RBU Journal of Library & Information Science, Vol. 27, 2025

ISSN: 0972-2750

RBU JOURNAL OF Library & Information Science

Volume 27, 2025

Blind Peer Reviewed

Previously in UGC-CARE
(Upto 10th February 2025)

DEPARTMENT OF LIBRARY & INFORMATION SCIENCE
Rabindra Bharati University
Kolkata

RBU Journal of Library & Information Science

ISSN: 0972–2750

(Official Organ of the Department of Library & Information Science, Rabindra Bharati University)

Volume 27, 2025

Online Content of Previous Volumes Available at:
https://lisrbu.wixsite.com/dlis/rbu-journal-of-lis

UGC-CARE **(Till 10th February 2025)**

Blind Peer Reviewed

RBU JOURNAL OF LIBRARY & INFORMATION SCIENCE

VOLUME 27, 2025

ISSN: 0972-2750

[Official organ of the Department of Library & Information Science]

Editor

Prof. Sudip Ranjan Hatua

Department of Library & Information Science

Kolkata, India

Visit: http://rbu.ac.in/home/page/102

About the Journal

The RBU Journal of Library & Information science is a scholarly communication for education, research and development of the Library & Information science field. It is published annually in print format only. This publication is fully funded by Rabindra Bharati University; therefore, it never asked any kind of charges or publication fees or donations from the author. The first volume was published in the year 1997. It received ISSN (0972-2750) in its 5th volume in the year 2001. From its 17th Volume published in the year 2015 the journal becomes peer-reviewed (follows blind peer review process) by the eminent experts across the country. This Journal was enlisted under *UGC List of Journals No. 45237, Sl. No. 2023 when UGC published a list of research journals published across the country in its website. Later this journal enlisted under UGC-CARE List w.e.f. 14.6.2019 to as long as the UGC maintain it (10th February 2025).* Present publication is its 27th Volume published in the year 2025.

The RBU Journal of Library and Information Science was published under following Editors-

- *Professor Pinakinath Mukhopadhyay: Vol.1 to Vol. 11*
- *Shri Salil Chandra Khan: Vol. 12 to Vol. 14*
- *Dr Sudip Ranjan Hatua: Vol. 15 to Vol. 25*
- *Dr. Ziaur Rahman: Vol 26*
- *Prof. Sudip Ranjan Hatua: Vol 27*

For details please visit: http://rbu.ac.in/home/page/102
To view the content of the previous volumes and the full text e-copy of VOLUME 21, 24, 25 please visit:
https://lisrbu.wixsite.com/dlis/rbu-journal-of-lis

UGC-CARE (Till 10th February 2025)

RBU JOURNAL OF LIBRARY & INFORMATION SCIENCE

(Blind Peer Reviewed)

Volume 27, 2025; ISSN: 0972-2750
(http://rbu.ac.in/home/page/102)

For details along with previous volumes' content
(Full text of Volume 21, 24 & 25)
visit: https://lisrbu.wixsite.com/dlis/rbu-journal-of-lis

Author's Guide

How to Send an Article?

Submission Guidelines

The RBU Journal of Library and Information Science publishes original research findings, review articles, practice outcome and survey results, integrate and critically examine new information accumulated in recent years in a particular subject field and specifically the following categories in order to cater to the diverse needs of its readership.

1. Research papers may describe completed research efforts with results, analyses and implications to practitioners, and should not exceed 15-20 pages in length.

2. Practice papers may describe new industry practices, tools and methodologies, with emphases on practical issues, problems and solutions, and should not exceed 10-15 pages in length. Papers may also discuss the relevance of theory to practice and applications.

3. Current Trends papers will reflect new or current trends, thinking, perspectives and opinions in research and practice, and should not exceed 8-12 pages in length.

4. Review papers should give a critical and analytical perspective of related books, publications, methodologies, practices, tools or systems, and should not exceed 5-6 pages in length.

Minimum standards for considering the submitted manuscripts for peer-reviewing Papers must be written in English Language. Text files should be prepared in MS Word format double line space.

Page setup: Page size A4, orientation portrait; Margins: mirror margins, top 3 cm, bottom 2 cm, inside 2.5 cm, outside 2 cm, gutter 0.5 cm;

Layout: header 1.7 cm (different odd and even), footer 1 cm, section start continuous, page alignment top; Font - Times New Roman, 12 points. First Heading -14 bold Times New Roman Second subheading -12 bold Times New Roman, Each next subheading- 12 italics Times New Roman.

Preparation of the manuscript

All original research articles should be structured in the following manner.

• Title: The title should be concise and reflect the entire work of the submitted manuscript. (Written in 14pt bold Times New Roman font size)

• Names of all authors need to be indicated below the title (Written in 12 bold Times New Roman font size)

• More than one author should display side by side with a tab space (5)

• Department/Affiliation: bellow of each author, 10pt Times New Roman, centered University/Institution: 10pt Times New Roman, centered Address: 10pt Times New Roman, centered Email: account@xxxxxx.xx (10pt Times New Roman, centered).

Abstract

iii

Abstracts should clearly state and represent in the following format Introduction-2-3lines Purpose-1-2 lines Research problem-1-2 lines Objective-1-2 lines Methodology-1-2 lines Findings-2-3 lines and NOT more than 500 words. (Written in 12pt italics Times New Roman font size).

Keywords

The author should provide 3 to 6 key words, characterizing the scope of the paper, the main plant material used and the central aspect of the work. Keywords should be presented below the (Written in 10pt bold Times New Roman font size).

Layout of article

Layout of article should maintain minimum following pattern:

• Introduction: State the background and mention clearly the objective of the present work.

• Literature Review: minimum last ten years

• Problem Identification: Research question and draw hypothesis

• Methodology: All methods used should be clearly mentioned

• Data collection and analysis: How and which method followed? How it has been analyzed?

• Findings: What specific finding have occurred?

• Conclusion: Give the major conclusion from the present study.

Acknowledgement: Acknowledge those persons who helped you in the present study by providing facilities, personal assistance and funding if any.

In text Citation

In text citation is mandatory. It should follow APA style only using surnames (year) approach.

For using Table Chart and Diagram

For each table text or data should use Arial font in 8 size. If a data table is good enough to represent the concept unnecessary pie, bar or any other diagram need not to be use. The caption of the table should be placed under the table in center using Table- No : title; format with Arial 9 font italics.

For chart use the same pattern and use Chart- No: tile ; and for figure user Figure- No.: title of the figure.

References & Footnotes

References to already published literature should be numbered consecutively in the text and placed within square brackets. Please adopt correct referencing methods. Papers with incorrect referencing and in-text citation are likely to be rejected.

The citations should be placed at the end of the paper in the sequence as they appear in the text. References to personal communication and unpublished literature should not be placed under references, but should be cited in the text in parentheses. Explanatory material should be given in the appendix. Examples of citations to different types of documents are given below:

(i) Journal Article

1. Gosh, B.K. (2004) Knowledge management policies options. RBU Journal of Library andInformation Science. 41(3): 145–150.

2. Neelameghan, A. & Gopinath, M. A. (1967). Research in library classification.LibraryScience with a Slant to Documentation. 4(2): 356–38

(ii) Book/Monograph

1. Ranganathan, S R. (1957) The Five Laws of Library Science. 2nd ed. Mumbai: Asia Publishing House, 456p.

(iii) Chapter from a Book

1. Neelameghan, A. & Raghavan, K.S. (2012). Frames of knowledge: a perspective of Vedic-Hinduism and Dravidian culture. In: Cultural frames of knowledge, edited by Richard, P Smiraglia & Hur-li Lee. Wursburg, Germany, 2012, 19–61.

(iv) Conference Paper

1. Ragahavan, K.S. & Neelameghan, A. Indic cultures and concepts: Implications for knowledge organization. In 12th International ISKO Conference , 6–9 August 2012, Mysore, India, edited by A. Neelameghan & K.S. Raghavan, 2012, pp. 176–182.

(v) Conference paper (online)

Cannan, J. (2008). Using practice based learning at a dual-sector tertiary institution: A discussion of current practice. In R. K. Coll, & K. Hoskyn (Eds.), Working together: Putting the cooperative into cooperative education. Conference proceedings of the New Zealand Association for Cooperative Education, New Plymouth, New Zealand. Retrieved from

http://www.nzace.ac.nz/conferences/papers/Proceedings_2008.pdf

(v) Report

1. Birkler, John; Smith, Giles; Kent, Gleen A. & Johns on, Robert V. (2000) An acquisition strategy, process, and organisation for innovative systems. National Defence Research Institute, RAND, USA, 2000. RAND-MR-1098-0SD.

2. Lindsay, R.S. (1999) Tests of level B suits-protection against chemical and biological warfare agents and simulants: Executive summary. Edgewood Chemical Biological Centre, Aberdeen Proving Ground, MD. July 1999. 14 p. AD-A3 68228; ECBC-TR-047.

(vi)

Serial / journal article (online from a database – e.g. EBSCO) Marshall, M., Carter, B., Rose, K., & Brotherton, A. (2009).Living with type 1 diabetes: Perceptions of children and their parents. Journal of Clinical Nursing, 18(12), 1703-1710. Retrieved from http://www.wiley.com/bw/journal.asp?ref=0962-1067

(vii) Internet – no author, no date

Pet therapy. (n.d.). Retrieved from http://www.holisticonline.com/stress/stress_pet-therapy.htm

(viii) Blog post

Liz and Ellory. (2011, January 19). The day of dread(s) [Blog post]. Retrieved from

http://www.travelblog.org/Oceania/Australia/Victoria/Melbourne/St-Kilda/blog-669396.html

(ix) Newspaper article

Matthews, L. (2011, November 23). Foodbanks urge public to give generously. Manawatu Standard, p. 4.

(x) Newspaper (online)

Rogers, C. (2011, November 26). Smartphone could replace wallets. The Dominion Post. Retrieved from http://www.stuff.co.nz/technology/gadgets/6038621/Smartphone-could-replace-wallets

(xi) Thesis (print)

Smith, T. L. (2008). Change, choice and difference: The case of RN to BN degree programmes for registered nurses (Master's thesis). Victoria University of Wellington, Wellington, New Zealand.

(xii)Thesis (online)

Mann, D. L. (2010). Vision and expertise for interceptive actions in sport (Doctoral dissertation, The University of New South Wales, Sydney, Australia). Retrieved fromhttp://handle.unsw.edu.au/1959.4/44704

(xiii) Wikis (including Wikipedia)

Moodle. (2011). Retrieved November 28, 2011, from Wikipedia: http://en.wikipedia.org/wiki/Moodle

The Reviewing Process

All submitted manuscripts are subjected to peer-review by independent reviewers. Primary review is made by Internal Editorial Team. Once the paper is selected by primary editorial team, it sends to the esteemed peers selected randomly across the country. Peer reviews are done by double blinding method where both the author and reviewer are unaware of each other. Final decision of accepting the article rests with the editor.

Final Selection

Verifying the Reviewer's comment, finally the editorial board will take decision to publish the paper. As there is limited space once the paper selected may publish in the next volume subject to the availability of space.

Author will be informed if paper is not selected for publication. No explanation will be given to the author for not selection of his/her paper.

Submission Process

All manuscripts must be submitted in MS-Word format through e-mail in the following

email address- lisrbu@gmail.com

After final selection of the article author has to send the corrected softcopy (through e-mail) with two hard copy (one side print) along with Plagiarism Report, Copyright and author of interest declaration in prescribed format, author's passport size picture with short biography to the following address.

To,
Head (Editor)
Department of Library and Information Science
Rabindra Bharati University
56A, B.T.Road
Kolkata-700050

Email: hod.lis@rbu.ac.in

Scholarly Journal (Blind Peer Reviewed)

(Previously UGC-CARE Enlisted)

RBU JOURNAL OF LIBRARY & INFORMATION SCIENCE

Volume 27, 2025; ISSN: 0972-2750

CONTENTS

RBU Journal of Library & Information Science
Volume 27, 2025
UGC – CARE **enlisted Journal w.e.f. 14.6.2019 to 09.02.2025**
Previously, it was under UGC List of Journals No. 45237, Sl. No. 2023
Visit: http://rbu.ac.in/home/page/102
Online Content of previous volumes (and full text e-copy of Volume 21, 24 & 22)
Visit: https://lisrbu.wixsite.com/dlis/rbu-journal-of-lis

Editorial

From Smart Machines to Intelligent Libraries: Why Human Judgment Still Matters

We are at a pivotal point in the evolution of knowledge. Generative artificial intelligence—through tools such as ChatGPT, Gemini, and DeepSeek—has revolutionized how information is produced, organized, and consumed. Tasks that once demanded hours of expert effort now take only seconds. This dramatic shift prompts a challenging question: If machines can provide every answer, what is left for librarians to contribute? What becomes of the librarian's role when machines seem to "know" it all?

Library and Information Science has continually adapted to technological change. The profession has weathered three significant transformations: from handwritten catalogues to automated systems, from automation to digital libraries, and from digital services to smart platforms. Each transition brought uncertainty, but also opened new possibilities. Today's shift—from smart tools to intelligent systems—shares the spirit of past changes, but stands apart in its unprecedented scale.

What makes this moment distinctive is not speed alone, but the illusion of intelligence.

Today's technologies are often described as "intelligent." Yet intelligence is not merely fast processing or pattern recognition. Smart devices can execute commands efficiently. AI systems can analyse vast datasets and generate fluent responses. But wisdom—the ability to judge, contextualise, and act ethically—remains firmly human.

True intelligence lies in understanding uncertainty, limitation, and consequence. It lies in recognising when information is incomplete, when sources are unreliable, and when answers require moral reflection. Libraries have preserved this tradition of wisdom for centuries. No algorithm has replaced it.

When search engines became dominant in the late 1990s, many predicted the decline of librarianship. Instead, librarians adapted. They embraced digital platforms, expanded user education, and redefined information services. The profession did not retreat; it advanced.

Generative AI presents a more complex challenge. It does not merely retrieve information—it produces it. This capability introduces three serious risks.

First, the problem of quality. Information is now abundant, but credibility is scarce. AI-generated content often lacks transparent sources, making verification difficult.

Second, the erosion of epistemic agency. When machines think on our behalf, users risk losing the habit of questioning, evaluating, and reasoning independently.

Third, the danger of confident error. AI systems frequently present inaccurate or biased information with persuasive authority. In academic, medical, and legal contexts, such "artificial ignorance" can be deeply harmful.

In this environment, the librarian's role becomes more essential.

Recent research makes one point clear—AI is not replacing librarians; it is redefining their professional mission.

Studies on AI literacy among LIS students in eleven countries, conducted by Zakir Hossain (2025), reveal widespread use of generative AI tools but limited awareness of ethical, legal, and critical implications. This gap highlights a central responsibility of contemporary librarianship: cultivating informed, reflective, and responsible users of technology.

Similarly, research by José Aguiñaga and colleagues (2025) on the use of generative AI in reference services demonstrates that AI functions most effectively as a complement rather than a substitute for professional expertise. While chatbots can provide rapid responses, they lack the contextual understanding and evaluative judgment that define high-quality reference work.

At the institutional level, studies from Latin America and Europe illustrate both the promise and limitations of AI. Research by Dr Diego Alejandro Patiño Guevara (Pontificia Universidad Javeriana, Colombia) shows how neural networks and clustering algorithms can integrate syllabus and circulation data to support evidence-based collection development. Meanwhile, scholars from the Universities of Glasgow and Edinburgh (2025) successfully transformed nineteenth-century catalogues into machine-readable data, while cautioning against the risks of hidden biases and undocumented datasets.

These examples underscore a critical reality: technological efficiency does not guarantee intellectual reliability. In this context, the librarian's role is evolving from information retriever to knowledge mediator. Today's professionals are increasingly responsible for evaluating AI-generated content, teaching digital and AI literacy, and providing ethical guidance on issues such as privacy, copyright, and data governance. As Rebecca Morris (2025) notes in her work on field-based learning, students are already engaging with AI in professional settings, reinforcing the need for structured, critical training.

Recognising this shift, professional bodies such as the American Library Association have introduced AI competency frameworks emphasising technical, critical, and transformational skills. Universities worldwide are integrating AI into LIS curricula. Yet evidence shows that many students remain underprepared, particularly in ethical and analytical domains. Infrastructure gaps, outdated syllabi, and limited faculty training continue to hinder progress. Today's librarians increasingly function as data curators, AI literacy specialists, and knowledge architects. Their task is not simply to manage information, but to design environments where both humans and machines can work responsibly.

Equally important is the changing information behaviour of younger users. Algorithm-driven platforms encourage passive consumption rather than active inquiry. Libraries must therefore reposition themselves as spaces for critical engagement, where users learn not only to access information, but to question and verify it.

History reminds us that technological revolutions do not destroy professions; they reshape them. Routine, predictable tasks are increasingly automated. But contextual judgment, ethical reasoning, and human connection remain irreplaceable. Librarianship rests precisely on these foundations.

An intelligent library is not defined by the number of digital tools it possesses. It is defined by how responsibly it uses them. It ensures access without sacrificing quality. It embraces innovation without abandoning values. It cultivates curiosity rather than passive consumption.

The future, therefore, does not belong to "smart libraries" alone. It belongs to intelligent libraries. As machines organise information, humans must create meaning. Library and Information Science students today are not merely preparing for technical careers. They are becoming bridge-builders between historical knowledge and digital futures. AI will be a powerful instrument in this process. But the direction, purpose, and moral framework of that instrument will remain the result of human choices.

Will libraries disappear? Yes, only if they refuse to evolve. Will librarians become irrelevant? Yes, only if they abandon their intellectual mission. To be truly intelligent is not to be fast. It is to be thoughtful. Not to generate answers, but to understand their implications. Not to accumulate data, but to safeguard knowledge. In the age of artificial intelligence, these human capacities matter more than ever.

A Warm wish to all

Professor Sudip Ranjan Hatua
Editor
V-27, 2025

Blind Peer Reviewed
(Previously UGC-CARE)
ISSN: 0972-2750

Manuscript Received on –
30th December, 2026
Primary Reviewed on –
6th January, 2026
Peer Reviewed on –
24th January, 2026
Accepted for publication
1st February, 2026

The Application and Limitation of DC, LOM and LRMI Metadata for archiving K-12 Learning Resources

Debasis Das & Sudip Ranjan Hatua

Debasis Das
Ph.D. Scholar
Dept. of Lib. & Inf. Science
Rabindra Bharati University, Kolkata
Librarian
Public Library, Govt. of West Bengal
debdas12312@gmail.com

Dr. Sudip Ranjan Hatua
Professor
Dept. of Lib. & Inf. Science
Rabindra Bharati University, Kolkata
sudiprh@rbu.ac.in

ABSTRACT

Introduction: *E-educational resources are key elements of the school education system. Semantically describing educational resources using standard metadata, such as DC, IEEE LOM, and LRMI, can improve the effectiveness of resource discovery and information dissemination. The Dublin Core (DC) and IEEE LOM standard, LRMI, are user-friendly metadata element sets that are specifically examined in this article for their application and limitations in archiving resources in a digital repository.*

Purpose: *This paper examines the most popular metadata standards, DC, IEEE LOM and LRMI, for Open Educational Resources for a school repository to highlight the advantages and limitations of these standards.*

Research Problem: *There are many metadata formats, and standards have been developed; however, some schemas have limitations. However, a single metadata schema cannot fulfil all the desired requirements for archiving educational resources. The work first reviews and analyses the DC, IEEE LOM, and LRMI metadata standards, comparing them for suitability for archiving educational resources.*

Objectives: *The current study aims to analyse DC, LRMI, and the IEEE LOM metadata standard and determine their suitability for storing educational content.*

Methodology: *This study and analysis of various metadata schemas and learning objects, as well as the analysis of board study materials. Investigate various metadata schemas, learning objects, and Study Material for the K-12 Standard, and provide a prototype metadata schema.*

Findings: *No existing standard satisfies the management of study material at the K-12 level. Therefore, it is very difficult to manage the library materials for this type of School library.*

Keywords: *Metadata Standard, Dublin Core (DC), IEEE LOM, LRMI, School Digital Repository*

1. Introduction

The word "learning materials" has an inadequate standard meaning, and there is growing recognition that this may be incorrect for school educational resources that serve multiple purposes and come in numerous varieties. The most obvious action enabled by metadata is resource discovery, which most people associate most strongly with metadata: learning material repositories, especially school repository archives, and high-quality learning materials for online learning. Several metadata standards have been developed for learning resource descriptions. To enable efficient exploration and recovery of materials tailored to the needs of school students, the teaching resources are tagged with metadata that describes the educational patterns of the documents, the class level of the materials, etc. However, the educational resources can be linked to a uniform metadata standard to make them easier to share and reuse across learning management systems.

A metadata standard is a set of accepted guidelines that define the kinds and structures of data, ensuring consistency in data management and description and facilitating interoperability and standardisation. The term "metadata" refers to the information that librarians have typically added to catalogs by collection of entries like by the title, by the author(s), creation date of materials, publication date, subject area, and the index number that indicate the item's location on the shelf, are use to describe a book or other library material in the library catalog. This information usually concerns the description of online resources.

2. Literature Review

This study reviews the literature on Metadata Standards, OER Metadata Standards, and their limitations for archiving educational resources.

In the paper by Getaneh Alemu (2024) titled "Metadata Standards and Models," the author elucidates the roles of various stakeholders in the generation of metadata. He investigates major standards such as Dublin Core, MARC21, IFLA LRM, FRBR, RDA, and BIBFRAME, emphasising the significance of controlled vocabularies and authority headings in maintaining consistency and facilitating information retrieval. They also discuss future perspectives and developing trends in metadata standards and models, such as enrichment, linking, openness, and filtering, which are being pushed by collaboration, the Semantic Web, open science, big data, and AI.

Christian, M., Klante, S., & Biel, C. (2024), in their paper " Metadata for Learning on Distributed Platforms", outline the most relevant metadata standards for the educational sector, as well as different scenarios of application. They also find that recent years have seen a shift towards greater interoperability, with the sharing of content across platforms becoming crucial. Learning object metadata standards have facilitated filtering, cataloguing, and dependency creation. Initially, cross-schema compatibility was limited, as learning content was primarily managed within individual systems.

Mosha, N. F., and Ngulube, P. (2023), in their review article, point out that research data preservation, discovery, and reuse have been central to the mission and vision of most higher education institutions (HEIs) and their entities, such as universities, libraries, archives, and repositories. They provide an outline of the factors to consider when developing and evaluating metadata standards to improve research data preservation, discoverability, and reusability. They found little evidence on how these metadata standards can be used to improve the discovery and reuse of research materials in repositories.

In an article "Metadata Analysis of Open Educational Resources", Tavakoli, M., Elias, M., Kismihók, G., & Auer, S. (2021), they highlighted from a learner point of view that OERs vary in terms of a large number of important features, like levels of education, topics or vocation. Open Educational Resources are available in a wide range of formats and languages, so it is increasingly important to analyse and control their quality, with OER metadata playing a key role.

Deus & Barbosa (2020) explained in an article, "The Use of Metadata in Open Educational Resources Repositories: An Exploratory Study", that OERs are stored in digital repositories and use metadata to describe their content. However, the metadata is not yet used correctly and frequently by users, which creates several problems in finding and retrieving OER. Considering this scenario, they intend to identify how metadata is used in the context of OER and the implications for finding resources in OER repositories.

Tavakoli, M et al., (2020) in their article "Quality Prediction of Open Educational Resources: A

Metadata-based Approach" explained, "Online learning environments have accumulated millions of Open Educational Resources (OERs). However, for learners, finding relevant and high-quality OERs is a complicated and time-consuming activity". They propose an OER metadata scoring model and build a metadata-based prediction model to estimate OER quality.

