from the Arcadia Fund to establish the platform (Borrego & Urbano, 2025). The long-term sustainability of this model remains a live question in the literature (Borrego & Urbano, 2025; Cao et al., 2025).

1.2 Genesis of OpenAlex

The immediate origin of OpenAlex lies in the retirement of Microsoft Academic Graph. MAG was a large heterogeneous graph of publications, citations, authors, institutions, venues and fields of study, built largely through automated web crawling and machine learning, and were freely available to researchers (Wang et al., 2020). By the late 2010s it had become an important open alternative to Web of Science and Scopus in bibliometric research (Visser et al., 2021). In May 2021 Microsoft announced that MAG would be discontinued at the end of that year, leaving many research groups, tools and services without a data source (Borrego & Urbano, 2025; Chawla, 2022).

OurResearch responded by announcing a replacement. OpenAlex was built to be backward-compatible with MAG, released a beta data dump in November 2021, and launched formally in January 2022 (Chawla, 2022). A comparison of the final MAG snapshot with an early OpenAlex snapshot found that practically all MAG works were transferred to OpenAlex with their bibliographic data, publication years, DOIs and reference counts preserved, and that more than 90% of MAG documents had equivalent document types in OpenAlex (Scheidsteger & Haunschild, 2023). OpenAlex therefore began life as a continuation of MAG, but it quickly broadened its foundations by integrating open sources such as Crossref, ORCID, the Research Organization Registry (ROR), PubMed, arXiv, DOAJ, Unpaywall and, later, DataCite (Borrego & Urbano, 2025; Priem et al., 2022). Crossref, as the largest DOI registration agency and a community-owned source of scholarly metadata, is especially central to this model (Hendricks et al., 2020).

1.3 Evolution of OpenAlex

Since 2022 OpenAlex has evolved from a stop-gap for MAG users into a strategic component of open research-information policy. Table 1 summarises the principal milestones reported in the literature.

Table 1. Milestones in the genesis and evolution of OpenAlex

Period	Milestone	Source
2015–2020	Microsoft Academic Graph matures into a large, freely available bibliographic knowledge graph used widely in scientometrics	Wang et al. (2020); Visser et al. (2021)
May 2021	Microsoft announces the retirement of MAG at the end of 2021	Borrego & Urbano (2025)
Nov 2021 – Jan 2022	OurResearch releases a beta data dump; OpenAlex launches formally as the MAG successor	Chawla (2022)
2022	Data model, sources and entity structure described in the founding paper	Priem et al. (2022)
2023	Systematic comparisons with MAG confirm continuity of core bibliographic metadata	Scheidsteger & Haunschild (2023)

Period	Milestone	Source
2024	French Ministry partnership; Arcadia grant; Barcelona Declaration signed; CWTS Leiden Ranking Open Edition built on OpenAlex; Sorbonne University and CNRS announce moves away from proprietary databases	Borrego & Urbano (2025); Kramer et al. (2024); van Eck & Waltman (2024)
2024–2025	Large-scale evaluations of coverage, references, language and affiliations published; DataCite metadata integrated; topic-based classification in use	Culbert et al. (2025); Céspedes et al. (2025); Zhang et al. (2024); Borrego & Urbano (2025)
Sep 2025	Our Research renames itself OpenAlex, uniting Unpaywall and Unsub under one brand	Borrego & Urbano (2025)

Two further developments deserve mention. First, the thematic layer has changed. OpenAlex inherited MAG's field-of-study tagging as concepts, each linked to a Wikidata identifier (Scheidsteger & Haunschild, 2023). It now also assigns topics through an automated system that uses titles, abstracts, sources and references, arranged hierarchically into subfields, fields and domains (Borrego & Urbano, 2025). Second, institutional adoption has grown. The open edition of the CWTS Leiden Ranking is based on OpenAlex, using a published methodology to identify core sources and publications (van Eck & Waltman, 2024), and several French research organisations have adopted it as an alternative to Scopus and Web of Science (Borrego & Urbano, 2025).

1.4 Need and significance of the study

Most published work on OpenAlex comes from scientometrics and research-policy communities and focuses on data quality or coverage. Much less attention has been paid to how academic libraries, as the institutional custodians of bibliographic services, can build services around OpenAlex. This gap matters especially in countries such as India, where access to Scopus and Web of Science varies widely between institutions, and where libraries are increasingly asked to supply research-performance evidence for accreditation, ranking and funding purposes. An open, programmable and free source could narrow this gap, but only if library professionals understand both its capabilities and its limits.