Villalba-Condori et al. (2019), in their work on Open Educational Resources (OER) in the article "A Methodology to Assign Educational Resources with Metadata Based on the Purpose of Learning", aggregate metadata that allows for considering the cognitive processes that correspond to the learning purpose. However, the issue of establishing a repository with OERs that fit the standards for teaching and learning remains. They propose adding metadata to each resource that specifies the action (verb), the cognitive processes associated with that verb, and the resource's content and context in relation to the student group.

Andrade, M., and Baptista, A. (2015), in their article "The Use of Application Profiles and metadata schemas by digital repositories: findings from a survey," demonstrate the objective of determining the use of application profiles and associated metadata standards.

Haneefa K, Mohamed, & Chembrakuzhi, Manu. (2014) in a conference article "Metadata Standards for Open Educational Resources" points out that "resource discovery of learning materials can be made effective by semantically tagging them with standard metadata". They focused on metadata use in e-learning systems, assessing the most commonly used metadata standards for Open Educational Resources.

Beak, J. (2014). In his article, "A Child-driven metadata schema: a holistic analysis of children's cognitive processes during book selection," he sought to construct a child-driven metadata schema by understanding children's cognitive processes and behaviours during book selection.

In a case study by Palavitsinis, Nikos & Manouselis, Nikos & Sanchez-Alonso, Salvador (2014) for "Metadata quality in learning object repositories", their work seeks to solve the issue of low-quality metadata records characterising instructional content in Learning Object Repositories (LORs). This seeks to increase the discoverability of learning objects in such LORs by implementing a systematic procedure that facilitates metadata development. This research also proposes a metadata quality assessment and certification process for LORs. The process was designed to be general and adaptable to a variety of application domains.

3. Objectives of the Study

The objective of the present study is to

- Analysis of DC metadata elements, LRMI and IEEE LOM metadata standard.
- Finds the usability of DC metadata elements, LRMI and IEEE LOM metadata to store educational resources.
- Try to identify the limitations of the DC, LRMI, and IEEE LOM metadata standards.
- Find out how other metadata schemas can fulfil the needs of archiving educational content, especially for school digital repositories.
- Build a new metadata schema for archiving educational resources for the school repository.

4. Metadata Schemas

Various standard schemes were created to meet specific demands related to the use and management of information, and the need for some interest groups to standardise their information classification led to these standards. Numerous user contexts and disciplines are developing a wide range of metadata methods. To better describe learning objects, several educational metadata standards have been proposed over time. Some widely adopted metadata element sets for this purpose include DC, IEEE Learning Object Metadata (LOM), and LRMI.

5. Dublin Core Metadata Initiative (DCMI)

The Dublin Core Metadata Initiative is an open forum dedicated to creating an interoperable metadata standard that enables a wide variety of resources for various applications. The fifteen-metadata elements i.e. Title, Subject, Description, Type, Source, Relation, Coverage, Creator, Publisher, Contributor, Rights, Date, Format, Identifier, and Language- are theelements used in Simple Dublin Core to express elements as attribute-value pairs and Audience, Provenance, and Rights Holder are three qualifierthat improve the elements' semantics in a way that could be helpful for resource discovery are included in the qualified Dublin Core. The main goal of the DCMI structure is to foster broad, inter- and intra- and global consensus around a key set of web assets, enabling various communities and areas of interest to restore and discover networking information. Although it is acknowledged that metadata has several uses beyond the objective of networked information retrieval and discovery, these needs are unrelated to DCMI's mission and do not directly influence its activities.

The DCMI Advisory Committee established the DC-Ed Education Working Group in August 1999, with the objective of "developing and discussing a proposal for the use of Dublin Core metadata in the description of educational resources." In that it was tasked with investigating potential new domain-specific DCMI elements and qualifiers, the establishment of the DC-Ed Working Group represented a significant shift from DCMI practice. To put it briefly, the creation of a standard set of DCMI-recommended elements and qualifiers for use with interoperable educational content was the initial aim of DC-Ed.

5.1 Limitation of Metadata Elements in Dublin Core

The Dublin Core metadata has features that are helpful for general-purpose applications; however, it lacks attributes that describe a document's pedagogical approach. The primary disadvantage of the Dublin Core is its simplicity, which limits specificity and makes it challenging to transfer additional data between systems and translate it for use in other systems (Beall, 2004). Additionally, it lacks standards for identifying and structuring data within its elements. The use of spaces between data inputs in each element poses a risk. Some of the DC elements, i.e., Type, Subject, Format, Description, Creator, and Identifier, are confusing in their usability, as discussed below.

DC Type:

The Dublin Core component "Type" explains the characteristics, genre, or category of a resource (such as a book, image, or sound) in standard phrases for interoperability, such as "Text," "Image," "Sound," or "Collection," which are often taken from the DCMI Type Vocabulary to ensure consistent discovery and use across various systems.

There is confusion in this DC element: the type cannot include an element that indicates whether it is a resource type or a content type. This element allows students and other users to search by both resource and content type; however, DC does not meet these criteria. Multiple DC.Type elements are used for resources with hybrid content (for example, Type="Image" and Type="Text" for a picture with a description). In contrast, qualifiers (such as Text.Collection) are used for more specialised kinds in qualified Dublin Core.

DC Subject:

Some limitations can be noticed in the DC element "Subject". DC provides the option to incorporate the subject using either dc.classification or dc.keyword, which are in fact interlinked. If we use the dc.subject keyword and include the metadata element, fetching based on the keyword will return all keywords. For example, the query "Delta" can be retrieved from any discipline, such as Physics, Geography or Mathematics. The user may require a resource on "Delta" from Geography, but the dc metadata element can fetch "Delta" from any subject.

DC Format:

The Format of a resource refers to the media, data format, and materials used in its instantiation, as well as its size, if applicable. In a DC element, the resource's "file format" is easily known to include, whereas the "duration" of the resource is confused with the file format, physical media, or the resource's dimensions.

DC Description:

An explanation of the resource's contents. Examples of descriptions include an abstract, a table of contents, a reference to a graphical representation of the material, or a free-text explanation of the topic, among many others. The phrase "description" is ambiguous because it might refer to an abstract, a table of contents, or a free-text explanation of the material. As a result, operators can use keywords or phrases while creating any description they want about the material in any style.

DC Creator:

The creator is primarily responsible for developing resource content. Creators might be individuals, companies, or services. In general, use the creator's name to refer to the entity. Dublin Core does not specify how the inventor should be expressed. The description does not specify the main source of the creator's information. The comment states that "typically, the name of a creator should be used to indicate the entity". The preceding clarification does not help to define the originator. Furthermore, it may encourage implementers to use the creator's name however they see fit, regardless of any official list of names.

DC Identifier:

An identifier is an authentic indicator of a resource inside a specific context. The best technique is to identify the resource with a string or number that follows a formal identification scheme. Formal identification methods include the Uniform Resource Identifier (URI), commonly known as the URL, the Digital Object Identifier (DOI), and the International Standard Book Number (ISBN).

A resource's URI contains information about its current location. The URL is not permanent. All links to a resource may break if its placement inside a database or server changes, or if it is moved to a new database or server. This is one of Dublin Core's downsides.

Therefore, based on the discussion above, we can point out the limitations of DC as follows.

i. **DC is** designed for general-purpose applications and lacks elements to specify learning resource types, educational objectives, or alignment with curricula.

ii. **It is too simple to sacrifice** specificity, which makes it challenging to transfer rich, detailed data between specialised systems.

iii. Several core elements are confusing due to vague or overloaded definitions, such as

DC Type:

▪ Ambiguity between **resource type** (e.g., image) and **content type**

(e.g., lesson plan). The DCMI Type Vocabulary (e.g., "Text," "Image")

does not adequately represent educational content types.

Requires multiple dc.type entries or qualifiers for hybrid resources.

(e.g., an image with descriptive text), which complicates consistent application.

DC Subject:

▪ Lacks a reliable mechanism to disambiguate subject terms by discipline. A keyword like "Delta" can be retrieved from Physics, Geography, or Mathematics without context, potentially returning irrelevant resources for a user's specific need.

DC Format:

▪ Confusion arises because the element mixes file formats (e.g., PDF), physical media, dimensions, and duration. There is no clear distinction, making a precise description difficult.

DC Description:

The term is overly broad and ambiguous. It can refer to an abstract, table of contents, or a free-text summary, allowing operators to create descriptions in any style. This variability undermines consistent search and discovery.

DC Creator:

▪ Does not specify a standardised way to express the creator's name (e.g., no guidance on authority files, name order, or corporate naming conventions). The comment that a "name should be used" is insufficient and encourages inconsistent, non-interoperable practices.

DC Identifier:

▪ While it recommends formal schemes (URI, DOI, ISBN), a critical flaw is its common reliance on non-persistent URLs. If a resource's location changes (e.g., moved to a new server), all links break, rendering the metadata obsolete and the resource unfindable.

6. IEEE LOM

The IEEE profile for learning object metadata aims to develop recognised technical requirements, suggested practices, and guidelines for learning technology. It outlines a data architecture that establishes the structure of a learning object's metadata instance. Any entity, digital or non-digital, that can be utilised for instruction, training, or learning is considered a learning object according to this standard. Semantics, vocabulary, and extensions are all explained in LOM. LOM is a valuable group of internationally accepted metadata elements organised into nine descriptive classes: life cycle, technical, relation, annotation, rights, educational, classification, general, and metadata.

The LOM data model comprises a hierarchy of data items, including simple and aggregate data elements. Aiming high, the LOM standard specifies about 80 distinct "elements" or characteristics for the management and description of learning resources. General informational items such as title, author, description, and keywords are among these components, along with technical details like file size and file type. Educational and interpretative factors, such as "typical learning time" or "educational context," are some metadata elements. However, it is well known that this metadata specification presents challenges for its implementers due to the sheer quantity and diversity of its constituents.

6.1 Limitation of IEE LOM

The Learning Object Metadata (LOM) standard defines a structured framework in which each core data element is associated with a prescribed value space and data type. This value space delineates the constraints for permissible data entries, ranging from unrestricted Unicode strings for some elements to strictly controlled vocabularies or syntactical patterns for others, such as formalised codes. A fundamental principle of LOM is its inherent flexibility; implementers—whether data or service providers—are not obligated to support every element within the conceptual schema, nor is the schema intended to restrict the nature of the information that can be described. This extensibility allows for the integration of elements from complementary metadata standards, the omission of non-pertinent LOM elements, and the augmentation of LOM's predefined vocabularies with community-specific values.

However, the LOM conceptual schema exhibits significant design limitations that impede broad interoperability. Primarily, it is not grounded in a foundational semantic framework, such as the Resource Description Framework (RDF), which is a common abstract model for modern metadata.

standards. This absence complicates semantic alignment and integration with other metadata ecosystems. Furthermore, challenges arise from its extensibility model. While communities frequently develop custom extensions to address local or consortium-specific needs, there is minimal evidence of cross-community adoption of these extensions. Consequently, similar functional requirements are often met through parallel yet distinct modifications, leading to a proliferation of incompatible solutions that ultimately fragment the interoperability landscape.

Therefore, the primary limitations of the LOM conceptual data schema are:

- Lack of a Foundational Semantic Model
- Fragmented and Non-Transferable Extensibility
- Proliferation of Incompatible Variants

7. Learning Resource Metadata Initiative (LRMI)

The Learning Resource Metadata Initiative (LRMI) was launched with the goal of facilitating the discovery of instructional materials for educators and students through specialist resource discovery services. Schema.org has been introduced to help understand the LRMI approach. Sponsors of Schema.org include Yandex, Microsoft, Yahoo, and Google. An agreed-upon ontology or hierarchy of resource categories, and a vocabulary for describing resource attributes, are the two parts of Schema.org. To provide a uniform vocabulary for defining learning resources, Creative Commons (CC) and the Association of Educational Publishers (AEP) are leading the Learning Resource Metadata Initiative (LRMI). The project has established an industry-standard metadata framework for labelling online learning resources. The vocabulary is the first industry-specific education framework to be independently established and compatible with schema.org, making it easier to find and search for educational content online. OER repositories will be more transparent and interoperable if they use a uniform structure for labelling and classifying educational materials.

7.1 LRMI Specification

The LRMI specification is a set of attributes used to characterise learning materials. It expands upon the rich language offered by Schema.org, primarily via Schema.org/CreativeWork and the Educational Audience developed by LRMI. The metadata schema for content on the Web, such as a "Creative Work" in the form of a "Book," which includes metadata like author, publisher, datePublished, etc., is already defined by Schema.org. Additionally, it specifies particular information for many types, including Article, ImageObject, AudioObject, VideoObject, and others. Five qualities pertinent to learning objects that are already covered by schema.org were identified by LRMI. LRMI builds on Schema.org's work to standardise website markup.

Therefore, the fundamental limitations of the Learning Resources Metadata Initiative (LRMI) can be summarised as follows:

1. LRMI is not a standalone metadata schema. It is explicitly defined as an extension of Schema.org. This dependence limits LRMI's autonomy and ability to evolve independently of Schema.org's priorities.

2. The specification does not create a novel or comprehensive framework for educational metadata. Instead, it primarily identifies and formalises a small number **(five)** of learning-relevant properties *within* the pre-existing, broadly defined Schema.org/CreativeWork type and its related types (e.g., Article, VideoObject).

3. LRMI characterises learning materials by anchoring them in the generic concept of a "CreativeWork" (e.g., a Book, Article, or VideoObject). This model 5 may not fully capture the pedagogical relationships in modern educational resources,

4. The text frames LRMI's are designed for incremental improvement of web resource discovery rather than for addressing deeper, systemic challenges in educational metadata.

8. Comparison among DC, IEEE LOM and LRMI Metadata Standard

Metadata standards aim to ensure the interoperability, portability, and reusability of this content to **address** the lack of interoperability in e-learning. Nonetheless, there is a great deal of misunderstanding regarding the use and application of metadata standards, as well as their function in their implementation. This is because various educational metadata have been created to address various needs. It can be difficult to comprehend most of this information because the elements have distinct sets of features and other characteristics.

Notably, the popular data almost exactly maps to the more straightforward and widely used Dublin Core metadata element set. Additional observations show that almost all sets had LOM profiles with specialised vocabulary to convey traits primarily pertinent to the communities from which they originated. As expected in the Metadata Standard, there was some difficulty in incorporating this vocabulary into LOM and ensuring that other users understood the meaning of the vocabulary items. It is not always easy to deal with the LOM Metadata record structure with basic tools.

The Dublin Core element type does not indicate the "learning" kind, such as "activity" or "assessment," but rather a specific content type, such as "text" or "audio."." In contrast, the IEEE LOM components include values for both media format and academic usage, academic purpose is defined by some curriculum vocabulary, such as "lesson plan," "online project," "Practice Lesson," and so on. Dublin Core is used for resource discovery in a variety of educational content, despite being designed with e-learning content in mind. It has occasionally been enhanced with components tailored to the educational field. Implementing GOL e-Learning metadata can benefit from understanding where and how DC has been altered to make it more relevant for educational materials.

This can be presented in a table format as below:

Parameter	Dublin Core	IEEE LOM	LRMI
Purpose	General-purpose, cross-domain resource discovery	Specifically engineered for describing e-learning objects and resources.	Designed to markup educational resources on the web for improved discovery via search engines.
Educational	Low. Describes generic	High. Includes elements	Moderate-High.
Specificity	format (e.g., "text," "audio") but not learning type.	for educational context, learning resource type (e.g., "lesson plan"), and interactivity.	Introduces educational properties (e.g., educational Alignment, learning Resource Type) into general web markup.
Structural; Foundation	Simple element set; not based on a formal abstract model like RDF in its core.	Conceptual data schema not based on a foundational model like RDF, limiting semantic interoperability.	Built as an extension of Schema.org, which uses a shared vocabulary and is inherently compatible with web standards.
Interoperability	Strong for basic discovery	Strong within closed or	Strong for web-scale

Strength	across diverse systems due to widespread adoption and simplicity.	compatible e-learning systems but weak in broader semantic web contexts.	discovery via major search engines that consume Schema.org markup.
Implementation Challenge	Requires enhancement with educational extensions to be fully useful for learning resources.	Complex structure difficult to handle with basic tools; challenges insharing and understanding community-specific vocabularies.	Dependent on search engine adoption and a critical mass of marked up content to realize its full potential.
Application	Digital libraries, repositories, and general cataloging of educational content among other materials.	Learning Object Repositories, Learning Management Systems (LMS), and formal e-learning ecosystems.	Markup of open educational resources (OERs) and any educational content on the public web.

Table 1: DC, IEEE LOM and LRMI: Comparative Study

9. Metadata Elements for K-12 Learning Resources

By analysing various digital school repositories and online learning platforms used at the school level, either for online or classroom teaching, it is found that 24 metadata elements are mostly required to describe any collection. The metadata elements- title, creator, publisher, subject, source, language, relation, coverage, description, identifier, abstract, edition and keyword elements available in DC, which are most usable for archival. LRMI and IEEE LOM primarily cover learning resource metadata that are not available in DC. The following shows 24 commonly used metadata elements for describing educational materials in any digital archive. The table also compares which metadata elements are available in which schema. Metadata variables describe the features of metadata components throughout the entire collection.

Sl No	Metadata Elements	Dublin Core	IEEE LOM	LRMI
1.	Title	Yes	Yes	No
2.	Subject	Yes	No	No
3.	Description	Yes	Yes	No
4.	Type	Yes	Yes	No
5.	Relation	Yes	Yes	No
6.	Source	Yes	Yes	No
7.	Creator	Yes	No	No
8.	Coverage	Yes	Yes	No
9.	Contributor	Yes	No	No
10.	Date	Yes	Yes	No
11.	Format	Yes	Yes	No
12.	Identifier	Yes	Yes	No
13.	Language	Yes	Yes	No
14.	Publisher	Yes	No	No
15.	Rights	Yes	Yes	No
16.	educational Alignment	No	No	Yes
17.	educational Role	No	No	Yes

18.	educational Use	No	No	Yes
19.	time Required	No	Yes	Yes
20.	interactivityType	No	No	Yes
21.	typical Age Range	No	Yes	Yes
22.	Classification	No	Yes	No
23.	learning ResourceType	No	Yes	Yes
24.	Keyword	No	Yes	No

Table 2: Comparison of Metadata Elements

10. Conclusion

Metadata practices diverge from conventional cataloguing norms in both form and application. Yet, a closer examination of their core objectives reveals a fundamental alignment: both systems are fundamentally engaged in the enterprise of bibliographic description. This conceptual continuity extends to the underlying principles governing metadata's development and architecture, which adhere to established cataloguing tenets even when applied to diverse digital objects. As noted by Milstead and Feldman (1999), the term "metadata" in the electronic context expansively denotes "data in the broadest sense—encompassing datasets, textual information, web pages, graphics, music, and any other entity likely to exist in digital form."

The strategic annotation of documents with standardised metadata enhances their reusability and interoperability across systems. Evaluation of prevalent metadata standards indicates, however, that no single schema is universally sufficient to meet all functional requirements. The nature of the document itself dictates the appropriate standard, as no monolithic metadata set can adequately serve the disparate needs of all agencies or communities of practice. Concurrently, through shared semantic, syntactic, and structural conventions, metadata standards facilitate cross-system sharing and reuse. In practice, achieving a comprehensive description often necessitates the complementary application of multiple standards. For open educational resources (OERs), the Learning Resources Metadata Initiative (LRMI) specification has emerged as particularly significant. Its formal integration into Schema.org and subsequent adoption by major search engines for semantic indexing positions LRMI as a pivotal standard. As search engines increasingly utilise LRMI and a critical mass of educational content becomes consistently marked, educators, learners, and other stakeholders will benefit from the greatly improved discoverability of relevant materials.

This evolution coincides with a growing scholarly recognition that educational resources are not merely discrete "learning objects," but complex aggregations of interrelated assets. These resources possess distinct attributes and are woven together through intricate relational networks, a reality that metadata schemas must increasingly capture to represent the nuanced ecology of digital learning materials fully.

References

Alemu, G. (2024). Metadata Standards and Models. https://doi.org/10.1016/B978-0-323-95689-5.00035-3 retrieved on December 05, 2025

Andrade, M and Baptista, A (2015). The Use of Application Profiles and Metadata Schemas by Digital Repositories: Findings from a Survey, available at https://dcpapers.dublincore.org/pubs/article/viewFile/3770/1960 retrieved on December 05, 2025

Beak, J. (2014). Children's perceptual cognitive factors in book selection and metadata schema: Pilot study. Proceedings of the American Society of Information Science and Technology, 49(10), 1–10.

Caplan, P. (2003). *Metadata fundamentals for all librarians*. Chicago: American Library Association.

Christian, M., Klante, S., & Biel, C. (2024). Metadata for Learning on Distributed Platforms. Retrieved from https://conference.pixel-online.net/files/foe/ed0014/FP/3797-ELRN6560-FP-FOE14.pdf

DCMI Usage Board. (2012). DCMI: Dublin coreTM metadata element set, version 1.1: Reference description. Retrieved from www.dublincore.org website: https://www.dublincore.org/specifications/dublin-core/dces/

Deus, W. S. de, & Barbosa, E. F. (2020). The Use of Metadata in Open Educational Resources Repositories: An Exploratory Study. 2020

IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC). Spain: IEEE. Retrieved from https://ieeexplore.ieee.org/document/9202802

Good, J. (2003). A Gentle Introduction to Metadata. Retrieved November 4, 2025, from http://www.language-archives.org/documents/gentle-intro.html

Haneefa, M., & Manu C. (2014). Metadata Standards for Open Educational Resources. *59th ILA International Conference on Managing Libraries in the Changing Information World: From Surviving to Thriving*. Retrieved from https://www.researchgate.net/publication/305298208_Metadata_Standards_for_Open_Educational_Resources

Heery, R. (1996). Review of metadata formats. *Program*, *30*(4), 345–373. https://doi.org/10.1108/eb047236

IEEE Standard for Learning Object Metadata. (2020). 1–50. https://doi.org/10.1109/ieeestd.2020.9262118

LRMI Terms (RDF). (2022). Retrieved November 4, 2025, from https://www.dublincore.org/ website: https://www.dublincore.org/specifications/lrmi/lrmi_terms/2022-06-14/

Milstead, J., & Feldman, S. (1999). Metadata: Projects & Standards. *Online*, *23*(1), 32–40.

Mosha, N. F., &Ngulube, P. (2023). Metadata Standard for Continuous Preservation, Discovery, and Reuse of Research Data in Repositories by Higher Education Institutions: A Systematic Review. *Information*, *14*(8), 427. https://doi.org/10.3390/info14080427

Palavitsinis, Nikos & Manouselis, Nikos & Sanchez-Alonso, Salvador. (2014). Metadata quality in learning object repositories: A case study. The Electronic Library. 32. 10.1108/EL-12-2011-0175.