1.5 Objectives of the study

The study pursues the following objectives:

- (i) To trace the genesis of OpenAlex from Microsoft Academic Graph and its evolution into open scholarly infrastructure.
- (ii) To describe the OpenAlex ecosystem of entities, relationships, access channels and tools.
- (iii) To review the recent literature on the coverage, metadata quality and use of OpenAlex.
- (iv) To identify the applications of OpenAlex in academic library services.
- (v) To propose a conceptual framework and a responsible-use protocol for the adoption of OpenAlex by academic libraries.
- (vi) To outline the opportunities and challenges that OpenAlex presents for academic libraries, with particular reference to the Indian context.

2. Review of Literature

The literature on OpenAlex has grown rapidly since 2022. Early studies documented its origins and data model; later studies tested its coverage and metadata quality against established databases; and the most recent work examines its suitability for research evaluation and its practical tooling. Ten studies that together represent these strands are reviewed below in broadly chronological order.

Wang et al. (2020) Wang and colleagues described the design of Microsoft Academic Graph, the predecessor of OpenAlex. They explained how MAG combined web-scale crawling with machine-learning methods for entity recognition, disambiguation and field-of-study tagging, producing a heterogeneous graph of papers, authors, institutions, venues and topics. Their account is important for understanding OpenAlex because the latter inherited MAG's graph architecture, its identifiers and much of its historical data. The study also shows the risk of relying on a corporate-owned free resource: the infrastructure could be, and was, withdrawn. **Visser et al. (2021)** Visser, van Eck and Waltman compared Scopus, Web of Science, Dimensions, Crossref and Microsoft Academic at scale. They found that Microsoft Academic had the broadest coverage of documents, but that its metadata were less complete and less accurate in several respects than those of curated databases. The study established the trade-off between breadth and curation that has since framed almost every evaluation of OpenAlex, and it demonstrated that open sources could compete with proprietary ones on coverage. **Priem et al. (2022)** In the founding paper of OpenAlex, Priem, Piwowar and Orr introduced it as a fully open index of scholarly works, authors, venues, institutions and concepts. They set out its data sources, entity model, open licensing, API and snapshot, and positioned it explicitly as a replacement for MAG and an alternative to proprietary indexes. This paper remains the primary reference for the design principles of OpenAlex and is the starting point for later evaluations. **Scheidsteger and Haunschild (2023)** Scheidsteger and Haunschild compared metadata relevant to bibliometrics in the final MAG snapshot with those in an early OpenAlex snapshot. They found that practically all MAG works were carried over with their core bibliographic data, DOIs and reference counts intact, and that more than 90% of documents had equivalent document types. Their analysis of subject classification showed that MAG's fields of study survived as OpenAlex concepts. The study provided early assurance that OpenAlex could support continuity for MAG-based research. **Delgado-Quirós and Ortega (2024)** Delgado-Quirós and Ortega assessed the completeness of publication metadata across eight free-access scholarly databases, including OpenAlex. Their comparison of fields such as abstracts, affiliations, references and identifiers showed that completeness varies considerably between databases and between fields. The study is useful for librarians because it places OpenAlex within the wider landscape of free bibliographic tools rather than comparing it only with commercial databases. **Zhang et al. (2024)** Zhang, Cao, Shang, Sivertsen and Huang examined missing institutional affiliations in OpenAlex. They reported that a large share of records lack institutional information entirely or partially, with the problem concentrated in older publications and in the social sciences and humanities. They discussed possible causes in source metadata and processing, and proposed solutions. The study has direct implications for any library using OpenAlex to measure institutional output, because missing affiliations lead to undercounting. **Culbert et al. (2025)** Culbert and colleagues conducted a large-scale comparison of reference coverage in OpenAlex, Web of Science and Scopus, using a cleaned set of 16.8 million recent publications shared by all three databases. They showed that OpenAlex had average source-reference numbers and internal coverage rates comparable to both commercial databases. Journal-level analysis, however, revealed

mixed results: OpenAlex captured more ORCID identifiers, fewer abstracts and a similar number of open-access indicators. This is one of the strongest pieces of evidence that OpenAlex is suitable for citation analysis of recent literature. **Céspedes et al. (2025)** Céspedes and colleagues evaluated the linguistic coverage of OpenAlex against Web of Science and manually validated a sample of 6,836 articles. OpenAlex offered a far more balanced coverage of languages than Web of Science, but its language metadata were not always accurate, leading it to overestimate English and underestimate other languages. The authors concluded that, used critically, OpenAlex can support representative analyses of multilingual publishing. This finding is significant for India, where scholarship appears in English, Hindi, Telugu and many other languages. **Aria et al. (2024)** Aria, Le, Cuccurullo, Belfiore and Choe presented openalexR, an R package that interfaces with the OpenAlex API. The package supports querying of works, authors, institutions, sources and concepts with filtering, searching, sorting and grouping, and formats results for use in bibliometrix (Aria & Cuccurullo, 2017). The paper illustrates how OpenAlex's open API lowers the technical threshold for reproducible bibliometric analysis, a point of direct relevance to library-based research-support services. **Borrego and Urbano (2025)** Borrego and Urbano analysed the features, sources, entities, advantages and limitations of OpenAlex from a library and information science perspective. They documented records lacking abstracts, affiliations and references, problems with document types and languages, duplicate author profiles, incorrect affiliations and version-control issues. They also described community curation initiatives such as the French Works-magnet project, in which librarians and research managers review and validate metadata. They recommended greater publisher commitment to complete metadata, coordination structures for institutional users, and more resources for quality control.

2.1. Synthesis and research gap

Three conclusions emerge from the review. First, OpenAlex has credible continuity with MAG and coverage of recent literature that rivals commercial databases (Culbert et al., 2025; Scheidsteger & Haunschild, 2023). Second, its metadata quality is uneven, particularly for affiliations, abstracts, document types, languages and citation versions (Borrego & Urbano, 2025; Céspedes et al., 2025; Mongeon et al., 2025; Zhang et al., 2024). Studies of its suitability for research-quality evaluation reach cautiously positive conclusions for most fields (Thelwall & Jiang, 2025), while field-normalised indicators agree with commercial databases at article level but diverge when broken down by university and subject (Scheidsteger et al., 2025). Third, an ecosystem of open tools has grown around the API (Aria et al., 2024). What is largely missing is a library-centred account that translates these findings into service models, skills and safeguards. The present study addresses that gap.

3. Methodology

The study follows a qualitative, descriptive and conceptual research design. It does not collect primary data; instead, it analyses and synthesises existing documentary evidence to build a conceptual framework for library adoption of OpenAlex.

3.1 Sources of evidence

Four categories of sources were consulted: (a) peer-reviewed journal articles in scientometrics and library and information science; (b) preprints deposited in arXiv, where much of the most current OpenAlex research first appears; (c) policy documents and declarations, namely the Barcelona Declaration, the UNESCO Recommendation on Open Science, the FAIR principles and the POSI

principles; and (d) official OpenAlex documentation and announcements, accessed through the literature that reports them.

3.2 Search and selection

Relevant literature was identified by searching OpenAlex itself, Google Scholar, arXiv and publisher platforms using the terms "OpenAlex", "Microsoft Academic Graph", "open research information", "open bibliographic data" and "open scholarly infrastructure", singly and in combination. Studies were included if they (i) described the origins, structure or development of OpenAlex or its predecessor, (ii) evaluated its coverage or metadata quality, (iii) described tools built on it, or (iv) set out policy principles for open research information. Only works with a verifiable DOI were retained, so that every reference in this paper can be independently checked.

3.3. Analysis

The selected documents were read closely and coded thematically under the headings of genesis, evolution, data model, access, tools, applications, limitations and good practice. The themes were then mapped onto the functions of academic libraries to produce the conceptual framework presented in Section 4. Because OpenAlex is updated frequently, the analysis reflects the state of the literature up to September 2026.

4. Conceptual Framework

The conceptual framework (Figure 1) represents OpenAlex as a layered infrastructure through which open data sources become library services and, ultimately, a more open research-information ecosystem. Five layers are linked vertically, and two cross-cutting pillars, governance and responsible use, frame the whole system.