Riley, J. (2017, January 1). Understanding Metadata: What is Metadata, and What is it For?: A Primer | NISO website. Retrieved September 4, 2025, from www.niso.org website: https://www.niso.org/publications/understanding-metadata-2017

Tavakoli, M., Elias, M., Gabor Kismihok, & Auer, S. (2020). Quality Prediction of Open Educational Resources: A Metadata-based Approach. ArXiv (Cornell University). https://doi.org/10.1109/icalt49669.2020.00007

Tavakoli, M., Elias, M., Kismihók, G., & Auer, S. (2021). Metadata Analysis of Open Educational Resources. *In LAK21: 11th International Learning Analytics and Knowledge Conference (LAK21)*. Irvine, CA, USA: Cornell University. https://doi.org/10.1145/3448139.3448208

Xie, I., & Matusiak, K. K. (2016). Metadata. *Discover Digital Libraries*, 129–170. https://doi.org/10.1016/b978-0-12-417112-1.00005-3

About Authors

Mr. Debasis Das is a Ph.D. scholar in the department of Library and Information Science of Rabindra Bharati University. He is also a librarian under Dept. of Mass Education Extension & Library Services, Govt. of WB. He worked as a Senior Project Officer in National Digital Library of India (NDLI) project at the central library, IIT Kharagpur from 2015 to 2024. Before that, he worked as Project Assistant at the central library, ISI Kolkata. He completed his M.Lib.Sc degree from Rabindra Bharati University in 2012 and his M.Phil. from the University of Calcutta in 2020. He has qualified UGC-NET in 2014. The present study is a part of his Ph.D. research work. He has published many articles in national and international journals. His area of interest is Digital Repositories and Metadata Schemas.

Dr. Sudip Ranjan Hatua, a DRTC alumni presently working as Professor and HOD in Department of Library & Information Science, Rabindra Bharati University, Kolkata, India. He has awarded PhD from Vidyasagar University. Prior to 20 years teaching experiences he served 7 years as working professional at ICAST, National Aerospace Laboratories, Bangalore, Karshak Engineering College, Hyderabad, followed by Central Library, Indian Institute of Technology, Kharagpur. He has authored around than 60 articles published in National, International Journals, seminars, conferences proceeding, book chapters etc. and 7 books. He was performed as a content writer of B.Lib.I.Sc course of NSOU and translator of a course content of IGNOU. 2 M.Phil Scholars and 9 Research Scholars awarded and six are pursuing Ph.D. till 2025 under his supervision.

Blind Peer Reviewed
(Previously UGC-CARE)
ISSN: 0972-2750

Manuscript Received on –
13th December, 2026
Primary Reviewed on –
6th January, 2026
Peer Reviewed on –
24th January, 2026
Revised on –
27th January, 2026
Accepted for publication
3rd February, 2026

Artificial Intelligence and Robotics in Libraries: Opportunities for Integrating AI in LIS Curriculum to Prepare Future Library Professionals

Mohd Shoaib Ansari, Sneha Tripathi, & Punam Chauhan

ABSTRACT

Introduction *- This study examines the integration of Artificial Intelligence and robotics into Library and Information Science education in India. It highlights the growing gap between global technological advancements and Indian LIS curricula. By reviewing existing programmes, the paper emphasizes the need for curriculum reform that addresses ethical, technical, and governance challenges in AI-driven information environments.*

Purpose *- The purpose of the study is to underline the need for LIS education in India to evolve in response to a technology-driven information society. It stresses curriculum flexibility, measurable learning outcomes, and a balance between traditional LIS syllabi and emerging areas such as AI, data curation, digital literacy, and ethical information governance.*

Research Problem *- Despite rapid technological changes in libraries and information services, LIS education in India has not adequately integrated AI and robotics. This study addresses the problem of curricular lag by examining whether current LIS programmes prepare future professionals to understand, apply, and critically evaluate emerging technologies in professional practice.*

Objective *- The study aims to assess the extent to which AI and robotics are included in Indian LIS curricula and to identify opportunities for curriculum revision. Specifically, it seeks to determine how many institutions address these technologies and how LIS programmes can be better aligned with emerging technological and professional requirements.*

Methodology *- The study adopts a qualitative content analysis of postgraduate LIS curricula from selected Indian universities. Course syllabi and learning outcomes were reviewed using defined inclusion criteria related to AI, digital technologies, data science, and ethics. The relevant subjects were analysed systematically to assess the depth and context of technological integration.*

Findings *- AI and robotics are largely absent as core subjects in Indian LIS programmes. While some courses briefly mention emerging technologies, practical and technical engagement remains limited. The study recommends revising the curriculum, fostering interdisciplinary collaboration, and enhancing faculty development to prepare LIS graduates for providing ethical, technology-enabled, and future-ready information services.*

Keywords: *Artificial Intelligence, Robotics, Education, LIS Curriculum, Curriculum Review, Technological Adaptation*

Dr. Mohd Shoaib Ansari
Assistant Director
Directorate of Higher Education,
Chhattisgarh

Dr. Sneha Tripathi
Deputy Librarian
Central Library, Banaras Hindu University,
Varanasi, Uttar Pradesh
sneha.tripathi10@bhu.ac.in

Dr. Punam Chauhan
Associate Professor-Library
Government Degree College,
Gosaikheda, Unnao, Uttar Pradesh

1. Introduction

Libraries and librarians have long been recognized for their innovative use of new technologies, particularly to enhance their services to users and refine their professional practices (Tenopir et al., 2013). Over the years, numerous new and emerging technologies have sparked discussions about how library and information science (LIS) education should evolve to keep pace with technological advancements (Parirokh, Daneshgar, & Fattahi, 2009). Now, as we move past the era of Web 3.0 and enter into the world of the metaverse, robots, and artificial Intelligence, these conversations have become more important than ever (Bakar et al., 2022). Once again, professionals are asking how LIS can remain relevant and how to prepare library staff with the right skills to lead and contribute in meaningful ways (Raju, 2020).

Artificial Intelligence (AI) and robotics have recently become major trends across many fields, including libraries and information centres (Jha, 2023). One area where libraries have made notable progress is in utilizing machine learning for chatbots and virtual assistants (Bawden & Robinson, 2012). However, much existing research and writing on AI and robotics in libraries remains speculative, often attempting to predict the future (Asadzadeh et al., 2022). There is a lack of a clear definition and understanding of how widely AI is used in libraries. People use the terms "AI" and "robotics" in different ways, and there are numerous possibilities for how these technologies can be applied in LIS practice (Ajani et al., 2024).

As AI and robotics grow and become more practical in everyday work, questions are being raised about what these tools will mean for the profession's future. These questions are part of an ongoing cycle in which librarians explore the best ways to utilize technology in their work (Tang & Sae-Lim, 2016). However, AI and robotics are totally different from past technologies. They are autonomous, data-driven, and come with their ethical concerns (Floridi et al., 2018). To prepare for these changes, the profession needs to understand how these tools might affect their profession. They need to learn how to use it responsibly and contribute to ethical conversations around them. This process will begin with education, specifically, what future professionals learn in LIS programs (Hossain, 2025).

This paper examines the relationship between developments in AI and robotics and LIS education in India. The goal is to discuss the opportunities and challenges of these technologies for professional training. To do this, the authors reviewed LIS programs at some Indian academic institutions.

1.1. AI and Robotics in the Context of Digital Transformation

AI and robotics have already begun transforming professional fields such as law and medicine. However, their impact on the LIS field has not received much attention, especially in India. Library professionals generally exhibit a positive attitude toward incorporating new technologies into everyday library services; however, integrating AI and robotics remains a challenging task (Yoon et al., 2021). It is essential to recognize that robots and AI are not merely machines performing tasks. They are deeply connected with the growing power of data and information systems (Cockshott & Renaud, 2016). As libraries in India continue to face significant changes due to digital transformation, there is a clear need for more academic studies and real-life examples from library practitioners. This would help ensure that Indian libraries not only follow trends but also actively lead and shape the future of library services (Winkler & Söllner, 2018).

Globally, professional library associations such as the American Library Association (ALA) have been encouraging greater involvement in AI-related developments. However, most progress has been observed in large, well-funded academic institutions, such as Stanford or MIT (Wheatley & Hervieux, 2019). In India, though there has been strong awareness regarding AI technologies and their potential applications in libraries, only a few central and technical universities are exploring AI applications in libraries (Subaveerapandiyan & Gozali, 2024). Despite this slow adoption, the profession is well-positioned to contribute to broader ethical discussions surrounding AI and robotics, particularly in areas such as data privacy, fairness, access to information, and the responsible use of technology (Senthilkumar, Jagajeevan, & Sangeetha, 2024).

Experts argue that libraries should move faster to adopt AI technologies (Arlitsch & Newell, 2017). However, the pace of adoption is uneven. While academic librarians in India are showing growing interest in tools like AI chatbots and virtual assistants, many public libraries are still limited by a lack of funds, infrastructure, and technical expertise (Kalbande et al., 2024). Several areas of library services are expected to undergo significant changes as a result of the adoption of AI. These include online services, digital reference support, automated cataloguing, and digital resource management. AI can also help improve data handling, enhance user experience, and support online learning. Librarians and institutions that adopt these changes early will likely become leaders in this field (Lund et al., 2020).

There are also creative ways to use humanoid robots in Indian libraries. For example, robots can aid in community engagement, support teaching, assist library staff, and even transform traditional library systems by introducing entirely new service models (Hossain, 2025). AI-based chatbots are already being used in some libraries globally to answer user queries in real-time, and this could be adopted more widely in Indian libraries (Saldeen, 2020). Besides improving efficiency, AI can reduce workloads and make library services more responsive and accessible to users (Igbinovia & Okuonghae, 2021).

1.2. Implications of AI and Robots for Roles and Skills

Whenever we discuss AI and robots, a common concern is whether these technologies will eventually replace humans. Librarians are also facing a similar dilemma, as robots are becoming more efficient (Wood & Evans, 2018). For instance, Calvert (2017) discusses two possibilities: either robots as a *helping hand* for librarians or robots as a *threat* to librarians. He envisions a symbiotic relationship between the two, albeit one led by humans. A hopeful viewpoint suggests that AI can take over repetitive and routine tasks, allowing librarians to focus on more meaningful and complex work (Fernandez, 2016). For example, AI tools can assist in generating metadata, such as subject headings or classifications, while librarians oversee the process to ensure privacy and adherence to ethical standards. Human supervision is important to ensure fairness and proper representation in information organization (Corrado, 2021).

This need comes at a time when the role of libraries in communities is being rethought. Libraries are now expected to act as community centres for lifelong learning and civic engagement. As libraries transition to more hybrid environments, where humans and machines collaborate, it becomes crucial to identify the skills librarians will need to work effectively and ethically in this new setup (Wilson & Daugherty, 2018). For new professionals entering the LIS field, understanding AI and its impact on information services will be essential (Cox, 2021). This means that library and information science curricula in universities need to be updated to reflect the current state of the field. At the same time, current librarians will need ongoing professional development opportunities to stay current with these rapidly evolving technologies. Beyond AI and robotics, emerging technologies such as drones, 3D printing, and augmented reality also hold great potential for library services in India. However, Indian LIS schools and training institutes must revise their curricula to truly benefit from these advancements and provide ongoing development programs for professionals in the field.

Global studies on AI and education have highlighted a lack of adequate training in AI in higher education (Hossain, Biswas & Khan, 2025). Recognising the need for a trained workforce, some top LIS institutions have begun integrating AI and related disciplines into their course curricula. For example, the Information School at the University of Sheffield explicitly highlights AI, data governance, and digital transformation as key themes in its courses. Likewise, the School of Information Science (iSchool) at the University of Illinois offers AI-adjacent electives, such as data science, natural language processing, information retrieval, etc. However, these examples are preliminary and insufficient to meet future needs.

LIS education in India is also not proactive in adopting AI and related technologies into its curriculum. It appears more scattered and vaguer due to a lack of a clear and shared understanding of what these technologies are and how they can be used. Moreover, the existing stakeholders and policymakers are not sufficiently trained in these emerging technologies. As a result, they remain skeptical and proceed cautiously as they attempt to adjust to the changing environment. Training the trainers should be the first step to encourage the adoption of AI in LIS education.

2. Recent Trends in LIS Education in India

Like many other countries, LIS education in India has faced several challenges in recent years. The COVID-19 pandemic brought sudden shifts in higher education, exposing gaps in online teaching, infrastructure, and curriculum delivery (Aslam, Sonkar & Owan, 2021). Although India has many LIS departments across central, state, and private universities, their course curriculum and quality of education are still not up to date (Kumar & Sharma, 2009). Similar to trends in countries such as Australia and the UK, there has been a gradual decline in comprehensive LIS programs, particularly those that strike a balance between theory and practice. This reduction raises concerns about the long-term sustainability of the LIS discipline, particularly as newer, fast-growing fields such as data analytics and digital humanities compete for attention and resources within academic institutions (Wang, 2018). Indian libraries are at different stages of their digital transformation. While many large academic and national libraries have adopted digital technologies, such as e-resources, digital repositories, and RFID systems, others struggle with basic automation, especially in rural or underfunded areas (Jeevan, 2004). Despite these challenges, Indian librarians have demonstrated great adaptability, continually learning and upgrading their skills to meet the changing demands (Negi & Srivastava, 2014).

The LIS education must evolve to prepare future professionals with the skills required for today's information-rich, technology-driven world. The curriculum should be dynamic, grounded in both theory and practical experience, and responsive to emerging challenges (Koltay, 2017). The LIS programs must clearly define their role and value, especially when working alongside newer interdisciplinary programs. Courses should include measurable outcomes and focus on both traditional library knowledge and emerging fields, such as data curation, digital literacy, and instructional design (Bawden & Robinson, 2012). Many LIS departments in India already offer specializations, such as digital libraries, knowledge management, or information systems, while still providing a strong foundation in core areas like cataloguing, reference services, and classification (Mushtaq & Ahmad, 2021). While much modern education focuses on data and algorithms, the LIS field brings a critical human dimension, governance, ethics, and user-centric information services. This is where LIS education in India can offer a unique contribution. For example, digital curation, preserving and organising data for the long term, is becoming increasingly important in academic and research settings. LIS professionals are well-suited for this task because they possess a deep understanding of metadata, documentation standards, and user needs (Paul, 2024).

Moreover, as new roles emerge, LIS education must expand its focus beyond library walls. Including subjects like digital ethics, inclusive information design, and responsible AI will be essential. This is especially important in a diverse country like India, where the design of information systems must respect local cultures, languages, and access issues (Warjri, Gogoi & Longkumer, 2024). LIS curricula in India must be reviewed and updated to facilitate this. Faculty must be trained in emerging technologies, and continuing education programs should be strengthened. Strong connections must also be maintained between academic institutions and the real-world needs of the profession (Gadagi, 2020). In short, the future of LIS education in India depends on how well it can balance tradition with innovation and theory with practice, so that tomorrow's information professionals are not just keeping up with change, but actively shaping it.

3. Research Questions

This study is based on the following key research questions, focusing on the Indian context of LIS education:

1. How many Indian academic institutions offer LIS courses that include topics related to AI and robotics in their syllabus or course descriptions?
2. What are the possible opportunities to revise and align LIS curricula in India to include emerging areas like AI and robotics better?

These questions aim to explore how well LIS education in India keeps pace with the technological advancements reshaping libraries and information services today. The goal is to identify gaps and suggest improvements that can help prepare future librarians and information professionals for the digital age.

4. Methodology

A review of postgraduate LIS course curricula was conducted to understand how topics such as AI and robotics are incorporated into LIS education in India. The focus was on analyzing the course content offered by selected Indian academic institutions that run master's programs in LIS or related fields.

Step 1: Selection of Academic Institutions and Courses

First, we identified master's degree programs in LIS from leading Indian academic institutions. Courses known for their structured curriculum and focus on emerging technologies were shortlisted. These included programs such as:

- Indira Gandhi National Open University (IGNOU)
- University of Delhi
- Documentation Research and Training Centre (DRTC), Bengaluru
- Banaras Hindu University
- Aligarh Muslim University
- Osmania University, Hyderabad

Step 2: Collection of Documents

Next, we examined documents related to LIS courses available in the public-domain to collect information on their syllabi, learning outcomes, and subject descriptors, in order to study the course structure and its intended objectives. The emphasis was to identify the extent to which AI and related technologies are integrated into the LIS curriculum.

Step 3: Analysis Criteria

All subjects in the selected programs were carefully reviewed according to the following inclusion criteria. A subject was selected for deeper analysis if its content mentioned at least one of the following themes:

- Artificial Intelligence or robotics
- General or emerging technologies in libraries
- Digital technology or digitization
- Social issues linked to digital technologies (e.g., digital divide, digital culture)
- Data science or data analytics
- Information ethics or related ethical concerns in technology use

Subjects that matched these criteria were listed and organised in spreadsheets for easy comparison, with notes and comments for reference.

Step 4: Content Analysis

Finally, the selected subjects were analysed using a content analysis method described by Patton (2002). The aim was to understand how these subjects discussed AI, robotics, and related technologies, and how these topics were presented in context, whether they were treated as technical skills, theoretical knowledge, or part of ethical discussions. This helped us evaluate how well Indian LIS education prepares future professionals for a technology-driven information environment.

5. Findings: Subject Descriptions

A total of 88 subject syllabi from 6 Indian academic institutions were reviewed based on the inclusion criteria. Among these, only one subject syllabus mentioned AI. It referred to the challenges of utilizing AI and the increasing reliance on algorithms. This subject also aimed to cover related ethical concerns. However, none of the subjects made any reference to robotics.

Study Inclusion Criteria	N	Percent
AI	1	01.14
Robotics Technology	0	00.00
Digital Technology	78	88.64
Digital-social phenomena	42	47.73
Digital Divide	25	28.41
Data Management and Analytics	59	67.05
Information and Data Literacy	22	25.00
Information ethics	46	52.27
Ontological Studies & Semantic Web	28	31.82
Total	**88**	**100.00**

Table 1: The number and percentage of subject descriptions mentioned AI, robotics, and allied terms

The most common topic in these subject descriptions was the broad term '*technology*', followed by more specific mentions of '*digital technology*' with 88.64%. This shows that LIS education in India is gradually adapting to the growing use of modern technologies in libraries and information services. This is especially important in the current scenario, where access to digital resources has become necessary during the COVID-19 pandemic, with libraries playing a critical role in supporting education and information delivery during lockdowns.

Around 67.05% of the subjects mentioned '*data*', mainly in the context of data management and data analytics. Some also referred to metadata, while fewer subjects addressed data preservation and the social implications of data, such as its role in culture or heritage. The focus remains on practical uses of data and how to manage, analyse, and use it effectively. The recurring phrase '*information ethics'* appeared in 52.27% of subject descriptions, indicating that educators are trying to highlight multiple challenges linked to technology use in libraries. This could reflect a growing awareness of data privacy, information ethics, and responsible technology use, topics that are becoming increasingly important in library and information science education. AI literacy for different stakeholders is a current mandate to understand the ethical issues associated with AI and embrace it in the most responsible, effective, and unbiased way (Lachheb, Leung, Abramenka-Lachheb, & Sankaranarayanan, 2025).

About 47.73% of the subject descriptions included '*digital-social phenomena*', which generally referred to the effects of technology on society and library practices. These include the digital divide, the information society, and legal or ethical challenges associated with digital technologies. These topics frequently arose in discussions about how technology is transforming the work culture in libraries, user services, and the broader information environment. 31.82% of the subjects mentioned the term '*Ontological Studies & Semantic Web'*, which was the key aspect of information management in the digital environment. 26.41% mentioned the '*digital divide'*, the most common information access problem in developing countries. 25.00% mentioned *'information and data literacy'*, which is crucial in this data-driven world.

There were also a few unique mentions, such as the impact of digital tools on school education, website design based on user behavior, and digital data as part of cultural heritage – but these appeared only once.

6. Discussion

One of the most important findings from this analysis is that almost none of the LIS courses in India include AI and robotics as core subjects. In many academic institutions, while these emerging technologies may not be taught as standalone topics, some subjects related to technological innovation may briefly touch upon them. Subjects like digital libraries, information systems, or IT in LIS might discuss AI in passing. Still, there is no strong focus on how AI and robotics practically transform the library profession (Yadav & Gohain, 2015). Many LIS subjects today discuss the influence of technology on society and its users, but the focus remains on the social and ethical impacts (Huang, Samek, & Shiri, 2021). While this is important, it is also necessary for students to understand how the technology itself functions, especially AI algorithms, machine learning models, and automated systems. For instance, understanding how AI affects users' information-seeking behaviour or how algorithms can shape what information is seen or missed is critical for future professionals (Biagini, 2025).

To truly grasp the impact of AI, students must understand how AI works and its limitations. For example, algorithms can be biased or flawed depending on the kind of data they are trained on. The AI will produce misleading results if the data is incomplete, outdated, or skewed. This is where concepts like Simpson's Paradox and "Garbage In, Garbage Out" come into play. These concepts highlight how poor data quality can lead to inaccurate or even harmful AI decisions (Domingos, 2015). Moreover, AI often performs well with training data but may fail when applied to real-world scenarios. This mismatch can be particularly hazardous in professional environments, especially when decisions made by AI impact users' access to information, privacy, or trust in the library system. Therefore, LIS professionals must be trained to use AI tools and critically evaluate them (Malakar, Manavalan, Jain & Sareen, 2025).

Another primary concern is bias in algorithms, which often begins right at the data preparation stage.

Unless we actively address these biases, they will continue to affect the results (Binns, 2018). Given the Government of India's push towards digital transformation under schemes like Digital India, it is only a matter of time before AI becomes a regular feature in public and academic libraries. Hence, LIS professionals must be prepared. Therefore, Indian LIS educators and professional bodies, such as RRRLF, IASLIC, and ILA, must strongly advocate for integrating AI-related content into academic syllabi. The emphasis must be on capacity-building to effectively use AI tools, along with the necessary ethical considerations. LIS professionals should learn about the technologies behind AI tools, auditing techniques to address biases/conflicts of interest, and implementation strategies that will enable them to adapt and coexist within the evolving technological landscape. (Veeranjaneyulu, 2023). Future applications of AI in Indian LIS practice could include:

- Automated cataloguing using AI and metadata
- Smart recommendation systems for users
- AI-based subject classification
- Voice-assisted library queries
- Predictive analytics for user needs
- Chatbots for digital reference services

However, another challenge arises: many of our classification systems and metadata schemes already carry historical and cultural biases. If these flawed systems are used to train AI models, the AI will replicate those biases (Noble, 2018). So, LIS professionals face a dual responsibility: first, to clean and improve existing data systems; and second, to ensure AI tools are trained and tested with inclusive, diverse, and contextually appropriate data. Hence, we must not only focus on removing biases in classification and subject headings, but also on promoting ethical and responsible use of AI in our professional practices (Floridi, 2013). As we move towards greater use of technology, we must equip LIS students and professionals with technical knowledge and ethical judgment (Senthilkumar, Jagajeevan & Sangeetha, 2024).

As AI and robotics continue to grow and influence society, their impact on the library and information profession will be significant as well. However, this impact may vary from person to person and from one workplace to another. That is why tomorrow's professionals must be well-prepared to understand how AI works and use their knowledge to shape the future of libraries and information services in India in an ethical, efficient, and people-centered manner. In conclusion, integrating AI and robotics into LIS education is not just a matter of keeping up with technology; it is about empowering future professionals to lead confidently, adapt to change, and ensure that technology serves the best interests of users and society.