4.1 Layer 1: Open data sources

OpenAlex does not create metadata from scratch. It aggregates, links and enriches metadata from open sources: the legacy MAG corpus, Crossref and DataCite for DOI-registered content, ORCID for researcher identity, ROR for institutional identity, and PubMed, arXiv, DOAJ, Unpaywall and institutional repositories for further works and open-access status (Borrego & Urbano, 2025; Priem et al., 2022). The quality of OpenAlex is therefore bounded by the quality of these upstream sources. Crossref, for example, does not clean the metadata deposited by publishers, and fabricated or incomplete reference metadata can propagate into downstream databases (Besançon et al., 2024; Hendricks et al., 2020). Libraries, as depositors of repository metadata and as users of ORCID and ROR, are themselves part of this supply chain.

4.2 Layer 2: The OpenAlex ecosystem of entities

At its core OpenAlex is a knowledge graph that connects key entities of the research system and maps the relationships between them. Table 2 describes the principal entities and their relevance to library work.

Table 2. Core OpenAlex entities and their relevance to academic libraries

Entity	Description	Relevance to libraries
Works	Articles, books, chapters, preprints,	Discovery, citation analysis, open-

Entity	Description	Relevance to libraries
	datasets, theses and other scholarly outputs	access status checking, repository gap analysis
Authors	Author profiles created through automated disambiguation, linked to ORCID where available	Faculty profiles, author-identity services, expertise directories
Sources	Journals, conference series, repositories and other venues	Journal evaluation, collection development, publishing advice
Institutions	Research organisations, linked to ROR identifiers	Institutional output reports, benchmarking, collaboration analysis
Topics / Concepts	Hierarchical subject classification (topics, subfields, fields, domains); earlier concept tags inherited from MAG	Research landscape mapping, subject-based collection planning
Funders	Funding organisations and grant information	Funder-compliance monitoring, open-access mandate support
Relationships	Links between works, authors, institutions, funders, sources and topics	Network analysis of co-authorship, citation and institutional collaboration

The value of the graph lies less in any single entity than in the relationships between them. Because entities carry persistent identifiers such as DOIs, ORCID iDs and ROR IDs, OpenAlex data can be joined with local library systems, institutional repositories and research-information systems, which is the basis for most of the applications discussed below.

4.3 Why OpenAlex matters

Six attributes explain why OpenAlex matters for the library profession. It is **open access**, free to use with no paywalls, which removes the subscription barrier that excludes many institutions from citation data. It supports **greater inclusion**, since its broader and more multilingual coverage improves the visibility of research from the Global South and from non-English publishing (Céspedes et al., 2025). It is **transparent**, with an openly documented methodology and data model that can be scrutinised (Priem et al., 2022). It is **programmable**, offering an API and bulk data that support automation and reproducibility (Aria et al., 2024). It **supports innovation**, because anyone can build new tools, analyses and services on top of it. Finally, it is **aligned with open science** as expressed in the UNESCO Recommendation, the FAIR principles and the Barcelona Declaration (Kramer et al., 2024; UNESCO, 2021; Wilkinson et al., 2016).

4.4 Layer 3: Access and use channels

OpenAlex can be reached through several channels suited to different levels of technical skill. The web interface at openalex.org supports searching and filtering without programming. The REST API, documented at docs.openalex.org, allows programmatic retrieval of entities with filtering, sorting and grouping; recent changes have introduced API keys with a free daily usage allowance. A complete data

snapshot can be downloaded for local analysis at scale. The source code and related projects are published on GitHub, and a global community of users contributes corrections, tools and discussion. For libraries this layered access means that a reference librarian, a systems librarian and a research-data specialist can each engage with OpenAlex at an appropriate level.

4.5 Layer 4: Tools and integrations

OpenAlex works with many analytical tools and programming languages. R users can use `openalexR` to query the API and pass results to `bibliometrix` for science mapping (Aria et al., 2024; Aria & Cuccurullo, 2017), and Python clients offer similar functionality. Network and visualisation tools such as VOSviewer (van Eck & Waltman, 2010) and Gephi, and business-intelligence tools such as Tableau and Power BI, can turn OpenAlex data into maps, networks, dashboards and reports. The typical workflow therefore runs from OpenAlex, through the API or bulk data, into R or Python, then into analysis tools, and finally into visualisations and insights. Libraries that already produce usage dashboards can extend the same skills to research-output dashboards.