6.1. Implications for Best Practices: Opportunities for LIS Education

As we move towards a future increasingly shaped by AI and robotics, there is growing concern about the role the LIS profession and its education system will play. However, this changing landscape also brings exciting opportunities. The LIS field in India can position itself as a key player at the crossroads of human information behaviour, ethical information use, and the responsible application of AI and robotics. Here are some timely and practical suggestions to make the most of these opportunities:

i. **Understand Stakeholder Perspectives:** A comprehensive study should be conducted to understand how various LIS stakeholders, including librarians, educators, students, and professional associations, perceive the opportunities and challenges presented by AI and robotics. This will help shape practical strategies and solutions that reflect the real needs of the profession.

ii. **Integrate AI Content into the Curriculum:** More AI-related content should be included in LIS courses. Students should learn how AI works, its limitations, and its benefits and risks. Collaboration with other fields, especially computer science and data science departments, will be essential to building a strong interdisciplinary foundation.

iii. **Encourage Research on AI and Robotics in LIS:** More research is needed on how AI and robotics impact library services, information behavior, and society. Such research will help design course content that is theoretical and relevant to everyday professional practice in India.

iv. **Create and Share Open Educational Resources (OER):** Developing freely accessible learning materials and resources on AI and robotics for LIS can help reduce the workload on individual teachers and institutions. Sharing these resources among universities, colleges, and libraries in India will promote collective learning and reduce duplication of efforts. India could also emerge as a global leader in developing and sharing OER for LIS education in the AI age.

v. **Strengthen Ethics and Data Literacy in LIS Education:** As AI systems rely heavily on data, teaching students how to handle data responsibly is important. Courses should include a strong foundation in information ethics, so that future professionals are prepared to address the ethical issues associated with the use of AI technologies in information services.

vi. **Promote Continuing Education and Skill Development:** Working professionals in libraries and information centres should be allowed to upgrade their skills through short-term courses, micro-credentials, or continuing professional development programmes focusing on AI, robotics, and digital tools relevant to LIS.

vii. **Build Communities of Practice:** Creating collaborative groups where professionals can share knowledge and experiences can help everyone learn and grow together. LIS associations can support these Communities of Practice and focus on digital preservation, AI in information services, or ethical data management. International examples, such as the Digital Preservation Coalition, can inspire similar networks in India.

Implementing these suggestions will require time, funding, and collaboration from many stakeholders, educational institutions, government bodies, library associations, and technology experts. Cross-institutional partnerships can be encouraged with seed funding and shared projects. Regular curriculum reviews in colleges and universities should include input from employers and industry professionals to ensure relevance and future-readiness. India's LIS community must actively participate in research, conferences, and Special Interest Groups on AI and technology in libraries to keep pace with global trends and local needs. With thoughtful planning and collaborative efforts, India's LIS profession can adapt to the AI revolution and lead it responsibly and ethically.

6.2. Limitations and Scope for Future Research

While this study offers valuable insights, it is essential to acknowledge certain limitations. The information used for the analysis was mostly collected from official course websites, which may not fully capture what is being taught in classrooms. New and emerging topics, such as AI, may not yet be clearly reflected in these online descriptions. Another limitation is that the study focuses on only a few LIS departments, and the sample size is relatively small. To create a comprehensive picture, an extensive and detailed work is required. For future studies, researchers could expand the scope by comparing LIS education systems across different countries, particularly in regions such as South Asia. This would help gain a deeper understanding of how AI and robotics are being introduced and taught globally in LIS courses. Such comparisons could also guide Indian academic institutions in updating their LIS curriculum to keep pace with international developments.

8. Conclusion

This paper examines the current status of how AI, robotics, and related topics are incorporated into postgraduate-level LIS courses in India. Our review of course curricula revealed that only one subject explicitly mentioned AI, and none mentioned robotics. However, many courses did address related themes, such as technology, digital tools, the digital divide, data analytics, and information ethics. This suggests that there is scope to integrate AI and robotics more intentionally into the LIS curriculum. We have also discussed the broader concerns and possibilities that AI brings to the LIS profession, its education system, and society in general. Based on our findings, we have suggested several opportunities for strengthening LIS education. In the Indian context, where digital transformation in libraries is still in its early stages of evolution, these suggestions are especially relevant. As AI becomes increasingly prevalent in all areas of life, the LIS profession must prepare to play a leading role in utilizing these technologies wisely and ethically. The future will undoubtedly see more advanced uses of AI in information management. Therefore, LIS education must equip future professionals with the necessary knowledge and skills to handle these changes and contribute meaningfully to society effectively.

References

Ajani, Y. A., Oladokun, B. D., Enakrire, R. T., Ogunjimi, B. E., Ilori, O. O., & Akin-Fakorede, O. O. (2024). Robots in public libraries: An empirical analysis of current applications and future opportunities. *Public Library Quarterly*, 1-20.

Arlitsch, K., & Newell, B. (2017). Thriving in the age of accelerations: A brief look at the societal effects of artificial Intelligence and the opportunities for libraries. *Journal of Library Administration*, 57(7), 789–798. https://doi.org/10.1080/01930826.2017.1362912

Asadzadeh, A., Mohammadzadeh, Z., Fathifar, Z., Jahangiri-Mirshekarlou, S., & Rezaei-Hachesu, P. (2022). A framework for information technology-based management against COVID-19 in Iran. *BMC Public Health*, *22*(1), 402.

Aslam, S., Sonkar, S.K., & Owan, VJ (2021). Changes in teaching and learning in higher education during COVID-19 lockdown: A study of LIS students in India. *Library Philosophy and Practice (e-Journal)*.

Bakar, A. A., Hussin, N., Angchun, P., & Seman, M. R. (2022). Roles of library in the metaverse. *Environment-Behaviour Proceedings Journal*, *7*(SI10), 269-273.

Bawden, D., & Robinson, L. (2012). *Introduction to information science*. London: Facet Publishing

Biagini, G. (2025). Towards an AI-Literate Future: A systematic literature review exploring education, ethics, and applications. *International Journal of Artificial Intelligence in Education*, 1-51.

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In *Conference on fairness, accountability and transparency* (pp. 149-159). PMLR.

Calvert, P. (2017). Robots, the quiet workers, are you ready to take over? *Public Library Quarterly*, 36(2), 167–172. https://doi.org/10.1080/01616846.2017.1275787

Cockshott, P., & Renaud, K. (2016). Humans, robots and values. *Technology in Society*, 45, 19–28. https://doi.org/10.1016/j.techsoc.2016.01.002

Corrado, E. M. (2021). Artificial Intelligence: The possibilities for metadata creation. *Technical Services Quarterly*, 38(4), 395–405. https://doi.org/10.1080/07317131.2021.1973797

Cox, A. (2021). *The impact of AI, machine learning, automation and robotics on the information professions: A report for CILIP*.

Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Basic Books.

Fernandez, P. (2016). Through the looking glass: Envisioning new library technologies' understanding artificial Intelligence. *Library Hi Tech News*, 33(3), 20–23. https:/doi.org/10.1108/LHTN-03-2016-0013

Floridi, L. (2013). *The ethics of information*. Oxford University Press (UK).

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and machines*, *28*(4), 689-707.

Frost, M., Goates, M., Cheng, S., & Johnston, J. (2020). Virtual reality: A survey of use at an academic library. *Information Technology and Libraries*, 39(1), 1. https://doi.org/10.6017/ITAL.V39I1.11369

Gadagi, M. V. S. (2020). Challenges and Future of LIS Education in India. *Aayushi International Interdisciplinary Research Journal*, *7*(3).

Hervieux, S., & Wheatley, A. (2021). Perceptions of artificial Intelligence: A survey of academic librarians in Canada and the United States. *Journal of Academic Librarianship*, 47(1), https://doi.org/10.1016/j.acalib.2020.102270

Hossain, Z. (2025). Integrating AI literacy into library and information science education and practice: a human-centered approach. *Library Hi Tech News*.

Hossain, Z., Biswas, M. S., & Khan, G. (2025). AI literacy of library and information science students: A study of Bangladesh, India and Pakistan. *Journal of Librarianship and Information Science*, 09610006241309323.

Huang, C., Samek, T., & Shiri, A. (2021). AI and ethics: Ethical and educational perspectives for LIS. *Journal of Education for Library and Information Science*, *62*(4), 351-365.

Igbinovia, M.O. & Okuonghae, O. (2021), Internet of Things in contemporary academic libraries: application and challenges. *Library Hi Tech News*, 38(5), 1–4. https://doi.org/10.1108/LHTN-05-2021-0019

Jeevan, V. K. J. (2004). Digital library development: identifying sources of content for developing countries with special reference to India. *The International Information & Library Review*, *36*(3), 185-197.

Jha, S. K. (2023). Application of artificial Intelligence in libraries and information centers services: prospects and challenges. *Library Hi Tech News*, *40*(7), 1-5.

Kalbande, D., Suradkar, P., Chavan, S., Verma, M. K., & Yuvaraj, M. (2024). Artificial intelligence integration in academic libraries: perspectives of LIS professionals in India. *The Serials Librarian*, *85*(5-6), 132-143.

Koltay, T. (2017). Data literacy for researchers and data curators. *Journal of Librarianship and Information Science*, 49(1), 3–10.

Kumar, K., & Sharma, J. (2009). *Library and information science education in India*. Har Anand Publications.

Lachheb, A., Leung, J., Abramenka-Lachheb, V., & Sankaranarayanan, R. (2025). AI in higher education: A bibliometric analysis, synthesis, and a critique of research. *The Internet and Higher Education, 67*, 101021. https://doi.org/10.1016/j.iheduc.2025.101021

Lund, B. D., Omame, I., Tijani, S., & Agbaji, D. (2020). Perceptions toward artificial Intelligence among academic library employees and alignment with the diffusion of innovations' adopter categories. *College & Research Libraries*, 81(5), 865–882. https://doi.org/10.5860/crl.81.5.865

Malakar, P., Manavalan, L., Jain, P., & Sareen, M.S. (2025). Embracing AI in libraries: a strategic approach for India's evolving library landscape. *Library Hi Tech News*, *42*(5), 15-18.

Mushtaq, M., & Ahmad, N. (2021). Curriculum based competencies imparted in lis schools: outlook of lis professionals working in special libraries. *Library Waves*, *7*(1), 31-39.

Negi, A. S., & Srivastava, J. P. (2014). Changing role of the academic library professionals in the digital era: trends, vision and challenges. *Indian Streams Research Journal*, *4*(1), 1-5.

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. In *Algorithms of oppression*. New York University Press.

Parirokh, M., Daneshgar, F., & Fattahi, R. (2009). A theoretical framework for development of a customer knowledge management system for academic libraries. *Spanish Journal of Agricultural Research*, 1-16. https://www.ifla.org/past-wlic/2009/195-parirokh-en.pdf

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Sage.

Paul, N. (2024). The Future of Information Science: Exploring the Role of Libraries In Data Management and Digital Curation. *Library of Progress-Library Science, Information Technology & Computer*, *44*(3).

Raju, J. (2020). Future LIS education and evolving global competency requirements for the digital information environment: An epistemological overview. *Journal of Education for Library and Information Science*, *61*(3), 342-356.

Saldeen, N. A. (2020). Artificial intelligence chatbots for library reference services. *Journal of Management Information and Decision Sciences*, 23(1).

Senthilkumar, K. R., Jagajeevan, R., & Sangeetha, S. (2024). Impact of AI on library and information science in higher institutions in India: A comprehensive analysis of technological integration and educational implications. In *AI-assisted library reconstruction* (pp. 21-33). IGI Global Scientific Publishing.

Subaveerapandiyan, A. & Gozali, A. (2024). AI in Indian Libraries: Prospects and Perceptions from Library Professionals. *Open Information Science*, 8(1), 20220164. https://doi.org/10.1515/opis-2022-0164

Tang, R., & Sae-Lim, W. (2016). Data science programs in US higher education: An exploratory content analysis of program description, curriculum structure, and course focus. *Education for Information*, *32*(3), 269-290.

Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2013). Academic librarians and research data services: preparation and attitudes. *IFLA journal*, *39*(1), 70-78.

Veeranjaneyulu, K. (2023). Restructuring LIS education to meet the emerging digital library and information technology environment. In *Proceedings of National Conference on Exploring the Past, Present, and Future of Library and Information Science* (pp. 38-47).

Wang, L. (2018). Twinning data science with information science in schools of library and information science. *Journal of Documentation*, *74*(6), 1243-1257.

Warjri, D., Gogoi, G., & Longkumer, S. T. (2024). Designing Inclusive Digital Learning Environments in Libraries for Patrons with Special Needs. Available at *SSRN 4924156*.

Wheatley, A., & Hervieux, S. (2019). Artificial Intelligence in academic libraries: An environmental scan. *Information Services & Use*, 39(4), 347–356. https://doi.org/10.3233/isu-190065

Wilson, H. J., & Daugherty, P. R. (2018). Collaborative Intelligence: Humans and AI are joining forces. *Harvard Business Review*, 96(4), 114–123.

Winkler, R., & Söllner, M. (2018). *Unleashing the potential of chatbots in education: a state-of-the art analysis*. Academy of Management Annual Meeting (AOM), Chicago, USA.

Wood, B. A., & Evans, D. (2018). Librarians' perceptions of artificial Intelligence and its potential impact on the profession. *Computers in Libraries*, 38(1), 116–124. https://doi.org/10.1177/0266382120952016

Yadav, A. K., & Gohain, R. R. (2015). Growth and development of LIS education in India. *SRELS Journal of Information Management*, *52*(6), 403-414.

Yoon, J. W., Andrews, J. E., & Ward, H. L. (2021). Perceptions on adopting artificial Intelligence and related technologies in libraries: Public and academic librarians in North America. *Library Hi Tech*, 37(3), https://doi.org/10.1108/LHT-07-2021-0229

About Authors

Dr. Mohd Shoaib Ansari is presently serving as Assistant Director at the Directorate of Higher Education, Chhattisgarh. He has over ten years of professional experience in the field of Library and Information Science. His work has made significant contributions to library automation, digital resource management, and the development of information systems, with a particular focus on underserved and marginalized communities. He was awarded a Ph.D. by Banaras Hindu University, with his doctoral research titled *"Developing a Need-Based Information System for the Weaver Community*," reflecting his strong commitment to socially relevant and community-oriented research. A prolific researcher, he has published over 20 peer-reviewed articles in reputed Scopus and UGC-CARE indexed journals. His research interests include e-learning, mobile learning applications, digital repositories, library social media, and traditional knowledge systems.

Dr. Sneha Tripathi is currently serving as Deputy Librarian at the Central Library, Banaras Hindu University, Varanasi, with over 18 years of professional experience. She holds a postgraduate degree in Documentation and Information Science from the Indian Statistical Institute and a Ph.D. in Library and Information Science from the University of Kalyani. Her research interests include digital libraries, metadata, semantic web technologies, and multilingual information retrieval systems. She has received the prestigious VLIR-UOS scholarship, Brussels, Belgium, and has published extensively in reputed Scopus-indexed journals. Her professional interests include library automation, ERM, research support services, information literacy, and scholarly communication, and she actively contributes to professional development through training and academic initiatives.

Dr. Punam Chauhan is an Associate Professor (Library) and Librarian at Government Degree College, Gosaikheda, Unnao. She holds postgraduate degrees in Chemistry and Library and Information Science, as well as a B.Ed. from Pt. Deendayal Upadhyay Gorakhpur University, Gorakhpur. A UGC-NET qualified scholar, she earned her Ph.D. from Banaras Hindu University. She has published over 21 peer-reviewed research articles in reputed Scopus and UGC-CARE indexed journals. Her research interests include information literacy, job satisfaction, interpersonal behaviour, e-learning, mobile learning, digital repositories, social media, and information-seeking behaviour. She has also contributed to the development of research scales and evaluative studies. She is the author of six books and has contributed conference papers and book chapters. Alongside her academic pursuits, she actively engages in social awareness initiatives.

Blind Peer Reviewed
(Previously UGC-CARE)
ISSN: 0972-2750

The Opportunity of Metaverse in the Domain of Indian Higher Education Perspectives

Rajib Roy & Ashis Kumar Samanta

Manuscript Received on –
25th December, 2026
Primary Reviewed on –
6th January, 2026
Peer Reviewed on –
24th January, 2026
Revised on –
27th January, 2026
Accepted for Publication
3rd February. 2026

ABSTRACT

Introduction: *The National Education Policy (NEP) 2020 represents a landmark reform in India's education system, emphasizing multidisciplinary learning, digital integration, and equitable access for diverse learners. Emerging technologies such as the Metaverse, powered by virtual reality (VR) and augmented reality (AR), provide new opportunities to actualize NEP's vision by transforming traditional classrooms into immersive, interactive learning environments.*

Purpose: *The purpose of this paper is to explore the potential of the Metaverse as a tool to advance NEP 2020 goals and strengthen India's educational ecosystem.*

Research Problem: *Despite its potential, the integration of Metaverse technology faces challenges related to infrastructure, high cost, data privacy, and teacher preparedness, which hamper its inclusive implementation.*

Objective: *The objective is to analyze the role of the Metaverse in enhancing experiential learning, collaboration, and personalization, while identifying barriers and proposing strategies for effective implementation.*

Methodology: *The research employs a combination of literature review, case study analysis, and policy evaluation to examine global practices and contextualize them within India's educational framework.*

Findings: *The study shows that the Metaverse can improve science, history, and vocational education by using simulations and interactive learning. It helps student engagement, skill development, global collaboration, and personalized learning. However, proper infrastructure, lower costs, ethical use and teacher training are essential for successful adoption under NEP 2020.*

Keywords: *Metaverse, NEP 2020, Immersive Learning, Experiential Education, Digital Transformation*

Dr. Rajib Roy
Assistant University Librarian
University of Calcutta Kolkata, India
rroy@caluniv.ac.in

Dr. Ashis Kumar Samanta
Database Administrator
University of Calcutta Kolkata, India
aksdba@caluniv.ac.in, aksresearchcu@gmail.com

Introduction

India's education landscape has undergone significant reforms with the introduction of the National Education Policy 2020, approved by the Union Cabinet on July 29, 2020. NEP 2020 aims to overhaul the education system to make it more holistic, flexible, multidisciplinary, and aligned with the needs of the 21st century. It envisions an education framework that fosters critical thinking, creativity, and innovation while ensuring equity and inclusion for all learners, including those from socio-economically disadvantaged groups (SEDGs). Central to NEP 2020 are principles such as technology integration, experiential learning, and multidisciplinary approaches, which aim to bridge gaps in access and quality between urban and rural areas (Education & Pedagogicstructures, 2024). The policy emphasizes the extensive use of technology in teaching, learning, and administration to enhance efficiency and accessibility. For example, it promotes digital infrastructure, such as the Digital Infrastructure for Knowledge Sharing (DIKSHA) platform, virtual labs, and online resources, to support foundational literacy, numeracy, and vocational education. NEP 2020 also advocates for coding, computational thinking, and digital literacy from the early stages, alongside experiential methods such as hands-on activities, arts-integrated learning, and internships (Education & Pedagogicstructures, 2024). Multidisciplinary approaches are a cornerstone, eliminating rigid disciplinary boundaries and encouraging flexible subject choices.

Figure 1: NEP 2020 and Metaverse in Education

The Metaverse, evolving from science fiction into a tangible, immersive digital universe that integrates physical and virtual worlds, enables persistent user interactions through avatars, supported by technologies such as VR, AI, blockchain, and NFTs. Initially dominated by computer science and engineering research in the 1990s–2000s, it expanded into an interdisciplinary field from the 2010s across business, the social sciences, education, health, law, the arts, and more, accelerated by post-truth societal shifts and post-2020 technological advancements (Figure 1). This ecosystem combines foundational infrastructure, device independence, decentralization, digital twins, and human-like interactions to reshape social life, economy, education, entertainment, and governance, presenting profound social, ethical, financial, and environmental implications while requiring ongoing interdisciplinary research to address emerging challenges (Qi, 2025).

This integrative literature review explores the transformative potential of the Metaverse in education, focusing on learning/teaching, peer learning, and social aspects, based on an analysis of more than 35 studies and interviews with researchers. It highlights benefits like gamified experiences, inclusion, personalized skills, virtual campuses, and immersion, while noting challenges including cybersecurity, distractions, cyberbullying, and usability. Despite these advantages, significant research gaps remain in understanding long-term educational effects and sociological impacts within Metaverse environments (Mauz & Diaz, 2024).

The benefits of Metaverse in education are well-documented. It promotes personalized learning by adapting to individual learning styles, and enhancing retention through interactive avatars and gamification. In India, where student diversity is high, the Metaverse can facilitate inclusive education for Divyang students via assistive VR tools, resonating with the NEP's equity focus. Furthermore, it enables global collaboration, allowing Indian students to interact with peers worldwide, supporting NEP's emphasis on global citizenship education (Alfaisal, Hashim & Azizan, 2024). Challenges are equally prominent in the literature. Ethical concerns, including data privacy, cyberbullying, and digital addiction, pose risks in Metaverse environments. Infrastructure barriers in India, such as limited high-speed internet access in rural areas, hinder widespread adoption, and potentially widen the digital divide, contrary to the NEP's inclusive aims. Teacher readiness is another issue; many educators lack skills in virtual facilitation, necessitating robust training programs (Mukred et al., 2025 & Dutta, 2025).

Literature Review

A paper by Acharjee, Bhattacharjee & Roy (2025) explores the role of Metaverse technologies in faculty development for NEP 2020's multidisciplinary education goals. The analysis of data from 450 faculty members across 15 Indian universities reports substantial gains in interdisciplinary teaching confidence (78%), collaborative research (65%), and technology skills (82%). Metaverse-based training cuts costs by 45% and boosts participation by 67%, offering a scalable framework for India's educational reform. A paper by Dutta (2025) analyzes the transformative potential of India's National Education Policy (NEP) 2020 through integration with generative AI and immersive technologies like AR/VR. It highlights how these advancements align Indian education with global Sustainable Development Goals, particularly SDG 4, while addressing local socio-economic challenges. By fostering personalized, engaging, and equitable learning experiences, NEP 2020 enables institutions to bridge urban-rural divides, enhance student outcomes, and position India as a global knowledge hub, offering practical implications for policymakers and educators in achieving sustainable educational transformation.

Kali (2025) explores through collaborative efforts among educators, policymakers, and technologists, how the Metaverse can evolve into a powerful tool for equitable and inclusive education. In an article, Panda (2025) explores the symbiotic relationship between technology and open distance learning (ODL), highlighting how globalization has driven advances from print to AI, VR/AR, the Metaverse, and blockchain, thereby enhancing accessibility, personalization, interactivity, and collaboration in lifelong learning and professional development. Despite these transformations, traditional and networked learning models coexist, particularly in the Global South. Shivani and Kumari (2025) proposed that India's higher education research landscape is evolving through massification, quality enhancement, and NEP 2020's push for innovation, multidisciplinary integration, technology adoption, industry collaboration, and equity amid rural-urban and funding disparities.

A paper by Sing (2024) emphasizes NEP 2020's holistic education vision in India, fostering values, duties, and global citizenship while cautiously integrating the metaverse/eduverse for immersive learning. It emphasizes ethical oversight, inclusivity, real-virtual balance, and educators' roles amid disparities, drawing on a Gen Z pilot study that advocates for mindful tech policies in higher education. A study by Vijitha and Anandan (2025) examines post-COVID digital education disparities in Tamil Nadu's Kanchipuram district, proposing Metaverse-driven solutions integrated with 5G, LoRaWAN, AI tools like Duolingo, and gesture-based authentication to bridge urban-rural gaps, enhance literacy, and ensure secure, inclusive learning for marginalized communities through pilot programs and supervised machine learning. An article by Krishnan, Lamba, and Mariappan (2024) highlights the transformative potential of metaverse technologies in Indian education for personalized, immersive learning and enhanced student engagement, while stressing the need to overcome barriers like inadequate teacher training, infrastructure deficits, data privacy risks, mental health concerns, and the digital divide through policy reforms and ethical safeguards. The qualitative study by Dhillon and Tinmaz (2024) investigates the impact of avatar design on virtual education in the Metaverse through structured interviews with 20 diverse participants (10 males, 10 females). Using thematic coding, it explores how avatar attributes influence learning experiences, engagement, and interactions. The research by Sa and Serpa (2023), using document analysis examines the Metaverse's inevitable role in a digitalized society and its profound impact on formal/informal teaching-learning across initial and continuing education. It reflects on the challenges and opportunities the Metaverse presents for creating more relevant, effective pedagogical processes.

Overall, the literature underscores the Metaverse's alignment with NEP 2020 but calls for balanced strategies to address challenges. This review synthesizes global and Indian perspectives, setting the stage for deeper discussion.

Outcomes of the Literature Review

Metaverse-based training led to a 78% increase in confidence in interdisciplinary teaching, a 65% increase in collaborative research, an 82% increase in technology integration, a 45% reduction in costs, and a 67% increase in participation rates (Acharjee, Bhattacharjee & Roy, 2025). Merging NEP 2020 with advanced digital technologies revolutionizes traditional teaching, enhances student engagement, bridges urban-rural educational gaps, and delivers personalized, immersive learning experiences that advance sustainable development (Dutta, 2025). The Metaverse provides continuous, immersive virtual environments that overcome geopolitical limitations, significantly enhancing individualized and collective learning. It has emerged as a key educational tool post-COVID, improving engagement through real-time, non-face-to-face interactions (Kali, 2025). Technological innovations in AI, cloud computing, VR/AR, and analytics significantly improve personalized, engaging ODDBL while addressing learning gaps and scaling operations (Panda, 2025). NEP 2020 drives a shift toward innovative, multidisciplinary, and industry-linked research to boost excellence and global impact (Shivani & Kumari, 2025). A qualitative pilot study with Gen Z learners highlights concerns about unchecked adoption of the eduverse, urging ethical, contextual tech integration to prevent unethical deviations and ensure learner well-being (Singh, 2024). Combining Metaverse with low-energy LoRaWAN, 5G, and AI personalization significantly improves accessibility, engagement, and literacy outcomes in rural Kanchipuram while addressing infrastructure limitations (Vijitha & Anandam, 2025). Metaverse integration offers significant opportunities to improve educational outcomes and meet diverse learner needs via engaging, personalized experiences in Indian schools (Krishnan, Lamba & Mariappan, 2024). Avatar design elements like gender, facial features, attire, and personalization significantly shape emotional connections, self-expression, engagement, and socialization in virtual learning, though opinions vary on educational efficacy (Dhillon & Tinmaz, 2024). The Metaverse offers significant opportunities to enhance teaching relevance and effectiveness through immersive, innovative experiences, while posing challenges requiring careful implementation and monitoring (Sa & Serpa, 2023).

Research Gap

The study highlights substantial benefits of Metaverse training; it lacks longitudinal data on sustained impact and broader scalability across diverse institutional contexts in India (Acharjee, Bhattacharjee & Roy, 2025). While NEP 2020 encourages technology adoption, the study identifies a need for more structured guidelines and empirical frameworks to effectively integrate generative AI and immersive tools for achieving SDG 4 (Quality Education) at scale (Dutta, 2025). The article identifies critical barriers, including cost, data privacy, security risks, and accessibility issues, that hinder widespread equitable adoption. It lacks in-depth empirical evidence or case studies on actual implementations, limiting insights into practical solutions and long-term outcomes in diverse educational contexts (Kali, 2025). While pedagogy remains central, the passage notes persistent challenges, such as privacy, surveillance, and unequal implementation, in developing countries, which require greater focus on contextual equity and the integration of ethical AI (Panda, 2025). Persistent challenges in funding, rural-urban inequities, and limited internationalization hinder the full realization of research potential and require targeted professional management strategies (Shivani & Kumari, 2025). Limited nationwide perspectives on accounting necessitate comprehensive policies for the thoughtful integration of technology into Indian higher education and address digital disparities and diverse learner needs (Singh, 2024). Persistent rural connectivity deficits, limited local-language AI tools, and security vulnerabilities in online platforms hinder equitable digital education, necessitating targeted hybrid technologies and policy interventions (Vijitha & Anandam, 2025). Critical shortcomings in teacher preparedness, infrastructure equity, ethical data handling, and access for underserved populations that hinder the widespread, inclusive adoption of the metaverse necessitate targeted interventions and further longitudinal research (Krishnan, Lamba & Mariappan, 2024). Persistent concerns regarding privacy, data security, and the Metaverse's effectiveness as an educational platform underscore the need for additional research on long-term impacts and the development of standardized ethical guidelines (Dhillon & Tinmaz, 2024). There is a critical need for systematic research studies to track, evaluate, and guide Metaverse integration in education, as current understanding relies heavily on theoretical reflection rather than empirical follow-up (Sa & Serpa, 2023).

Statement of the problem

The optimal level of Metaverse integration into India's education system will be determined within a specific framework, maximizing educational benefits (such as enhanced experiential learning, personalization, multidisciplinary collaboration, student engagement, and skill development) while minimizing costs, risks, and challenges (such as infrastructure deficiencies, high implementation costs, data privacy risks, teacher unpreparedness, increased digital divide, and ethical concerns).

Objectives of the Study

- To explore how the Metaverse can improve higher education in India by making learning more engaging, interactive, and aligned with NEP 2020.
- To identify the main challenges, such as cost, infrastructure, teacher training, and data privacy, that affect the use of the Metaverse in universities.
- To propose practical solutions and strategies for successfully using Metaverse technology technology in Indian higher education in an inclusive and effective manner.

Research Methodology

This article employs a qualitative, conceptual research approach based on a systematic literature review and policy analysis. Given the exploratory nature of Metaverse integration in NEP 2020, a review methodology is appropriate for synthesizing existing knowledge and proposing frameworks (Figure 2).

Data sources include peer-reviewed articles, policy documents, and reports accessed via web searches on topics like "Metaverse applications in education, India NEP 2020," and "Benefits and challenges of Metaverse in implementing NEP 2020." The NEP 2020 document was summarized for key themes using browsing tools.

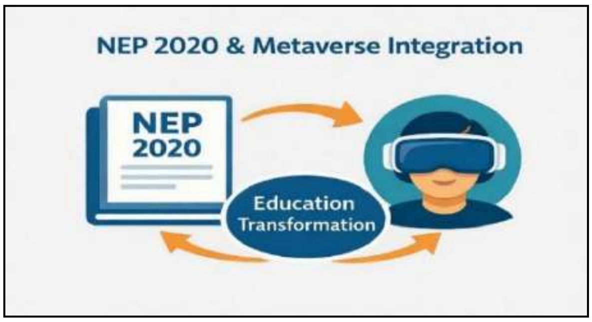


Figure 2: NEP 2020 and Metaverse in Integration

Inclusion criteria: Publications from 2020-2025 focusing on Metaverse in education, with relevance to India or NEP 2020. Exclusion: Non-educational or pre-2020 studies.

The analysis involved thematic coding of benefits (e.g., immersion, personalization), challenges (e.g., ethics, access), applications (e.g., experiential learning), and case studies. This conceptual framework maps Metaverse features to NEP pillars, informing recommendations.

Limitations include reliance on secondary data; future research could involve empirical surveys of Indian educators.

Let the following symbols represent key quantifiable factors:

c: Total implementation cost (including infrastructure, VR/AR devices, content development in Indian languages, and training programs).

b: Per-student benefit factor (capturing improvements in engagement, retention, critical thinking, and learning outcomes from immersive experiences).

q: Quality enhancement multiplier (representing personalization and experiential learning gains, e.g., from virtual simulations and multidisciplinary projects).

n: Scale of adoption (number of students or institutions reached).

r: Aggregate risk factor (encompassing data privacy breaches, cyberbullying, health issues like VR fatigue/addiction, and ethical concerns).

e: Equity index (measure of inclusive access, especially for socio-economically disadvantaged groups (SEDGs), rural areas, and Divyang students; ranges from 0 to 1).

i: Infrastructure availability index (e.g., high-speed internet and device access; ranges from 0 to 1).

t: Teacher training/readiness level (minimum required for effective facilitation).

The optimization must satisfy NEP 2020-aligned inequality constraints:

Equity constraint to prevent widening the digital divide and ensure inclusion for

SEDGs/rural areas, $\geq \mathbf{0.8}$.

Infrastructure constraint (addressing gaps in internet/devices): $\mathbf{i} \geq \mathbf{k} \cdot \mathbf{n}$ where k is the proportionality constant.

Teacher preparedness constraint: $\geq t_{min}$, Minimum training level, as the article highlights the need for robust faculty development programs.

Risk mitigation bound: $\mathbf{r} \leq r_{min}$, Upper limit on acceptable risks, tied to ethical guidelines and regulations like the DPDP Act.

Where non-negativity, $\mathbf{c}, \mathbf{b}, \mathbf{q}, \mathbf{n}, \mathbf{r}, \mathbf{e}, \mathbf{i}, \mathbf{t} \geq \mathbf{0}$

Optimization of Metaverse Integration in India's Education System under NEP 2020

The goal is to maximize the Net Benefit (NB) of adopting Metaverse technologies in alignment with NEP 2020. Values are derived directly from the article's literature review and findings:

$$\mathbf{NB} = b.n.q - c - r$$

Subject to:

$$\mathbf{c} \geq \mathbf{60n}$$

$$\mathbf{r} \leq \mathbf{15}$$

$$\mathbf{n} \leq \mathbf{500,000}$$

$$\mathbf{c} \leq \mathbf{80,000,000}$$

$$\mathbf{n}, \mathbf{c}, \mathbf{r} \geq \mathbf{0}$$

The balances enhanced learning outcomes against costs and risks while respecting constraints such as infrastructure, equity, and budget.

Per-student benefit b=150 (arbitrary units): Averaged from reported gains 78% increase in interdisciplinary teaching confidence, 65% in collaborative research, 82% in technology skills [6]. Approximate average ~75%, scaled up to 150 units per student for modeling impact on engagement, personalization, and outcomes. Additional benefits, like 67% higher participation and long-term cost savings, contribute to this positive factor.

Minimum cost per student = 60 units: Reflects high initial costs (infrastructure, VR devices, content in Indian languages, teacher training) but incorporates 45% cost reduction from Metaverse-based training (Acharjee, Bhattacharjee & Roy, 2025).

Risk penalty = 30 per unit risk: Accounts for challenges like data privacy, ethical concerns, digital divide, and health risks (e.g., addiction, VR fatigue). Risk is capped low in optimal scenarios through guidelines.

These values are illustrative, grounded in the article's quantitative findings (e.g., percentages from faculty surveys) and qualitative challenges.

The Optimal Solution

Optimal scale (n): 500,000 students

Optimal cost (c): 30,000,000 units

Optimal risk (r): 0 units (fully mitigated)

Maximum Net Benefit (NB): 45,000,000 units

This suggests pushing to the maximum feasible scale (500,000 students) while minimizing risk and using the minimum required cost. In practice, this represents a balanced pilot phase focusing on high-equity regions.

Trade-off Illustration: The net benefit increases with scale but is constrained by costs and risks. In a sensitivity analysis (varying n from 0 to 500,000 with minimum c=60n and low risk), NB peaks near the maximum scale (~45,000,000 at n=500,000).

Applications of the Metaverse in the Context of NEP 2020

Metaverse technologies offer versatile applications in implementing NEP 2020's core elements. In experiential learning, virtual simulations enable hands-on activities without physical constraints. For instance, students can explore historical sites in VR, aligning with NEP's storytelling and discovery-based pedagogy. This supports bagless days and internships by providing virtual exposure to vocations such as carpentry and agriculture.

For multidisciplinary approaches, Metaverse platforms allow seamless integration of subjects. A virtual environment could combine physics with art, where students design 3D models using AI tools, fostering creativity and critical thinking as per NEP. In higher education, multidisciplinary universities can use the Metaverse to support collaborative projects across disciplines, simulating real-world scenarios like environmental policy debates.

Technology integration is amplified through the Metaverse's XR features. NEP's virtual labs on DIKSHA/SWAYAM can evolve into full Metaverse ecosystems, offering immersive coding sessions or language learning via avatars. In digital education, the Metaverse addresses rural access by enabling low-bandwidth VR experiences, supporting ODL, and bridging the digital divide.

In faculty development, Metaverse-based programs train teachers in NEP's flexible curricula, using simulated classrooms for practice. Overall, these applications position Metaverse as a catalyst for NEP's transformative vision.

Benefits of Metaverse Adoption in Higher Education

The benefits of the Metaverse in the implementation of NEP 2020 are multifaceted. Enhanced engagement through immersion leads to better learning outcomes; students in virtual worlds retain information longer due to interactive elements. Personalization adapts content to learning styles, enriching experiences for diverse learners (Figure 3).

Figure 3: Experiential Learning in the Metaverse

Figure 4: Multidisciplinary Education via Virtual Worlds

Equity is a key benefit: the Metaverse democratizes access, enabling rural students to attend elite virtual classes, thereby aligning with NEP's inclusivity goals. It fosters skills like empathy and collaboration via avatar interactions, supporting NEP's holistic development (Figure 4).

Cost-effectiveness is realized over the long term; virtual labs reduce physical resource requirements. In India, this could scale education to millions, enhancing employability through skill simulations.

Challenges and Issues

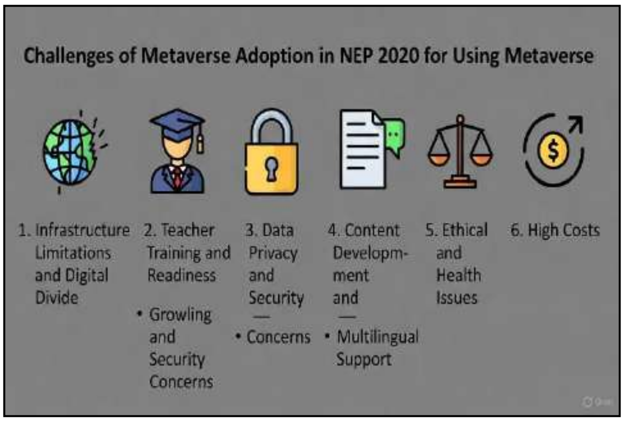


Figure 5: Challenges of Metaverse Adoption in NEP 2020

Despite benefits, challenges abound. Infrastructure gaps in India's limited internet and device access hinder access, potentially exacerbating inequalities. Ethical issues, such as privacy breaches in the data-rich Metaverse, require robust regulations.

Teacher preparedness is critical; many lack VR skills, necessitating training. Health concerns like VR-induced fatigue and addiction must be mitigated (Figure 5).

Implementation costs and content creation in Indian languages pose barriers. Distinguishing real from virtual learning to avoid detachment is essential.

Addressing these through policy, like NETF's role in technology induction, is vital.

Simulated Results

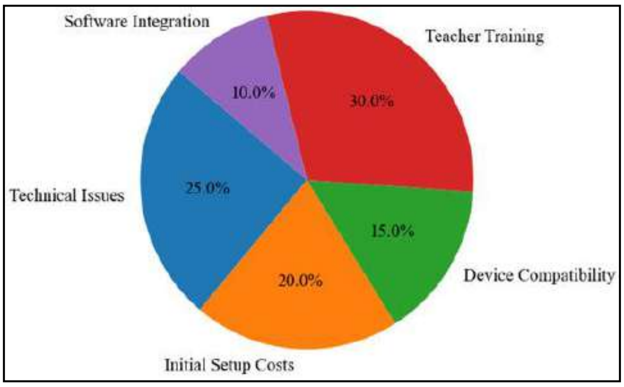


Figure 6: Percentage Increases in Key Areas from Metaverse-Based Training

Figure 6 indicates that Teacher Training (30.0%) is the most significant challenge, consistent with the article's repeated emphasis on teacher unpreparedness and the need for robust training programs. The Technical Issues (25.0%) are substantial barriers, including connectivity, cybersecurity, and usability (as mentioned in). The Initial Setup Costs (20.0%) entail significant upfront expenses for infrastructure and content development, despite long-term savings (linked to 45% reductions in training costs). The Device Compatibility (15.0%) issues with hardware access, especially in rural areas, and for diverse devices. The Software Integration (10.0%) challenges include integrating Metaverse platforms with existing systems such as DIKSHA/SWAYAM.

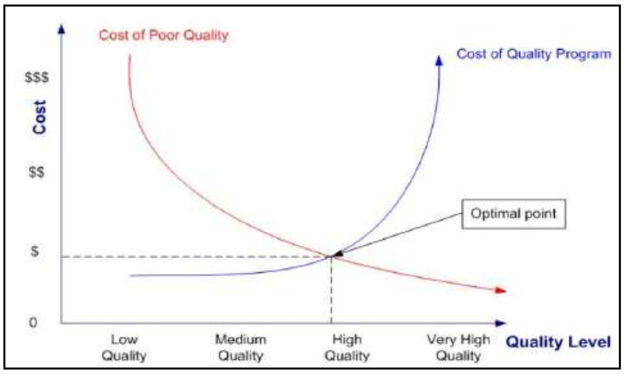


Figure 7: Sensitivity of Net Benefit to Variations in Benefit Parameter

(b=150 base vs. b=100 lower)

Multiple lines showing (Figure 8) how NB changes if benefits are lower (e.g., due to partial implementation challenges), highlighting the robustness of the scale recommendation. Key Data Points (Base b=150 vs. Lower b=100):

n=0: 0 vs. 0

n=101,010: 9,090,909 vs. 4,040,404

n=202,020: 18,181,818 vs. 8,080,808

n=303,030: 27,272,727 vs. 12,121,212

n=404,040: 36,363,636 vs. 16,161,616

n=500,000: 45,000,000 vs. 20,000,000

Findings

The Metaverse holds significant potential to transform Indian higher education by supporting the National Education Policy 2020's vision for multidisciplinary, digital, and inclusive education. Through immersive simulations and interactive experiences, it can enhance science, history, and vocational education, promote global collaboration, and create personalized learning pathways. However, challenges such as infrastructural limitations, high costs, data privacy concerns, and inadequate teacher preparation must be addressed. Strategic investment in technology, ethical safeguards, and teacher training is essential. Through a balanced approach, the integration of the Metaverse can build a flexible, future-ready education system that strengthens engagement, skill development, and equitable opportunities.

Recommendations and Policy Implications

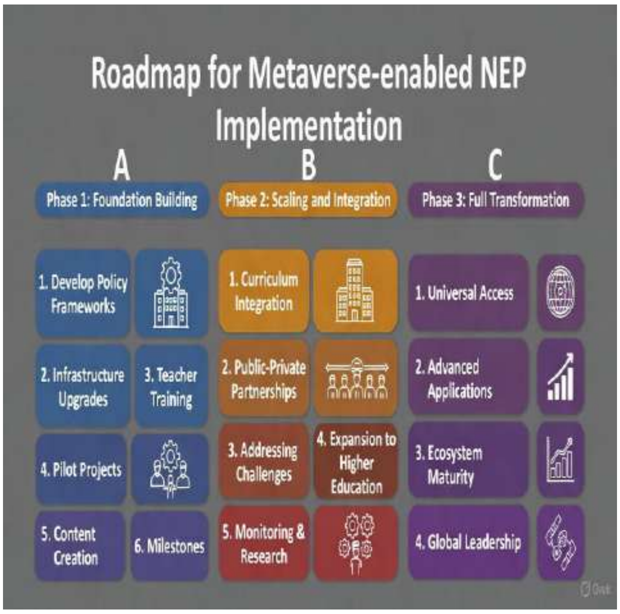


Figure 8: Roadmap of Metaverse Adoption in NEP 2020

To maximize benefits, a multi-stakeholder approach is recommended (Figure 6):

- Policy and Infrastructure: The government, through NETF and bodies like NCERT, should prioritize investments in 5G rollout, subsidized VR hardware, and indigenous platforms (e.g., building on BharOS). Public-private partnerships, like the CBSE-Meta model, should expand nationwide.
- Curriculum and Content: Develop multilingual, NEP-aligned Metaverse modules for DIKSHA/SWAYAM, focusing on vocational simulations and multidisciplinary projects. Encourage open-source contributions to reduce costs.
- Teacher Empowerment: Mandate Metaverse training in faculty development programs, using virtual platforms for scalable, low-cost workshops.

- Ethical Guidelines: Formulate regulations on data privacy (aligned with DPDP Act), age-appropriate usage, and mental health safeguards. Pilot health impact studies.
- Research and Monitoring: Fund empirical studies on learning outcomes, with longitudinal tracking of pilots like Gyandhara Metaverse.
- Inclusive Implementation: Prioritize SEDGs, Divyang students, and rural areas through low-bandwidth alternatives and assistive XR tools.

Conclusion

The integration of Metaverse technologies into India's education system, as envisioned by NEP 2020, holds transformative potential. By enabling immersive, personalized, and equitable learning experiences, the Metaverse aligns seamlessly with the policy's pillars: experiential pedagogy, multidisciplinary education, technology induction, and inclusion for SEDGs. Virtual simulations foster critical thinking and skill development, virtual collaborations promote global citizenship, and accessible platforms bridge urban-rural divides.

However, realization of this potential requires addressing persistent challenges. Infrastructure gaps, particularly in high-speed internet and affordable VR devices, in rural areas risk exacerbating inequalities. Ethical concerns, like data privacy, cyber risks, and screen addiction, demand robust frameworks. Teacher training remains crucial, as many educators lack proficiency in virtual tools. Content development in Indian languages and cost-effective scaling are additional hurdles.

By adopting these strategies, India can leverage the Metaverse to create a future-ready, holistic education system that honors its ancient knowledge traditions while embracing 21st-century innovation. As NEP 2020 progresses toward full implementation, thoughtful adoption of the Metaverse could position India as a global leader in the digital education transformation.

References

Acharjee, T., Bhattacharjee, D., & Roy, A. (2025). Metaverse-enhanced faculty development programmes: Catalyzing multidisciplinary competence in the NEP 2020 framework. http://dx.doi.org/10.2139/ssrn.5278187

Alfaisal, R., Hashim, H., & Azizan, U. H. (2024). Metaverse system adoption in education: A systematic literature review. *Journal of Computers in Education*, *11*(1), 259-303.

Dhillon, P. K. S., & Tinmaz, H. (2024). Academic augmentation: Analyzing avatar design in educational metaverse. *Journal of Metaverse*, *4*(1), 54-70.

Dutta, S. (2025). NEP 2020 and generative AI, immersive technologies-transforming Indian education for a sustainable future. *Global Goals, Local Impacts*, 38-49.

Education, L., & Pedagogicstructures, C. (2024). National education policy 2020. *CEERA.* Retrieve from https://www.education.gov.in/shikshakparv/docs/Background_note_vocational_education.pdf.

Kali, G. (2025). Significance of the metaverse in modern education system. Metaverse and education. In E. Ande Ivorgba (ed.), *Metaverse and Education Virtual Worlds for Teaching and Learning*, pp. 334–342. London.

Krishnan, C., Lamba Sahdev, S., & Mariappan, J. (2024). Navigating complexity: Thematic insights into ethical challenges and metaverse integration in Indian education institutions. *Cogent Education*, *11*(1), https://doi.org/10.1080/2331186X.2024.2428110

Mauz, V., & Díaz, C. (2024). Learning and teaching in the Metaverse: A paradigm shift in education in the developed world. *International Journal of Innovation in Education*, *9*(1), 47-60.

Mukred, M., Mokhtar, U. A., Hawash, B., AlSalman, H., Zohaib, M., & Abuzawayda, Y. I. (2025). Exploring the potential of metaverse adoption in higher education: A diffusion of innovation model approach to enhancing student engagement. *SAGE Open*, *15*(3). https://doi.org/10.1177/21582440251363668

Panda, S. (2025). Technology and ODL for continuing professional development and lifelong learning. *Journal of Learning for Development*, *12*(3), i-vi.