4.6 Layer 5: Applications of OpenAlex in the library field

The eight categories of use highlighted in the infographic correspond closely to established and emerging library functions. Table 3 maps each to concrete library services.

Table 3. Applications of OpenAlex mapped to academic library services

Use of OpenAlex	What it involves	Library service application
Literature discovery	Searching and exploring scholarly outputs across disciplines	Reference and research consultation; supplement to discovery layers; locating open-access versions through Unpaywall-derived status
Systematic and narrative reviews	Identifying and retrieving relevant literature	Evidence-synthesis support; reproducible, documented search strategies; citation chasing through reference links
Bibliometric analysis	Productivity, citation and impact analysis	Departmental and institutional publication reports; data for accreditation and ranking submissions; author-level metrics advice
Collaboration analysis	Co-authorship and institutional networks	Mapping national and international partnerships; identifying potential collaborators for grant applications
Research mapping	Visualising research landscapes and trends	Topic trend reports to inform collection development, curriculum planning and new-programme proposals
Institutional intelligence	Research profiles, benchmarking and monitoring	Faculty expertise directories; peer-institution benchmarking; monitoring open-access share of institutional output
Science and society studies	Studying research output, collaboration and inclusion	Analysing language diversity, gender and regional visibility; supporting LIS research by faculty and

Use of OpenAlex	What it involves	Library service application
		doctoral scholars
Building innovative tools	Developing dashboards, services and new analyses	Research-output dashboards; repository enrichment and gap detection; alerting services for new faculty publications

Beyond these eight uses, the literature suggests four further library applications. The first is **collection and serials assessment**: OpenAlex data on where institutional authors publish and what they cite can inform subscription decisions, a use closely related to the Unsub tool now under the OpenAlex brand (Borrego & Urbano, 2025). The second is **open-access monitoring**, building on the long-standing work of OurResearch on the prevalence of open access (Piwowar et al., 2018). The third is **repository and research-information management**, where OpenAlex records can be harvested to identify faculty publications missing from an institutional repository. The fourth is **metadata stewardship**: librarians can correct author profiles and affiliations, register their institutions in ROR and encourage faculty to maintain ORCID records, contributing to the community curation described by Borrego and Urbano (2025).

4.7 Points to consider: limitations of OpenAlex

Like any large-scale data source, OpenAlex has limitations, and users should validate data, especially for high-stakes analyses. Table 4 summarises the main limitations reported in the literature and suggests how libraries can mitigate them.

Table 4. Limitations of OpenAlex and library mitigation strategies

Limitation	Evidence	Library mitigation
Affiliation data	Many records lack affiliations; some affiliations are incorrectly assigned (Zhang et al., 2024; Borrego & Urbano, 2025)	Supplement with local faculty lists and ORCID; manually verify institutional sets before reporting
Document types	Misclassification of editorials and other items as articles (Mongeon et al., 2025)	Filter and sample-check document types; report the classification used
Language identification	Language metadata can overestimate English (Céspedes et al., 2025)	Validate language for regional-language journals before analysis
Citations and references	References to non-indexed documents are lost; citation counts can differ between versions (Borrego & Urbano, 2025; Culbert et al., 2025)	Use citation counts comparatively rather than absolutely; triangulate with other sources
Retractions	Not all retracted records are flagged	Cross-check key works against retraction databases before citing in evaluation reports

Limitation	Evidence	Library mitigation
Abstracts	About 40% of standard records lack abstracts, partly because of publisher restrictions (Borrego & Urbano, 2025)	Avoid abstract-dependent text mining without checking coverage
Content quality	Inclusive indexing means questionable or predatory journals may be present; source metadata can be manipulated (Besançon et al., 2024)	Teach source evaluation; combine OpenAlex with journal-quality checks

4.8 Good practices for responsible use

The Leiden Manifesto holds that quantitative evaluation should support, not replace, expert judgement and that data-collection and analytical processes should be open, transparent and simple (Hicks et al., 2015). Applied to OpenAlex, this principle translates into a six-step protocol that libraries can adopt as a standard operating procedure:

- **Define:** state the research or evaluation question clearly, including the unit of analysis and period.
- **Retrieve:** document the query, filters, API parameters and date of data extraction.
- **Clean:** deduplicate and normalise records, particularly author names and affiliations.
- **Validate:** check critical records manually and triangulate with other sources where feasible.
- **Analyse:** use methods and indicators appropriate to the question and discipline.
- **Report:** be transparent about data sources, methods, filters and limitations.