Qi, M. (2025). Shaping the future of metaverse research. *Journal of Metaverse*, *5*(2), 73-91.

Sa, M. J., & Serpa, S. (2023). Metaverse as a learning environment: Some considerations. *Sustainability*, *15*(3), 1-16.

Shivani, D., & Kumari, D. A. (2025). Changing landscape of research in Indian higher education. *International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)*, *2*(4), 98-105.

Singh, M. (2024). Exploring the possibilities to implement metaverse in higher education institutions of India. *Education and Information Technologies*, *29*(15), 20715-20728.

Vijitha, S., & Anandan, R. (2025). Metaverse technology to empower students literacy rate using modern educational resources in Kanchipuram district. *Journal of the Institution of Engineers (India): Series B*, 1-13.

About Authors

Dr. Rajib Roy currently serves as an Assistant University Librarian at the University of Calcutta. With over eight years of experience at the University Library, University of Calcutta, he has also worked for around 10 years as a library professional at the University of North Bengal. Dr. Roy completed his PhD in Library and Information Science (LIS) from the University of Calcutta. He has authored 10 research papers published in various academic journals. With more than fifteen years of experience in library services, his areas of interest include Application of Information Technology in Libraries, Development and Innovation in Library Services and Community Information Services.

Dr. Ashis Kumar Samanta is a seasoned academic technologist and university administrator with over two decades of experience in higher education, information systems, and academic governance. He is currently serving as Database Administrator at the University of Calcutta and has previously held senior administrative positions including Registrar (Acting) and Dean, Students' Welfare at Rabindra Bharati University. His technical expertise spans database administration, enterprise software development, academic automation, and web technologies across multiple platforms and operating systems. He has led and contributed to numerous university-level software projects, particularly examination and student information systems. Dr. Samanta has been closely associated with the conduct and management of approximately 760 university examinations over the past 15 years. His publications & research interests include Blockchain, cybersecurity, smart contracts, Hyperledger frameworks, NoSQL databases, and AI-enabled systems.

Blind Peer Reviewed
(Previously UGC-CARE)
ISSN: 0972-2750

Manuscript Received on –
27th December, 2025
Primary Reviewed on –
9th January, 2026
Peer Reviewed on –
18th January, 2026
Revised on –
17th February, 2026
Accepted for publication:
25th February, 2026

A systematic review of emerging RDM services and policies (2020-2024)

Rahul Rana & Dimple Patel

ABSTRACT

Introduction- *Research data management has become an essential component of the research ecosystem, supporting data preservation, sharing, and reuse across disciplines. Libraries are now increasingly expected to play an active role in the research life cycle by providing dedicated data-driven RDM services.*

Purpose- *This study aims to examine the provision of RDM services and associated policy frameworks through a systematic review of existing literature.*

Research Problem- *Existing studies have addressed specific aspects of RDM; some address RDM services or the policy framework independently, and they are often limited in scope, region, or discipline.*

Objective- *The main objective of this study is to assess the current status of library-led RDM services, policy implementation, and to provide a five-year snapshot of global RDM development.*

Methodology- *A total of 4756 items were retrieved from four scholarly databases: SCOPUS, J-Gate, Google Scholar, and Ebsco-Host. Using the PRISMA framework, 58 data-driven empirical studies published between 2020 and 2024 are included, while editorial, bibliometric, review papers, and non-English publications were excluded.*

Findings- *Collectively, this systematic review reveals significant regional disparities in RDM service provision, with rich-resource institutions in developed regions offering advanced technical services supported by mature policy frameworks, in contrast to those in developing regions. Strong policy support and systematic capacity-building emerge as critical factors in sustainable RDM implementation.*

Keywords- *Research data management, RDM Policy, RDM Services, Systematic literature review, PRISMA*

Study type- *Systematic Literature Review*

Mr. Rahul Rana
Research Scholar
Central University of Himachal Pradesh
rrana8227@gmail.com

Dr. Dimple Patel
Professor
Central University of Himachal Pradesh
dimplerp2011@gmail.com

1 Introduction

In recent years, research data management (RDM) has become more vital for scientists and academics due to the exponential growth of digital research data. Universities and research institutes are the primary sources of this data, especially in the sciences (Boiko et al., 2021). Research data generated during the research life cycle is an essential asset of academic institutions. It is generated in various formats, including numerical data, text, graphical, audiovisual material, and source code by students, scholars, scientists, and faculty members (Fitriana et al., 2022; Ismail et al., 2022). It is imperative that research data be managed appropriately, as it facilitates the validation of research outputs and provides a basis for future research (M. Mthembu & Ocholla, 2022). However, in practice, the research data is often maintained at the individual level without following standard practices of data management. RDM is a continuous process throughout the research cycle, encompassing the curation, organisation, processing, and long-term preservation of data. Key activities encompass the drafting of data management plans (DMPs), ensuring data documentation and formatting, archiving, and facilitating data sharing (Lafferty-Hess et al., 2020). Researchers, however, face numerous challenges in managing the growing volume of research data. They need tools and institutional support to manage and analyse the research data (Castro et al., 2020). Academic libraries are well-positioned to meet this need by providing support and services, including the facility of data storage, user rights management with features such as data linkage, data version control, and assistance in standard file formatting.

Libraries have transformed significantly in terms of library space, services, and resources. This paradigm shift is due to evolving technological modalities, changing information-seeking needs and behaviours, multicultural communities in higher education, and evolving services from competitors of libraries (Ashiq et al., 2022). As a result, modern libraries extend beyond traditional roles to support research by offering workshops on database searching, reference management, research ethics, academic integrity, access to digital resources, and plagiarism-detection tools. Libraries in academic institutions have been involved in research data management services for more than a decade (Griffin & Janz, 2024), and mostly provide advisory services on RDM, but there is increasing demand in today's data-intensive environment for libraries to also provide technical services and hands-on support to researchers. RDM services primarily fall into two categories: advisory and technical services. Advisory services encompass data literacy, create awareness among researchers with legal and ethical aspects of data, and provide avant data a guide to relevant data sources. The technical services include data analysis, data visualisation, data curation activities, aiding in creating DMPs, and running data repositories.

There are recent literature reviews on RDM that offer valuable insight into practices, services, and frameworks. (Ashiq et al., 2022) conducted a systematic review on services and practices of RDM. The study highlights that RDM is poorly practised in libraries, that multiple challenges arise in providing services, and that collaboration with stakeholders is needed to implement an RDM system. Similarly, (Zibani et al., 2022) examined faculty research repositories in higher education, suggesting that repositories provide global visibility, collaboration, and opportunities for data reuse, their effectiveness depends on strong institutional commitment and investment. (Andrikopoulou et al., 2022) explore the impact of RDM on library role and image, emphasising that the visibility of librarians' contributions is crucial; otherwise, their roles will be downgraded in terms of status and salary. They noted that RDM requires skills, collaboration, leadership, and policy support, while practices vary widely across institutions. In the Canadian context (Tayler & Jafary, 2021), the literature review investigated universities' programmes for support staff

to uplift their skills for providing effective RDM service, identifying good practices followed by universities, but also pointing to persistent gaps in institutional policies. (Mavodza, 2022) review of RDM in UAE universities, observing that development remains slow, and emphasised that libraries should lead the initiative of policy development and infrastructure building to strengthen RDM services. (Donner, 2023) examined organisational factors to be considered when implementing RDM systems; findings revealed that RDM systems can be influenced by institutional culture and strategies; all stakeholders must understand their roles and the specific needs of their target users when libraries implement RDM services. (Alomi et al.,2022) reviewed pharmaceutical RDM policies in Saudi

Arabia. The review provides comprehensive insights into policy design, including data collection, reliability, and storage, but is limited in terms of domain and region. (Dorcas et al., 2023) highlighting the role of librarians in knowledge sharing, Open Science enhancement and initiatives, and following best RDM practices in Africa and developing countries. (Azami et al., 2023) identified four principal dimensions of RDM: data, organisation, researcher, and technology. (Safdar et al., 2022) explore RDM service status globally and observed that developed countries provided better services than developing nations, though without analysing the underlying cultural and policy factors driving these disparities. (Birkbeck et al., 2022) focused on challenges in RDM practices and highlighted issues such as data documentation, data storage, human resources, education on RDM, data openness, and ethical use of data, but their analysis was limited to researchers' perspectives, neglecting the views of other stakeholders. (Ho et al., 2025) conducted a systematic review of RDM services in academic libraries, emphasising the importance of national frameworks and institutional contexts as important factors, although their analysis overlooked policy dimensions and other critical aspects shaping service provision.

This study aims to analyse the current landscape of RDM services and policies, providing valuable insights for libraries considering implementing RDM services in their institutions. Although some recent studies have addressed specific aspects of RDM, they are often limited in scope, region, or disciplinary focus. A review of the above-cited literature reveals a research gap: although previous studies have examined either RDM services or RDM policies, no comprehensive systematic review has addressed both.

2 Research Questions

1. What is the current status of the Research Data Management services provided by libraries?
2. What is the current status of RDM policy implementation?

3 Methodology

This study used a systematic literature review methodology following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. PRISMA can be used for reporting systematic reviews of other types of research, as well as for evidence-based systematic reviews and meta-analysis, to identify, select, and critically analyse studies (Moher et al., 2009). PRISMA was primarily used in the field of health science. However, it is equally applicable in other fields, enabling a well-structured, deep insight into any area of study across various aspects. PRISMA is a hierarchical flowchart with four main operations: identification, screening, eligibility, and inclusion of the study for analysis (Ashiq et al., 2022). Traditionally, authors used narrative and critical reviews in literature reviews, which differ from systematic literature reviews and do not follow predefined protocols, making it impossible to replicate the findings (Hemingway & Brereton, 2009). The selection of articles in traditional reviews depends entirely on the author's perception of what should be included in the study, and the selection of the next article for review is based on the citations and references of the first article. The purpose of using PRISMA in this systematic literature review is to overcome the biases (such as author perception, citation bias) and follow a standard protocol for including studies in the review.

3.1 Search Strategies

A systematic search was carried out from 1st to 8th January 2025 across four platforms: SCOPUS, J-Gate, Ebsco-Host, and Google Scholar on these four platforms by using the Boolean operator 'AND' for this study: the following search strings, "RESEARCH DATA MANAGEMENT", "RESEARCH DATA MANAGEMENT AND SERVICES, RESEARCH DATA MANAGEMENT AND POLICY, RESEARCH DATA MANAGEMENT AND STAKEHOLDERS, and RESEARCH DATA MANAGEMENT AND REPOSITORIES were used for retrieving items. Search results were filtered by publication date (2020 - 2024), language (English), and type (research articles) across SCOPUS, J-Gate, and Ebsco-Host. For Google Scholar, an exact match of the search string within titles was applied as a filter. The review period from 2020 to 2024 has been selected to include the recent studies on RDM. Since 2020, the COVID-19 pandemic has increased the demand for open science, data sharing, and the development of strong data management infrastructures.

3.2 Inclusion Criteria

- Must be a scientific research article in the English language.

- Published between January 2020 and December 2024.
- Full text must be available.
- Article must discuss RDM services and RDM policy.

3.3 Exclusion criteria

- Studies that cover other aspects of RDM and not RDM services or RDM policy.
- Studies published before January 2020 and after December 2024.
- Review articles, Bibliometric studies, and books are excluded.
- Items that are not published in the English language.

3.4 Selection of Study

In this study, the researcher reviewed all articles in two phases: the first phase involved reading the title and abstract, and in the second phase, skimming the methodology, discussion, and conclusion sections of the remaining studies to get relevant articles for full-text review. In this study, research data management services and policy are taken as the main theme for the selection of articles for systematic review. A total of 4756 documents were retrieved from these platforms, and their bibliographic information was retrieved in the open-source reference management software Zotero. SCOPUS database readily provided the search result in CSV file format, while in the case of the other three databases, the search results had to be processed through Zotero to first export them in XML file format and import them to the spreadsheet in CSV format. Items published between January 2020 and December 2024 were included; duplicate items were

removed by utilising the spreadsheet's remove duplicate function. After studying the titles, abstracts, and full text of selected items, it was determined that there are 58 items that fit the criteria for systematic review in the studies.

4 Scope

While this study provides a comprehensive overview of RDM services and policies, certain limitations should be acknowledged. First, the review focuses on selected dimensions of research data management, which may not capture all aspects of RDM practices. Second, some important domain-specific library databases, including LISA and LISTA, were not included in the search, which may have excluded relevant studies. Third, the review is restricted by its time frame, language restrictions, publication bias, and reliance on published literature, which further constrain the scope of the review.

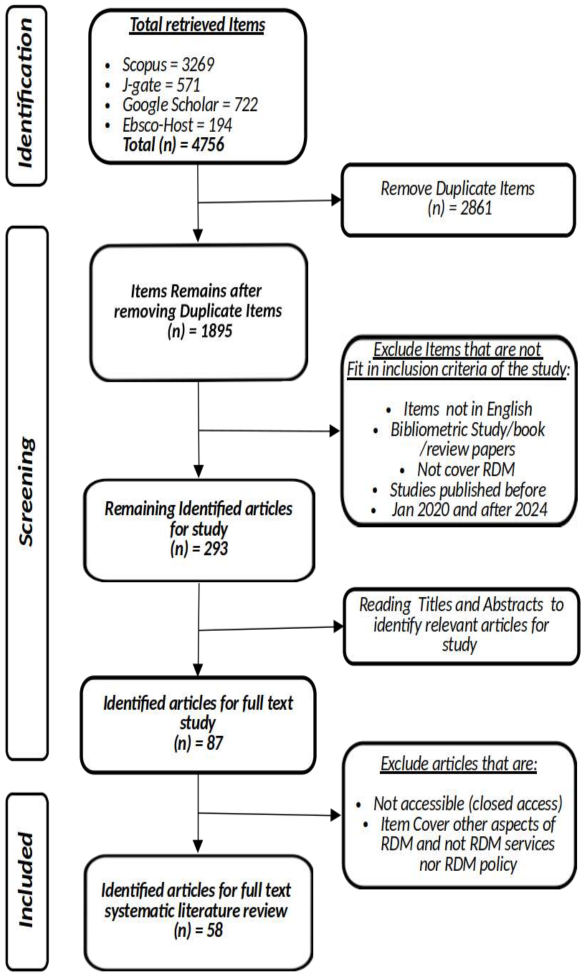


Figure1: PRISMA framework of the stud 1

5 Results

5.1Overview of Selected Studies

Among the included studies for systematic review, twelve were published in the year 2020, followed by nine in 2021, twelve in 2022, fourteen in 2023, and eleven in 2024. The selected studies were published in forty-four different journals. Geographical coverage of included studies is twenty-eight, which were focused on countries of North America, Oceania, and Europe, twenty-five studies on countries of Africa and Asia, one in South America, and two had a global scope. Among the studies, nineteen articles dealt with research data management services and policy; thirty-three articles dealt only with RDM services, and six dealt only with RDM policy are shown in Table 2. All the studies included in this review focus on research data management services and policy. Studies included for systematic review are diverse in nature, including case studies, surveys, content analysis, documentary reviews, and tools used for data collection, such as interviews, questionnaires, website analysis, feedback forms, and reviews of existing documents and datasets.

5.2 Research Data Management Services

The implementation of RDM services in libraries is still at an initial stage (M. Mthembu& Ocholla, 2022). At present most libraries primarily identified that most libraries provide consultative data services for their researchers, such as assistance in data management plans, training and workshops on RDM, data literacy, support for search and retrieval of external data sources, and maintaining a web guide (Amanullah & Abrizah, 2023; Baudin & Mapulanga, 2023; Gowen & Meier, 2020; Griffin & Janz, 2024; Kim & Syn, 2021; Liu et al., 2020; Machimbidza et al., 2022; Martin-Melon et al., 2023; Nahotko et al., 2023; Sinha et al., 2023), and some libraries from developed nations provide technical services such as running a data repository, data analysis and visualization, creation of metadata, long-term preservation of data, providing a data management plan template, and accessioning of data. (Howlett et al., 2024; Huang et al., 2021). Most frequent barriers reported for not providing technical services are a lack of technical skills among library staff, limited infrastructure, insufficient funds, and inadequate policy (Griffin & Janz, 2024; Jose et al., 2024; Martin-Melon et al., 2023; Masinde et al., 2021; Maurya & Madhusudhan, 2023).

Several studies point out that library professionals lack essential skills required to provide effective RDM services. This includes knowledge of data management practices, data analysis, data visualisation, locating relevant datasets from different sources, understanding of diverse research methodology, and familiarity with various metadata schemas related to data management (Badenhorst & Raju, 2023; Griffin & Janz, 2024; Masinde et al., 2021; M. S.Mthembu & Ocholla, 2024; Nahotko et al., 2023; Rod, 2023; Yidavalapati et al., 2022). To address these gaps, it is recommended that libraries provide training and workshop programs on RDM for their professionals in order to enhance their technical competence. (Borkakoti & Singh, 2021; M. S. Mthembu & Ocholla, 2024; Triasih et al., 2020). Researchers frequently manage large-scale research data independently at their own level (Cheung et al., 2022; Ismail et al., 2022; Machimbidza et al., 2022).

Such data are often generated in multiple formats and handled through personal data management practices (Bradley et al., 2021; Cheung et al., 2022; Fu, 2021; Ismail et al., 2022; Maurya & Subaveerapandiyan, 2022; Reichmann et al., 2021). Data sharing is rare among researchers due to concerns about privacy, sensitive data, following non-standardised metadata practices, and the possibility of reusing that data for more publications (Cheung et al., 2022; Masinde et al., 2021). Jose et al. identified that researchers did not store data with backup and security, lacked knowledge of data processing, and had less awareness about the legal aspects of data (Jose et al., 2024).

Researchers' perspectives and needs on RDM vary from discipline to discipline, institute to institute, and region to region (Fu et al., 2022; Fu, 2021; Mosha & Ngulube, 2023; Mushi, 2021). Some studies highlight a low level of knowledge about RDM among researchers (Avuglah, 2020; Kalová, 2020; Masinde et al., 2021; M. Mthembu & Ocholla, 2022; Mushi, 2021), and many researchers avoid RDM practices because of time constraints, cultural practices, and lack of knowledge on RDM (Donner & Huber, 2024). Libraries can bridge this gap by providing support tools, workshops, and training programmes on RDM (Baudin & Mapulanga, 2023; Bourke, 2022; Cheung et al., 2022; Fu et al., 2022; Kalová, 2020; Maurya & Madhusudhan, 2023; Mosha & Ngulube, 2023; Mushi et al., 2020; Zain et al., 2023) that enhance the knowledge of researchers on RDM (Xu et al., 2022).

RDM services are not provided solely by libraries because they require collaboration with other departments and stakeholders (Kim & Syn, 2021; Masinde et al., 2021) and even with other institution libraries (Lafferty-Hess et al., 2020). Few studies indicate that students are often unaware of RDM services provided by libraries (Mosha & Ngulube, 2023; Zhou et al., 2023), and emphasise the necessity to promote their services within the academic community (Dessa & Dani, 2024; Maurya & Madhusudhan, 2023; Sheffield & Burton, 2022). In this study, it is found that RDM is in very early stages in developing countries compared with developed countries, where researchers demonstrate confidence in following RDM practices (Zhou et al., 2023). Development of data repositories emerges as a core RDM service (Huang et al., 2021), and most members of Dataverse are from Western countries as compared to Asian and African continents (Chen et al., 2023).

RDM is a complex practice with numerous challenges. Common challenges that were identified are the heterogeneous nature of data, lack of data sharing cultural across disciplines and institutions, the sensitivity of some data (Griffin & Janz, 2024; Martin-Melon et al., 2023), limited tools and infrastructure for data analysis, lack of secure data communication channels (Martin-Melon et al., 2023; Zain et al., 2023), unclear institutional policies, as well as the roles and responsibilities of stakeholders (Bellgard et al., 2023), low awareness among students/researchers about RDM (Avuglah, 2020; Donner & Huber, 2024; Mosha & Ngulube, 2023; M. Mthembu & Ocholla, 2022), and lack of specified location for data submission (Baek et al., 2024). Challenges that libraries faced in providing effective RDM services include limited staff, lack of technical skills among library professionals, limited knowledge of RDM, lack of funding (Amanullah & Abrizah, 2023; Avuglah, 2020; Kuruppu, 2023; Masinde et al., 2021; M'kulama et al., 2022; Triasih et al., 2020; Yidavalapati et al., 2022), data security, inadequate institutional policy (Maurya & Madhusudhan, 2023), lack of infrastructure to support RDM services (Jose et al., 2024; Machimbidza et al., 2022; M. Mthembu & Ocholla, 2022).

5.3 Research Data Management Policy

Policy is an important driver for developing and sustaining an effective RDM system in an academic environment. It also plays an important role in cultural change towards RDM (Avuglah, 2020). However, many institutions lack formal RDM policies (Amanullah & Abrizah, 2023; Machimbidza et al., 2022; M'kulama et al., 2022), while in others it is in the initial stages of development (Martin-Melon et al., 2023; M. Mthembu & Ocholla, 2022; Nahotko et al., 2023). Existing policies often lack enforceable standards and are not clear on various aspects of RDM, such as accessibility of sensitive data, data ownership, and data sharing (Asok et al., 2024; Liu et al., 2020; Read et al., 2021; Si & Liu, 2024). RDM policy development is a collaborative effort in which every stakeholder should participate. (Borkakoti & Singh, 2021; Bunkar & Bhatt, 2020; M. Mthembu & Ocholla, 2022), and the library can play an important role in policy development (Bishop et al., 2022). Beckles emphasises that policy development committees should address the key issues, such as the exemption of sensitive data from open data (Beckles, 2020), data format and relevant repositories for data submission, legal aspects of the data, data availability statement, and the data management plan (Jackson, 2021; Read et al., 2021). Since every discipline deals with different data and different needs for data management, policy should be designed to address the needs of researchers within a specific field (Reichmann et al., 2021). Researchers want policies on various aspects of RDM, including data access policy, data retention policy, data archival, and data disposal policy. Institutions should focus on policy development before implementing the RDM system (Bunkar & Bhatt, 2020; Mushi et al., 2020).

Mosha and Ngulube identified that funder policy on data submission to the open domain encourages researchers to follow RDM practices (Mosha & Ngulube, 2023) because funding agencies mandated that researchers must include the DMP with the proposal, submit data to open repositories with metadata, and provide a Data Availability Statement (DAS) (Cooper & Costanzo, 2024)

Hrynaszkiewicz et al. mention that a clearly structured policy is necessary to address the multifaceted challenges of RDM (Hrynaszkiewicz et al., 2020). A well-designed policy defines clear roles and responsibilities for stakeholders, and establishes guidelines for effective RDM (Njagi et al., 2024), and improving the re-usability of data (Hrynaszkiewicz et al., 2020). Such policies create an environment that fosters collaboration, exchange of ideas, and access to data (Manco, 2024).

6 Discussion

The study reviewed fifty-eight studies on RDM services and policy. Libraries worldwide acknowledge the importance of RDM, studies from European, North American, and Oceanian regions reported technical data-oriented RDM services and policy frameworks (Howlett et al., 2024; Huang et al., 2021) as compared to those from Asian and African countries, where mostly services are consultative in nature (Amanullah & Abrizah, 2023; Baudin & Mapulanga, 2023; Liu et al., 2020; Machimbidza et al., 2022; Sinha et al., 2023). Regional disparities are explained by stronger advocacy and support for open data, open-science policies, funding agencies' policies on data management, and institutional support in developed regions. In contrast, Asian and African countries experience weaker institutional support, a lack of awareness and misconceptions about RDM, insufficient professional training, and the absence of dedicated policy frameworks (Amanullah & Abrizah, 2023; Machimbidza et al., 2022; Masinde et al., 2021, 2021; M. Mthembu & Ocholla, 2022) that hinder the RDM implementation. The analysis reveals that strong advocacy programs, institutional support, and well-established policy are the decisive factors in transforming libraries from advisory to technical RDM service hubs.

A common finding across the studies is the lack of technical knowledge and expertise among library staff (Avuglah, 2020; M'kulama et al., 2022; M. Mthembu & Ocholla, 2022), which demonstrate a clear need for capacity building for library professionals to enhance their technical data skills in data curation, data storage, retrieving datasets from various sources, using data visualization and analysis tools, and effectively managing institutional repositories. Persistent skill gaps highlight a fundamental weakness in library and information science (LIS) curricula, which frequently fail to integrate RDM and data science competencies in a structured way. Researchers, on the other hand, manage their data independently (Cheung et al., 2022; Ismail et al., 2022; Machimbidza et al., 2022), but their practices are inconsistent (Jose et al., 2024), discipline-specific and constrained by cultural attitudes or inadequate institutional support (Cheung et al., 2022; Masinde et al., 2021). Some studies highlight that researchers show a clear need for assistance in effective RDM practices, including data analysis, development of data management plans, data visualisation, and maintaining standard file formats (Fu et al., 2022; Mosha & Ngulube, 2023; Mushi, 2021). By responding to researchers' RDM needs, libraries have the opportunity to transform their role as data service providers and position themselves as indispensable stakeholders in the research data life cycle.

Policy frameworks are identified as important drivers of RDM systems (Asok et al., 2024; Mosha & Ngulube, 2023). The reviewed studies highlight that a comprehensive policy should cover all the relevant aspects of research data management that include the responsibility of data curation, provide guideline for data storage, outline short-term and long-term data preservation strategies, specific data retention period, clarify data ownership among researcher, institutions, and funding agencies, established rules for data access, intellectual property rights management, and the roles of various stakeholders in the RDM system (Jackson, 2021; Njagi et al., 2024; Read et al., 2021; Si & Liu, 2024).

Researchers across various disciplines have diverse requirements from an RDM policy, such as medical and health science demands, which require the protection of highly sensitive data. In contrast, social science emphasises confidentiality and ethics. Further, this literature review identified that studies from Asia and Africa reported an absence of dedicated institutional RDM policies (Amanullah & Abrizah, 2023; Machimbidza et al., 2022; Masinde et al., 2021, 2021), whereas studies from North America, Oceania, and Europe emphasised either mature policy frameworks or ongoing policy development (Cooper & Costanzo, 2024; Liu et al., 2020; Martin-Melon et al., 2023; Si & Liu, 2024). This highlights that regional disparities exist not only in service provision but also in policy development and implementation.

7 Suggestions

- Libraries with limited resources can enhance their technical services by resource sharing, adopting open-source tools, and hiring skilled technical staff.
- RDM systems implementation is a long process, so its development should be carried out in various phases from strategy to implementation.
- Before implementing RDM systems, libraries should conduct a study on the factors that influence RDM, such as researchers' needs, legal issues surrounding research data, and national-level policy on data.
- Institutional policies must move beyond draft statements to enforceable frameworks that address issues of data ownership, security, archival, and reuse. Data accessibility, open, closed or restricted, must be based on the data's nature and ethical and legal considerations.

- Accessibility of data should (i.e., open/closed/restricted) depend on the nature of the data.
- Capacity building among library professionals should be prioritised through both institutional training workshops and restructuring of LIS curricula to include data science and RDM competencies.
- Researchers should be offered refresher courses on RDM during the early stages of their research careers to enhance their data management skills and promote best practices.

8 Conclusion

Overall, this systematic review demonstrates libraries as a key stakeholder for the development of a sustainable RDM system, while also revealing the regional and structural factors that shape the library-led RDM services. The synthesis of the reviewed literature highlights three key interrelated findings that successfully address the research questions. First, RDM service provision is uneven across institutions; rich resources libraries deliver more data-driven technical services. Second, policy frameworks are a key determinant for the implementation of a robust RDM system. Third, capacity building is essential; without systematic training of researchers and library professionals, RDM efforts remain unsustainable. Collectively, the findings support and extend the previous studies by demonstrating that strong policies, institutional resources, and skilled professionals are critical for providing more technical services. By strengthening these dimensions, libraries can more effectively support the research data life cycle and enhance their strategic role in the research environment. As this study is limited to published academic literature focusing primarily on services and policy, other dimensions of RDM remain underexplored. Future studies could examine additional aspects of RDM, including repository functionalities, stakeholder perspectives, and legal and ethical challenges, to provide further insights into how these factors influence the implementation of RDM services and policies. Empirical research employing surveys, interviews, or case studies would further enhance understanding of the current RDM landscape and support libraries in designing more efficient and effective RDM services.

References

Alomi, Y. A., Al-Najjar, A. H., Abushomi, H. Q. A., Almadany, M. H., & Almasoudi, A. H. (2022). Research policy of the research data management in pharmacy practice. *International Journal of Pharmacology and Clinical Sciences*, 11(1), 20–24. https://doi.org/10.5530/ijpcs.2022.11.5

Amanullah, S. W., & Abrizah, A. (2023). The landscape of research data management services in Malaysian academic libraries: Librarians practices and roles. *The Electronic Library*, 41(1), 63-86. https://doi.org/10.1108/EL-06-2022-0135

Andrikopoulou, A., Rowley, J., & Walton, G. (2022). Research data management (rdm) and the evolving identity of academic libraries and librarians: A literature review. New Review of Academic Librarianship, 28(4), 349–365. https://doi.org/https://doi.org/10.1080/13614533.2021.1964549

Ashiq, M., Usmani, M. H., & Naeem, M. (2022). A systematic literature review on research data management practices and services. Global Knowledge, Memory and Communication, 71(8/9), 649–671. https://doi.org/10.1108/GKMC-07-2020-0103

Asok, K., Gupta, D. K., & Shrivastava, P. (2024). Research data policy: A library and information science publishers perspective. *Quality & Quantity*, 1–22. https://doi.org/10.1007/s11135-024-01994-8

Avuglah, B. K. (2020). Research data management (RDM) at the University of Ghana: Myth or reality? *International Journal of Digital Curation*, 15(1), 1–25. https://doi.org/10.2218/ijdc.v15i1.670

Azami, M., Sadatmoosavi, A., & Chashmyazdan, M. (2023). Research data management frameworks: A systematic literature review. *International Journal of Information Science and Management (IJISM*), 21(3), 1–18. https://doi.org/10.22034/ijism.2023.1977759.0

Badenhorst, P., & Raju, J. (2023). Research data management competencies for academic libraries: Perspectives from two universities in South Africa. *African Journal of Library, Archives & Information Science*, 33(2), 215–229. https://doi.org/10.4314/ajlais.v33i2.7

Baek, J., Lawson, J., & Rahimzadeh, V. (2024). Investigating the roles and responsibilities of institutional signing officials after data sharing policy reform for federally funded research in the United States: National survey. JMIR Formative Research, 8(1), 1–10. https://doi.org/10.2196/49822

Baudin, H., & Mapulanga, P. (2023). A qualitative assessment of the eresearch knowledge centres support practices in the human sciences research council in Pretoria, *South Africa. Digital Library Perspectives*, 39(4), 628–648. https://doi.org/10.1108/DLP-02-2023-0012

Beckles, Z. (2020). Extending support for publishing sensitive research data at the University of Bristol. *International Journal of Digital Curation*, 15(1), 1–5. https://doi.org/10.2218/ijdc.v15i1.712

Bellgard, M., Bennett, R., Wyborn, Y., Williams, C., Barner, L., & Zeps, N. (2023). RDM+ PM checklist: Towards a measure of your institutions preparedness for the effective planning ofresearch data management. *Data Science Journal*, 22(1), 1–13. https://doi.org/10.5334/dsj-2023-036

Birkbeck, G., Nagle, T., & Sammon, D. (2022). Challenges in research data management practices: A literature analysis. *Journal of Decision Systems*, 31(supl), 153–167. https://doi.org/10.1080/12460125.2022.2074653

Bishop, B. W., Orehek, A. M., Eaker, C., & Smith, P. L. (2022). Data services librarians responsibilities and perspectives on research data management. *Journal of eScience Librarianship*, 11(1), 1–32. https://doi.org/10.7191/jeslib.2022.1226

Boiko, A., Kramarenko, O., & Shabanov, S. (2021). The state of dissemination of open research data in Ukraine and the world: Bibliometric analysis. *European Journal of Management Issues*, 29(4), 209–217. https://doi.org/10.15421/192120

Borkakoti, R., & Singh, S. K. (2021). Research data management in central universities and institutes of national importance: A perspective from North East India. *Library Philosophy and Practice*, 5848, 1–14. https://digitalcommons.unl.edu/libphilprac/5848/

Bourke, T. (2022). Bibliographic control of research datasets: Reflections from the EUI Library. *Journal of Library and Information Science, 13(1), 321–334*.

https://doi.org/10.4403/jlis.it-12723

Bradley, B., Wade, Collier, H., Orehek, A. M., & Ihli, M. (2021). Potential roles for science librarians in research data management: A gap analysis. *Issues in Science and Technology Librarianship*, (98). https://doi.org/10.29173/istl2602

Bunkar, A. R., & Bhatt, D. D. (2020). Perception of researchers; academicians of Parul University towards research data management system; role of library: A study. *DESIDOC Journal of Library & Information Technology*, 40(3), 139–146. https://doi.org/10.14429/djlit.40.3.15302

Castro, J. A., Landeira, C., da Silva, J. R., & Ribeiro, C. (2020). Role of content analysis in improving the curation of experimental data. *International journal of digital curation*, 15(01), 1–15. https://doi.org/10.2218/ijdc.v15i1.705

Chen, H.-l., Chiu, T.-H., & Cline, E. (2023). Academic libraries and research data management: A case study of dataverse global adoption. *Information Discovery and Delivery*, 51(2), 166–178. https://doi.org/10.1108/IDD-04-2022-0028

Cheung, M., Cooper, A., Dearborn, D., Hill, E., Johnson, E., Mitchell, M., & Thompson, K. (2022). Practices before policy: Research data management behaviours in Canada. *Partnership*, 17(1),1–80. https://doi.org/10.21083/partnership.v17i1.6779

Cooper, A., & Costanzo, L. (2024). Developing institutional research data management strategies in Canada: Setting the foundation for stronger partnerships and collaborations. *IASSIST Quarterly*, 48(2), 1–18. https://doi.org/10.29173/iq1096

Dessa, H., & Dani, E. (2024). An investigation of current research support services in

Hungarian academic libraries. *The Canadian Journal of Information and Library Science*, 47(2), 39–48. https://doi.org/10.5206/cjils-rcsib.v47i2.17533

Donner, E. K. (2023). Research data management systems and the organization of universities and research institutes: A systematic literature review. *Journal of Librarianship and Information Science*, 55(2),261–281. https://doi.org/10.1177/09610006211070282

Donner, E. K., & Huber, C. M. (2024). Research data management in german academia from a multiple logic perspective. M@n@gement, 28(2),72–87. https://doi.org/10.37725/mgmt.2024.9700

Dorcas, I., Jiyane, V., Ibinaiye, J., & Tijani, R. (2023). Analyses of a decade of research gap in data management practices in academic libraries: A social science perspectives. *International Journal of Social Science Research and Review*, 6(5),271–289. https://doi.org/10.47814/ijssrr.v6i5.1108

Fitriana, F., Sukarni, S., & Zulkifli, Z. (2022). Complexity of web-based application for research and community service in academic. Complexity, 13(10),131–135. https://nelitiuap.aisyahuniversity.ac.id/uploades/penelitian/Paper_17Complexity_of_Web_based_Application_for_Research.pdf

Fu, P., Blackson, M., & Valentino, M. (2022). Developing research data management services in a regional comprehensive university: The case of Central Washington University. *IFLA Journal*, 49(2),443–451. https://doi.org/10.1177/03400352221116923

Fu, P. (2021). A research project of developing a research data management service (RDS) program at CWU: Needs assessment and recommendations. *Library Scholarship*, (118).

https://digitalcommons.cwu.edu/libraryfac/118

Gowen, E., & Meier, J. J. (2020). Research data management services and strategic planning in libraries today: A longitudinal study. *Journal of Librarianship and Scholarly Communication*, 8, 1–8. https://doi.org/10.7710/2162-3309.2336

Griffin, T. M., & Janz, M. (2024). Research data management sustainability: Services, infrastructure, accountability, and planning. portal: Libraries and the Academy, 24(2), 305–342.https://doi.org/10.1353/pla.2024.a923709

Hemingway, P., & Brereton, N. (2009). What is a systematic review. Evidence-based medicine: Second edition,(1–8). https://familymedicine.med.wayne.edu/mph/project/what_is_a_systematic_review.pdf

Ho, R. C. Y., Wong, S. N., Chia, P., Tang, C., & Ng, M. T. T. (2025). Research data management services in academic libraries to support the research data life cycle: A systematic review. *Journal of the Association for Information Science and Technology*, 1–29. https://doi.org/10.1002/asi.70008

Howlett, A., Colla, E., & Joyce, R. (2024). Toward redefining library research support services in Australia and Aotearoa New Zealand: An evidence-based practice approach. New Review of Academic Librarianship, 30(1), 24–51. https://doi.org/10.1080/13614533.2024.2306360

Hrynaszkiewicz, I., Simons, N., Hussain, A., Grant, R., & Goudie, S. (2020). Developing a research data policy framework for all journals and publishers. *Data Science Journal*, 19(5), 5–15. https://doi.org/10.5334/dsj-2020-005

Huang, Y., Cox, A. M., & Sbaffi, L. (2021). Research data management policy and practice in Chinese university libraries. *Journal of the Association for Information Science and Technology*, 72(4), 493–506. https://doi.org/10.1002/asi.24413506

Ismail, M. I., Che Jaafar, C. R., Azmi, N. A., Makhtar, M. M. Z., Samsuddin, S. F., & Abrizah, A. (2022). Eliciting researchers' behaviour as the foundation of research data management service

development. LIBRES: *Library & Information Science Research Electronic Journal*, 32(1), 44–63. https://doi.org/10.32655/LIBRES.2022.1.4

Jackson, B. (2021). Open data policies among library and information science journals. Publications, 9(25), 1–12. https://doi.org/10.3390/publications9020025

Jose, A., Dianelis, O., Manuel, M., Sadia, V., Cesar, L., Celso, G., Josefa, P. M., Arelexys, G., Andrea, B., Grizly, M., et al. (2024). State of research data management in latin American Universities 2022. Procedia Computer Science, 249, 209–215. https://doi.org/10.1016/j.procs.2024.11.066

Kalová, T. (2020). Creating needs-based metadata and research data management services: Exploring the requirements of scientists. *Young Information Scientist*, 5(3), 31–46. https://doi.org/10.1002/asi.24413

Kim, S., & Syn, S. Y. (2021). Practical considerations for a library's research data management services: The case of the national institutes of health library. *Journal of the Medical Library Association: JMLA*, 109(3), 450–458. https://doi.org/10.5195/jmla.2021.995

Kuruppu, R. I. (2023). Users satisfaction with factors affecting research data management systems in subject-related libraries in Colombo district, Sri Lanka. *Multidisciplinary International Journal of Research and Development*, 3(01), 8–20. https://www.mijrd.com/papers/v3/i1/MIJRDV3I10002.pdf

Lafferty-Hess, S., Rudder, J., Downey, M., Ivey, S., Darragh, J., & Kati, R. (2020). Conceptualizing data curation activities within two academic libraries. *Journal of Librarianship and Scholarly Communication*, 8(1), 1–17. https://doi.org/10.7710/2162-3309.2347

Liu, G., Zotoo, I. K., & Su, W. (2020). Research data management policies in USA, UK and Australia universities: An online survey. Malaysian Journal of Library & Information Science, 25(2), 21–42. https://doi.org/10.22452/mjlis.vol25no2.2

Machimbidza, T., Mutula, S., & Ndiweni, M. (2022). Exploring the feasibility of research data management services in Zimbabwean academic libraries: The case of zimbabwe school of mines. *Library Philosophy and Practice*, 6760, 1–34. https://digitalcommons.unl.edu/libphilprac/6760

Manco, A. (2024). Open science policies as regarded by the communities of researchers from the basic sciences in the scientific periphery. Online Information Review, 48(6), 1065–1087. https://doi.org/10.1108/OIR-03-2023-0135

Martin-Melon, R., Hernández-Pérez, T., & Martnez-Cardama, S. (2023). Research data services (RDS) in Spanish academic libraries. The Journal of Academic Librarianship, 49(4), 1–11. https://doi.org/10.1016/j.acalib.2023.102732

Masinde, J. M., Chen, J., & Muthee, D. W. (2021). Researchers' perceptions of research data management activities at an academic library in a developing country. *International Journal of Library and Information Services (IJLIS)*, 10(2), 1–17. https://doi.org/10.4018/IJLIS.20210701.oa11

Masinde, J., Chen, J., Wambiri, D., & Angela, M. (2021). Research librarians' experiences of research data management activities at an academic library in a developing country. Data and Information Management, 5(4), 412–424. https://doi.org/10.2478/dim-2021-0002

Maurya, A., & Madhusudhan, M. (2023). Research data management services in the Indian Institute of Technology libraries. *Journal of Advancements in Library Sciences*, 10(3), 33–40. https://doi.org/10.37591/JoALS

Maurya, A., & Subaveerapandiyan, A. (2022). Research data preservation practices of library and information science faculties. *DESIDOC Journal of Library & Information Technology*, 42(4),259–264. https://doi.org/10.14429/djlit.42.4.17538

Mavodza, J. (2022). Research data management: A review of uae academic library experience. Open Information Science, 6(1), 16–27. https://doi.org/10.1515/opis-2022-0128

M'kulama, A., Zulu, Z., Chewe, P., & Mwiinga, T. M. (2022). Preparedness for open science through research data management at the University of Zambia in covid-19 and post-covid eras. *Library Philosophy & Practice*. https://digitalcommons.unl.edu/libphilprac/7268

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), 1–6. https://doi.org/10.1371/journal.pmed.1000097

Mosha, N. F., & Ngulube, P. (2023). Teaching research data management courses in higher learning institutions in Tanzania. Library Management, 44(1/2), 166–179. https://doi.org/10.1108/LM-04-2022-0033

Mthembu, M., & Ocholla, D. (2022). Development strategies as catalysts for provision of the rdm services in the South African higher education institutions. South African Journal of Libraries and Information Science, 88(1), 1–8. https://doi.org/10.7553/88-1-2127

Mthembu, M. S., & Ocholla, D. N. (2024). Research data management competencies of researchers in selected public universities in South Africa. Library Management, 45(3/4), 208–225. https://doi.org/10.1108/LM-03-2023-0022

Mushi, G. E. (2021). Research data management and services: Resources for different data practitioners. *IASSIST Quarterly*, 45(3–4), 1–8. https://doi.org/10.29173/iq995

Mushi, G. E., Pienaar, H., & van Deventer, M. (2020). Identifying and implementing relevant research data management services for the library at the University of Dodoma, Tanzania. *Data Science Journal*, 19(1), 1–9. https://doi.org/10.5334/dsj-2020-001

Nahotko, M., Zych, M., Januszko-Szakiel, A., & Jaskowska, M. (2023). Big data-driven investigation into the maturity of library research data services (RDS). *The Journal of Academic Librarianship*, 49(1), 1–16. https://doi.org/10.1016/j.acalib.2022.102646

Njagi, P. R., Njoroge, G., & Bituka, R. (2024). Understanding the legal framework in research data management: A study of selected academic libraries in kenya. *Regional Journal of Information and Knowledge Management*, 9(1), 93–102.

Read, K., Campbell, A., Kitchin, V., MacDonald, H., & McKeown, S. (2021). Embracing the value

of research data: Introducing the jchla/jabsc data sharing policy. The Journal of the CanadianHealth Libraries Association, 42(1), 6–13. https://pmc.ncbi.nlm.nih.gov/articles/PMC9327608/

Reichmann, S., Klebel, T., Hasani-Mavriqi, I., & Ross-Hellauer, T. (2021). Between administration and research: Understanding data management practices in an institutional context. *Journal of the*

Association for Information Science and Technology, 72(11), 1415–1431. https://doi.org/10.1002/asi.24492

Rod, A. B. (2023). It takes a researcher to know a researcher: Academic librarian perspectives regarding skills and training for research data support in canada. *Evidence Based Library and Information Practice*, 18(2), 44–58. https://www.erudit.org/en/journals/eblip/2023-v18-n2-eblip08222/1101594ar/abstract/

Safdar, M., Rehman, S. U., Arif, M., & Ashiq, M. (2022). Research data services in libraries: A systematic literature review. Information Discovery and Delivery, 51(2), 151–165. https://doi.org/10.1108/IDD-04-2021-0044

Sheffield, M., & Burton, K. B. (2022). Research data management needs assessment of Clemson university. *Journal of Librarianship and Scholarly Communication*, 10(1), 1–28. https://doi.org/10.31274/jlsc.13970

Si, L., & Liu, X. (2024). Nip risks in the bud: Research data ethics governance framework and collaborative network from the perspective of UK policy. *Aslib Journal of Information Management*, 77(3), 599–619. https://doi.org/10.1108/AJIM-07-2023-0238

Sinha, P., Subaveerapandiyan, A., & Sinha, M. K. (2023). Research data management services in academic libraries: A comparative study of South Asia and Southeast Asia. Global Knowledge, Memory and Communication, 74(3/4), 777–793. https://doi.org/10.1108/GKMC-01-2023-0033

Tayler, F., & Jafary, M. (2021). Shifting horizons: A literature review of research data management train-the-trainer models for library and campus-wide research support staff in Canadian institutions. *Evidence Based Library and Information Practice*, 16(1), 78–90. https://doi.org/10.18438/eblip29814

Triasih, H., Rahmi, R., & Devi, K. S. (2020). Providing research data management services and practices at PDDI-LIPI: Preparedness, roles, challenges, and training. Baca*: Journal Dokumentasi Dan Informasi*, 41(2), 169–178. https://doi.org/10.14203/j.baca.v41i2.707

Xu, Z., Zhou, X., Kogut, A., & Clough, M. (2022). Effect of online research data management instruction on social science graduate students' RDM skills*. Library & Information Science Research*, 44(4), 1–8. https://doi.org/10.1016/j.lisr.2022.101190

Yidavalapati, J., Sinha, P., & Subaveerapandiyan, A. (2022). Research data management and services in South Asian academic libraries. *Library Philosophy and Practice*, 6457. https://digitalcommons.unl.edu/libphilprac/6457/

Zain, M. W., Ahmad, Z. S., Paula, L., Kamarudin, N., & Nordin, S. K. (2023). Academic library opportunity: Examining the research data management challenges facing Malaysian information science researchers. *Journal of Information and Knowledge Management (JIKM),* 2, 94–106. https://ir.uitm.edu.my/id/eprint/87357

Zhou, X., Xu, Z., & Kogut, A. (2023). Research data management needs assessment for social sciences graduate students: A mixed methods study*. Plos one*, 18(2), 1–17. https://doi.org/10.1371/journal.pone.0282152

Zibani, P., Rajkoomar, M., & Naicker, N. (2022). A systematic review of faculty research repositories at higher education institutions. *Digital Library Perspectives*, 38(2), 237–248. https://doi.org/https://doi.org/10.1108/DLP-04-2021-0035

About Authors

Mr. Rahul Rana is research scholar in the Department of Library and Information Science, CUHP. His areas of research interest are Machine Learning, Natural Language Processing (NLP), Research Data Management, Academic Integrity. His contributions include literature review, methodology design, data collection, data analysis, and visualization of findings.