Responsible use of this kind leads to more reliable insights and stronger research outcomes, and it gives libraries a defensible basis for the reports they supply to institutional leadership.

4.9 Opportunities for academic libraries

OpenAlex opens five significant opportunities for academic libraries. First, it offers **cost-effective access** to research information for institutions that cannot afford commercial citation indexes, which is especially relevant for many state universities and affiliated colleges in India. Second, it enables libraries to **reposition themselves as research-intelligence units**, producing publication analyses, collaboration maps and open-access reports that were previously outsourced or unavailable. Third, it creates a **metadata stewardship role**, in which librarians improve the global record by curating affiliations, identifiers and repository metadata, as the French Works-magnet initiative demonstrates (Borrego & Urbano, 2025). Fourth, it calls for **continuing professional development**: API use, basic R or Python, data cleaning and responsible metrics are now practical competencies for LIS professionals, and library schools, professional associations and faculty development centres can build them into training programmes. Fifth, it gives libraries a platform for **policy advocacy**, encouraging their institutions to align with the Barcelona Declaration and to value open, auditable research information in internal evaluation.

These opportunities come with challenges. Data-quality gaps require time and skills to address, the sustainability of OpenAlex depends on continuing funding (Cao et al., 2025), and heavy reliance on any

single infrastructure, open or not, carries risk. Libraries should therefore treat OpenAlex as one component of a diversified open research-information strategy that also includes Crossref, OpenCitations, ORCID, ROR and national systems.

4.10 Towards a more open future

OpenAlex helps break down barriers to knowledge, supports research from around the world, and enables new opportunities for discovery, collaboration and innovation. Its trajectory, from an emergency replacement for MAG to a platform underpinning an open university ranking and national open-science policies, shows how quickly open infrastructure can gain legitimacy when research communities choose to support it. Academic libraries, as long-standing advocates of open access, are well placed to carry this movement forward by using, improving and advocating for open research information.

5. Conclusion

This paper has examined OpenAlex as open scholarly infrastructure from a library and information science perspective. OpenAlex originated as the successor to Microsoft Academic Graph, launched in January 2022, and has since evolved into a central resource of the open research-information movement, supported by public funding, philanthropic grants and policy commitments such as the Barcelona Declaration. Its entity graph of works, authors, sources, institutions, topics and funders, accessible through a web interface, an API and bulk data, gives academic libraries a free and programmable foundation for services that were previously dependent on costly subscriptions.

The literature shows that OpenAlex offers coverage of recent literature comparable to Web of Science and Scopus and more balanced linguistic coverage, but that it still suffers from gaps and errors in affiliations, abstracts, document types, language metadata and citation versions. The proposed five-layer conceptual framework links open data sources, the core entity graph, access channels, tools and library applications, framed by governance principles and a six-step responsible-use protocol. Within this framework, libraries can support literature discovery, reviews, bibliometric analysis, collaboration and research mapping, institutional intelligence, open-access monitoring and metadata stewardship.

For Indian academic libraries in particular, OpenAlex offers a practical route to research-information services that are open, inclusive and affordable. Realising this potential will require investment in the data skills of library professionals, institutional policies that value open data, and active participation in improving the metadata on which OpenAlex depends. The main limitation of this study is its conceptual design. Future research should test the framework empirically, for example by comparing OpenAlex-based institutional reports with those derived from commercial databases for Indian universities, by assessing the coverage of Indian regional-language journals, and by surveying the competencies and training needs of library professionals for working with open research information.

References

1. Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
2. Aria, M., Le, T., Cuccurullo, C., Belfiore, A., & Choe, J. (2024). openalexR: An R-tool for collecting bibliometric data from OpenAlex. *The R Journal*, 15(4), 167–180. <https://doi.org/10.32614/RJ-2023-089>

3. Besançon, L., Cabanac, G., Labbé, C., & Magazinov, A. (2024). Sneaked references: Fabricated reference metadata distort citation counts. *Journal of the Association for Information Science and Technology*, 75(12), 1368–1379. <https://doi.org/10.1002/asi.24896>
4. Bilder, G., Lin, J., & Neylon, C. (2020). The principles of open scholarly infrastructure. <https://doi.org/10.24343/C34W2H>
5. Borrego, Á., & Urbano, C. (2025). OpenAlex: Features, advantages and limitations of an open database for retrieving and analysing scholarly outputs [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2512.16434>
6. Cao, Z., Zhang, L., Huang, Y., & Haunschild, R. (2025). How does the academia refer to open research information data sources? A review study based on OpenAlex and Microsoft Academic series. *Scientometrics*. <https://doi.org/10.1007/s11192-025-05347-6>
7. Céspedes, L., Kozłowski, D., Pradier, C., Sainte-Marie, M. H., Shokida, N. S., Benz, P., Poitras, C., Ninkov, A. B., Ebrahimi, S., Ayeni, P., Filali, S., Li, B., & Larivière, V. (2025). Evaluating the linguistic coverage of OpenAlex: An assessment of metadata accuracy and completeness. *Journal of the Association for Information Science and Technology*, 76(6), 884–895. <https://doi.org/10.1002/asi.24979>
8. Chawla, D. S. (2022). Massive open index of scholarly papers launches. *Nature*. <https://doi.org/10.1038/d41586-022-00138-y>
9. Culbert, J. H., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., & Mayr, P. (2025). Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. *Scientometrics*, 130(4), 2475–2492. <https://doi.org/10.1007/s11192-025-05293-3>
10. Delgado-Quirós, L., & Ortega, J. L. (2024). Completeness degree of publication metadata in eight free-access scholarly databases. *Quantitative Science Studies*, 5(1), 31–49. https://doi.org/10.1162/qss_a_00286
11. Hendricks, G., Tkaczyk, D., Lin, J., & Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies*, 1(1), 414–427. https://doi.org/10.1162/qss_a_00022
12. Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431. <https://doi.org/10.1038/520429a>
13. Kramer, B., Neylon, C., & Waltman, L. (2024). Barcelona Declaration on Open Research Information. Zenodo. <https://doi.org/10.5281/zenodo.10958522>
14. Mongeon, P., Hare, M., Riddle, P., Wilson, S., Krause, G., Marjoram, R., & Toupin, R. (2025). Investigating document type, language, publication year, and author count discrepancies between OpenAlex and Web of Science [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2508.18620>
15. Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
16. Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of open access articles. *PeerJ*, 6, e4375. <https://doi.org/10.7717/peerj.4375>

17. Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts [Preprint]. arXiv.
<https://doi.org/10.48550/arXiv.2205.01833>
18. Scheidsteger, T., & Haunschild, R. (2023). Which of the metadata with relevance for bibliometrics are the same and which are different when switching from Microsoft Academic Graph to OpenAlex? *Profesional de la información*, 32(2), e320209.
<https://doi.org/10.3145/epi.2023.mar.09>
19. Scheidsteger, T., Haunschild, R., & Bornmann, L. (2025). How similar are field-normalized citation impact scores obtained from OpenAlex and three popular commercial databases? An empirical comparison based on large German universities. *Scientometrics*, 130(7), 3537–3569.
<https://doi.org/10.1007/s11192-025-05338-7>
20. Thelwall, M., & Jiang, X. (2025). Is OpenAlex suitable for research quality evaluation and which citation indicator is best? *Journal of the Association for Information Science and Technology*, 76(12), 1660–1681. <https://doi.org/10.1002/asi.70020>
21. UNESCO. (2021). UNESCO Recommendation on Open Science. UNESCO.
<https://doi.org/10.54677/MNMH8546>
22. van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
23. van Eck, N. J., & Waltman, L. (2024). A methodology for identifying core sources and core publications in OpenAlex. Zenodo. <https://doi.org/10.5281/zenodo.13879947>
24. Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2(1), 20–41. https://doi.org/10.1162/qss_a_00112
25. Wang, K., Shen, Z., Huang, C., Wu, C.-H., Dong, Y., & Kanakia, A. (2020). Microsoft Academic Graph: When experts are not enough. *Quantitative Science Studies*, 1(1), 396–413. https://doi.org/10.1162/qss_a_00021
26. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
27. Zhang, L., Cao, Z., Shang, Y., Sivertsen, G., & Huang, Y. (2024). Missing institutions in OpenAlex: Possible reasons, implications, and solutions. *Scientometrics*, 129(10), 5869–5891. <https://doi.org/10.1007/s11192-023-04923-y>