Dr. Dimple Patel is Professor in the Department of Library and Information Science, CUHP. She has published research articles in both national and international journals. Her areas of research interest are Digital Libraries, Multilingual Information Retrieval in Indian Languages, Open Source Software for Libraries, Open Access, Knowledge Organization methods and Semantic Web Technologies. She contributed to conceptualizing the study and provided overall supervision throughout the research process

Sl.no	Authors	Title	Journal	Geographical scope
1	Lafferty-Hess et al., (2020)	Conceptualizing Data Curation Activities Within Two Academic Libraries	Journal of Librarianship and Scholarly Communication	United States
2	Hrynaszkiewicz et al., (2020)	Developing a Research Data Policy Framework for All Journals and Publishers	Data Science Journal	Global
3	Avuglah, (2020)	Research Data Management at the University of Ghana: Myth or Reality?	International Journal of Digital Curation	Ghana
4	Beckles, (2020)	Extending Support for Publishing Sensitive Research Data at the University of Bristol	International Journal of Digital Curation	United Kingdom
5	Bunkar & Bhatt, (2020)	Perception of Researchers & Academicians of Parul University towards Research Data Management System & Role of Library: A Study	DESIDOC Journal of Library & Information Technology	India
6	Gowen & Meier, (2020)	Research data management services and strategic planning in libraries today: A longitudinal study	Journal of Librarianship and Scholarly Communication	United States
7	Kalová, (2020)	Creating Needs-Based Metadata and Research Data Management Services: Exploring the Requirements of Scientists	Young Information Scientist	Austria
8	Liu et al., (2020)	Research data management policies in USA, UK and Australia universities: An online survey	Malaysian Journal of Library & Information Science	USA/UK/ Australia
9	Mushi et al., (2020)	Identifying and implementing relevant research data management services for the library at the University of Dodoma, Tanzania	Data Science Journal	Tanzania
10	Triasih et al., (2020)	Providing research data management services and practices at PDDI-LIPI: Preparedness, roles, challenges, and training	Baca: Jurnal Dokumentasi Dan Informasi	Indonesia
11	Read et al., (2021)	Embracing the value of research data: Introducing the JCHLA/JABSC Data Sharing Policy	Journal of the Canadian Health Libraries Association	Canada
12	Jackson (2021)	Open data policies among library and information science journals	Publications	Global
13	Fu, Ping (2021)	A Research Project of Developing a Research Data Management Service (RDS) Program at CWU: Needs Assessment and Recommendations	Library Scholarship	United States
14	Huang et al., (2021)	Research data management policy and practice in Chinese university libraries	Journal of the Association for Information Science and Technology	China

Table 1: Brief Description of included studies

Sl.no	Authors	Title	Journal	Geographical scope
15	Kim & Syn, (2021)	Practical considerations for a library's research data management services: the case of the National Institutes of Health Library	Journal of the Medical Library Association (JMLA)	United States
16	J. Masinde et al., (2021)	Researchers' Perceptions of Research Data Management Activities at an Academic Library in a Developing Country	International Journal of Library and Information Services (IJLIS)	Kenya
17	Masinde et al., (2021)	Research librarians' experiences of research data management activities at an academic library in a developing country	Data and Information Management	Kenya
18	Mushi, (2021)	Research data management and services: Resources for different data practitioners.	IASSIST Quarterly	Tanzania
19	Reichmann et al., (2021)	Between administration and research: Understanding data management practices in an institutional context	Journal of the Association for Information Science and Technology	Austria
20	Borkakoti & Singh, (2021)	Research data management in central universities and institutes of national importance: A perspective from North East India	Library Philosophy and Practice	India
21	Xu et al., (2022)	Effect of online research data management instruction on social science graduate students RDM skills	Library and Information Science Research	United States
22	M. S. Mthembu & Ocholla, (2023)	Research data management competencies of researchers in selected public universities in South Africa	Library Management	South-Africa
23	Sheffield & Burton, (2022)	Research data management needs assessment of Clemson university	Journal of Librarianship and Scholarly Communication	United States
24	Bishop et al., (2022)	"Data Services Librarians Responsibilities and Perspectives on Research Data Management	Journal of eScience Librarianship	United States
25	Bourke, (2022)	Bibliographic Control of Research Datasets: reflections from the EUI Library	Bibliographic control in the digital ecosystem.	Italy
26	Cheung et al., (2022)	Practices before policy: Research data management behaviours in Canada	Partnership	Canada
27	Fu et al., (2022)	Developing research data management services in a regional comprehensive university: The Case of Central Washington University	IFLA Journal	United States
28	Machimbidza et al., (2022)	Exploring the feasibility of Research Data Management services in Zimbabwean academic libraries: the case of Zimbabwe School of Mines	Library Philosophy and Practice	Zimbabwe

Table 1: Brief Description of included studies

Sl.no	Authors	Title	Journal	Geographical scope
29	Maurya & Subaveerapandiyan, (2022)	Research Data Preservation Practices of Library and Information Science Faculties.	DESIDOC Journal of Library and Information Technology	Asia
30	Mkulama et al., (2022)	Preparedness for Open Science through Research Data Management at the University of Zambia in COVID-19 and Post-COVID Eras.	Library Philosophy and Practice	Zambia
31	Mthembu & Ocholla, (2022)	Development strategies as catalysts for provision of the RDM services in the South African higher education institutions	South African Journal of Libraries and Information Science	South-Africa
32	Yidavalapati et al., (2022)	Research Data Management and Services in South Asian Academic Libraries.	Library Philosophy and Practice	South-Asia
33	Ismail et al., (2022)	Eliciting Researchers' Behaviour as the Foundation of Research Data Management Service Development	"LIBRES: Library and Information Science Research Electronic Journal	Malaysia
34	Rod, (2023)	It Takes a Researcher to Know a Researcher: Academic Librarian Perspectives Regarding Skills and Training for Research Data Support in Canada.	Evidence Based Library and Information Practice	Canada
35	Amanullah & Abrizah, (2023)	The landscape of research data management services in Malaysian academic libraries: librarians practices and roles	The Electronic Library	Malaysia
36	Baudin & Mapulanga, (2023)	A qualitative assessment of the eResearch Knowledge Centres support practices in the Human Sciences Research Council in Pretoria, South Africa	Digital Library Perspectives	South-Africa
37	Nahotko et al., (2023)	Big data-driven investigation into the maturity of library research data services (RDS)	The Journal of Academic Librarianship	Poland
38	Martin-Melon et al., (2023)	Research data services (RDS) in Spanish academic libraries	The Journal of Academic Librarianship	Spain
39	Zhou et al., (2023)	Research data management needs assessment for social sciences graduate students: A mixed methods study	Plos one	United States
40	Mohd Zain et al., (2023)	Academic library opportunity: examining the research data management challenges facing Malaysian information science researchers	Journal of Information and Knowledge Management (JIKM)	Malaysia
41	Bellgard et al., (2023)	RDM+ PM Checklist: Towards a Measure of Your Institutions Preparedness for the Effective Planning of Research Data Management	Data Science Journal	Australia

Table 1: Brief Description of included studies

Sl.no	Authors	Title	Journal	Geographical scope
42	Badenhorst & Raju, (2023)	Research Data Management Competencies for Academic Libraries: Perspectives from Two Universities in South Africa	African Journal of Library, Archives & Information Science	South-Africa
43	Sinha et al., (2023)	Research data management services in academic libraries: A comparative study of South Asia and Southeast Asia	Global Knowledge, Memory and Communication	South-Asia/ South-East Asia
44	Mosha & Ngulube, (2023)	Teaching research data management courses in higher learning institutions in Tanzania	Library Management	Tanzania
45	Kuruppu, (2023)	Users Satisfaction with Factors Affecting Research Data Management Systems in Subject-Related Libraries in Colombo District, Sri Lanka"	Multidisciplinary International Journal of Research and Development	Sri Lanka
46	Maurya & Madhusudhan, (2023)	Research Data Management Services in the Indian Institute of Technology Libraries	Journal of Advancements in Library Sciences	India
47	Chen et al., (2023)	Academic libraries and research data management: a case study of Dataverse global adoption	Information Discovery and Delivery	Global
48	Baek et al, (2024)	Investigating the Roles and Responsibilities of Institutional Signing Official After Data Sharing Policy Reform for Federally Funded Research in the United States: National Survey	JMIR Formative Research	United States
49	Si & Liu, (2024)	Nip risks in the bud: research data ethics governance framework and collaborative network from the perspective of UK policy	Aslib Journal of Information Management	United Kingdom
50	Manco, (2024)	Open science policies as regarded by the communities of researchers from the basic sciences in the scientific periphery	Online Information Review	Brazil and Peru
51	Griffin & Janz, (2024)	Research Data Management Sustainability: Services, Infrastructure, Accountability, and Planning	portal: Libraries and the Academy	United States
52	Asok et al., (2024)	Research data policy: a library and information science publishers perspective	Quality & Quantity	Global
53	Howlett et al., (2024)	Toward Redefining Library Research Support Services in Australia and Aotearoa New Zealand: An Evidence-Based Practice Approach	New Review of Academic Librarianship	Australia/ New Zealand
54	Dessa & Dani, (2024)	An Investigation of Current Research Support Services in Hungarian Academic Libraries	The Canadian Journal of Information and Library Science	Hungary
55	Cooper & Costanzo, (2024)	Developing Institutional Research Data Management Strategies in Canada: Setting the Foundation for Stronger Partnerships and Collaborations	IASSIST Quarterly	Canada
56	Njagi et al., (2024)	Understanding the legal framework in Research Data Management: a study of selected academic libraries in Kenya	Regional Journal of Information and Knowledge Management	Kenya
57	Jose et al., (2024)	State of Research Data Management in Latin American Universities 2022	Procedia Computer Science	Latin America
58	Donner & Huber, (2024)	Research Data Management in German Academia from a Multiple Logic Perspective	M@n@gement	Germany

Table 1: Brief Description of included studies

Table 1: Brief Description of included studies

Sl.no	Authors	Data collection tools	Target population	Sample size	RDM services Results	RDM policy results
1	Lafferty-Hess et al., (2020)	Review University services	Libraries	2	Need data curation services. Assess the effectiveness of service with both qualitative and quantitative tools. Collaboration across institutions is essential for the provision of successful data services.	
2	Hrynaszkiewicz et al., (2020)	Review policies	Publishers	4		Clear and structured policy is necessary and should take care of exemption of . sensitive data, list of data repositories for data submission, data availability statement data format standards for submission
3	Avuglah, (2020)	Interview	Librarians, Professors Research Administrators ICT Managers	7	Limited knowledge on RDM. Need RDM for security, integrity, and reputation. Trust is the main factor in data sharing.	Policy is an important factor for cultural change towards RDM.
4	Beckles, (2020)	QAMyData & sdcMicro	Library	1	Sensitive data should have restricted access. Publication of sensitive data requires discussion with experts, establishment of a dedicated sensitive data support service, with specialist staff	
5	Bunkar & Bhatt, (2020)	Questionnaire	Research Scholar/ Academic Faculty	88	Researchers need service regarding managing, archiving, and storing research data. Library should organize a workshop and training programs on RDM.	Research cell should also take part in policy development.
6	Gowen & Meier, (2020)	Existing Dataset/ Interview	Library Professionals	60	Recent increment in the number of data repositories. Libraries mostly provide advisory services.	
7	Kalová, (2020)	Interview	Researchers	8	Lack of skills in metadata development. Need specific staff for RDM.	
8	Liu et al., (2020)	Review policies	Universities website	100	Libraries provide access to online data repositories.	Funding agency's policy takes the upper hand over institutional policy. Institutional policy focuses on data storage, access, IPR, DMP, sharing, and ownership of data.
9	Mushi et al., (2020)	Interview/ Questionnaire	Directors/ Researchers/ PG students	14	Researchers show interest in DMP and need training for its development.	
10	Triasih et al., (2020)	Questionnaire	Librarins IT staff PDDI-LIPI staff	36	Institutional initiative, researcher requests, data sharing, access, and preservation drive RDM services. Lack of staff, technological infrastructure, and policy make RDM service difficult. The most crucial skills for offering services are data curation, domain understanding, and IT skills.	
11	Reed et al., (2021)	Interview, Feedback Review policies	Authors, Stakeholders, Journals	9/10/1		During policy development, these aspects should be addressed: complexity and support for sensitive data, a repository for data submission, legal issues with data, and data format for submission
12	Jackson (2021)	Website Analysis Review Policy	LIS Journals	201		Elements that explicitly addressed in journal research data policy: Mention the type of research data submitted in repositories. Exemption of sensitive data from open data policy. Licensing & data format use for data submission. Specify data repositories for data submission. Data availability statement before and after submission.

Table 2: Description of results (included studies)

Table 2: Description of results (included studies)

Sl.no	Authors	Data collection tools	Target population	Sample size	RDM services Results	RDM policy results
13	Fu (2021)	Questionnaire	Academic Staff/ Researchers/ Graduate students	69	Researchers want aid in data visualization, finding data repositories, and finding resources and datasets. The library does not provide any RDM service. Different academic disciplines generate data in a variety of formats.	
14	Huang et al., (2021)	Website Analysis/ Questionnaire/ Interview	Librarians	42	Implementation of RDM services in libraries are still in the early stages. Most preferred service is to run a data repository. Drivers of RDM services are the knowledge and skill of librarians and raising awareness of RDM and researchers needs.	
15	Kim & Syn, (2021)	Interview	Librarians	4	Only educational services are provided by libraries: curation, storage, backup, preservation, and security of data. Need support from other stakeholders for providing efficient RDM service.	
16	Masinde et al., (2021)(a)	Questionnaire	Researchers	330	Data is generated in various formats, and researchers store it at their personal level. Researchers have lack knowledge of RDM and are not satisfied with the services of libraries. Technological support is not up to the mark.	There is no dedicated policy on RDM.

Table 2: Description of results (included studies)

Sl.no	Authors	Data collection tools	Target population	Sample size	RDM services Results	RDM policy results
17	Masinde et al., (2021)(b)	Interview	Librarians	7	Accessibility depends on the nature of the data. Need skilled professionals, and high computing infrastructure to provide RDM services. RDM service needs collaboration with other stakeholders such as researchers, and IT cell Challenges in RDM services include non-collaboration, funding, staff, and IT infrastructure.	Policy exists but is not clear on data ownership.
18	Mushi, (2021)	Interview/ Questionnaire	Directors/ Researchers/ PG students	2/6/6	Lack of knowledge on RDM.	
19	Reichmann et al., (2021)	Interview/Questionnaire	Researchers	18/258	Researchers store their data in their local storage. Researchers perspectives on RDM vary from discipline to discipline.	Development of policy should be discipline specific.
20	Borkakoti & Singh, (2021)	Questionnaire	Library professional	28	Libraries do not provide any dedicated RDM services. RDM services enhance the library's image. Professionals feel the need for training on RDM.	Policy development is a collaborative effort, of all stakeholders
21	Xu et al., (2022)	Intervention	Graduate Students	84	Training on RDM enhances knowledge of students. Need RDM training program for graduate students.	

Table 2: Description of results (included studies)

Sl.no	Authors	Data collection tools	Target population	Sample size	RDM services Results	RDM policy results
22	M. S. Mthembu & Ocholla, (2023)	Questionnaire/ interview	"Researchers/ Librarians/ Academic Faculty	38/23	RDM service is on Initial stages. Most researchers are unaware about RDM services library provided. Technical skills required for RDM is data curation, metadata skills, data management and IT skills. Researcher lack knowledge and skills in RDM. Research visibility, data sharing, data storage, and data availability for reuse are all advantages of RDM for researchers.	Policy development is in initial stages.
23	Sheffield & Burton, (2022)	Questionnaire	Academic Staff/ Graduate students	259	Respondents indicated that RDM is not a problem, but the time it takes to do it is a barrier. Respondents want assistance from the library in long-term storage and identifying repositories for data submission. Libraries can promote their RDM service to inform users.	
24	Bishop et al., (2022)	Interview	Data librarians	10	Most libraries have fewer human resources. Data is mostly generated in tabular form. Professionals want to learn new skills on RDM and need training.	Library professionals play an important role in policy development.
25	Bourke, (2022)	Review data services Website Analysis	Library	1	Libraries should assist researchers in DMP, data submission to repositories, metadata creation, legal issues with data, and standard file naming.	
26	Cheung et al., (2022)	Questionnaire	Researchers	2658	Data is generated during research in various formats. Researchers store various types of research data at their personal level and share data on personal request. Researchers want RDM services with a focus on aid in DMP.	

Table 2: Description of results (included studies)

Sl.no	Authors	Data collection tools	Target population	Sample size	RDM services Results	RDM policy results
27	Fu et al., (2022)	Questionnaire	Academic Staff/ Researchers/ Graduate students	2025	RDM implementation is a step-wise process. Data is generated in various formats, and researchers store data at their personal level. Different discipline researchers had different needs for data services.	
28	Machimbidza et al., (2022)	Interview	IT manager Librarians Assistant Librarians	1/1/3	No technical infrastructure. Only provided advisory services, data curation, and dissemination. Researchers store data at their personal level. Professional expressed training needed on RDM.	There is no legal policy on RDM.
29	Maurya & Subaveerapandiyan, (2022)	Questionnaire	Library professionals	125	Data is generated in various formats, and researchers store it on their personal level. Expected services by researchers are data discovery, data preservation, data sharing, aid in metadata generation, and data visualization.	
30	Mkulama et al., (2022)	Interview	Librarians, Directors Registrar	5	Library professionals lack knowledge about RDM. Also reported a lack of technical skills required for providing RDM services.	
31	Mthembu & Ocholla, (2022)	Interview	Librarian, Technician, HOD Deputy V.C.	2/1/4/1	Lack of human resources and technical infrastructure. Professional show training is needed on RDM.	Policy development is a collaborative effort among stakeholders. Policy is in the development phase.

Table 2: Description of results (included studies)

Sl.no	Authors	Data collection tools	Target population	Sample size	RDM services Results	RDM policy results
32	Yidavalapati et al., (2022)	Questionnaire	Library professional	67	Need technical skills and knowledge of legal aspects to provide RDM services. Challenges associated with RDM are lack of skilled staff, legal issues with data, technical support, and technical infrastructure. Library staff should update their skills.	
33	Ismail et al., (2022)	Interview/ Questionnaire	Researchers	12/113	Researchers store data without metadata. Confidentiality and privacy are the major issues in data sharing. RDM workshop that library organized only deals with advocacy programmes.	
34	Rod, (2023)	Interview	Librarians	12	Identified four skills for the role of data librarians: Experience of conducting original research. Proficiency in data analysis through computation coding. MLIS-related skills including metadata understanding, documentation, preservation, and curation. Ability to learn new skills quickly.	
35	Amanullah & Abrizah, (2023)	Website Analysis/ Online Survey/ Semi-structured interview	Libraries/ Librarians	20/3	Libraries mostly provide advisory services Lack of knowledge about RDM, and needs training on technical skills required for providing RDM services. Librarians play a leadership role in implementing RDM in institutions.	Many institutions lack policies that hinder the deployment of RDM implementation.

Table 2: Description of results (included studies)

Sl.no	Authors	Data collection tools	Target population	Sample size	RDM services Results	RDM policy results
36	Baudin & Mapulanga, (2023)	Semi-Structured Interview	Library Professionals	9	Identified services related to data that researchers require: support, training, and tools required in data analysis. assistance in the creation of metadata. assistance in making data submission in the open domain. Libraries only provide advisory service on RDM.	
37	Nahotko et al., (2023)	Website Analysis	Libraries	80	Libraries offer mostly advisory RDM services. Librarians must improve their competencies in data analysis, data visualization, and data processing. A small number of libraries provide advanced RDM services.	
38	Martin-Melon et al., (2023)	Website Analysis	Libraries	48	Libraries provided advisory services on RDM. Technical services related to data processing, data analysis, and data visualization are not offered. Lack of adequate infrastructure for RDM implementation. Challenges in RDM: the heterogeneous nature of data, culture of data sharing is lacking, sensitive data.	A few libraries have recently developed RDM policies, that are focused on open research data.

Table 2: Description of results (included studies)

Sl.no	Authors	Data collection tools	Target population	Sample size	RDM services Results	RDM policy results
39	Zhou et al., (2023)	Interview	Graduate Students	98	International students reported lower confidence compared to United States students. Most participants are unaware of the RDM service on campus. Students need training on various concepts of RDM: data processing and organization, data preservation, data sharing, and data preservation.	
40	Zain et al., (2023)	Interview	Researchers	5	Challenges in RDM data storage include the misuse of open data Need tools and support for data analysis and visualization and a secure channel for communicating research data. Libraries can offer support, training, and tools to researchers on effective RDM.	
41	Bellgard et al., (2023)	Developed a Tool RDM-PM checklist	-	-	Challenges in RDM: Due to variation in discipline-wise data and data formats, clear policy, stakeholder do not know clear role and responsibilities. RDM checklist is an effective tool for RDM and its whole process from data generation to data submission.	Researchers are unaware of their institutional policy.
42	Badenhorst & Raju, (2023)	Interview/ Content analysis	Librarians Job advertisement (RDM)	10/31	Essential competencies and skills are needed for librarians to deliver effective RDM services: Knowledge of data management practice, data curation, various research methods, and funding agencies policies. ICT skills in data analysis, visualization, digitization, and preparing datasets for submission. Willingness for continuous learning.	

Table 2: Description of results (included studies)

Sl.no	Authors	Data collection tools	Target population	Sample size	RDM services Results	RDM policy results
43	Sinha et al., (2023)	Questionnaire	Library Professionals	228	Libraries provide mostly advisory services and less focus on technical services. Skills required to provide effective RDM are knowledge of metadata, locating data from data sources, data analysis, and data visualization. Challenges associated with providing RDM: lack of awareness, lack of funding, lack of skilled professionals.	
44	Mosha & Ngulube, (2023)	Questionnaire/ Interview	Postgraduate Students	70	Need for RDM courses in higher education institutions. RDM courses should include theoretical and . hands-on activities Students are not aware of RDM and its associated concepts.	Funder policy regarding the submission of data to the open domain is an important factor that can encourage users for effective RDM.
45	Kuruppu, (2023)	Questionnaire	Library Users	293	Factors influencing user satisfaction with RDM services include IT infrastructure, technical knowledge of RDM and the IT skills of library staff.	
46	Maurya & Madhusudhan, (2023)	Questionnaire	Library Professionals	74	Institutional servers function as secure places for data storage. Challenges in providing RDM services are data security, IPR issues, lack of trained professionals, and inadequate institutional RDM policy. Libraries must market their services.	
47	Chen et al., (2023)	Website Analysis	Academic/ Research Institutions	67	Most members of Dataverse repositories are developed nations. Top-ranking universities provide some kind of RDM services. Majority of data repositories were developed recently.	

Table 2: Description of results (included studies)

Sl.no	Authors	Data collection tools	Target population	Sample size	RDM services Results	RDM policy results
48	Baek et al, (2024)	Questionnaire	Stakeholders (Signing officials)	56	Lack of clarity regarding the location for data submission following project completion. Need dedicated staff to verify the authorized users of data. Need to store data in a centralized location.	
49	Si & Liu, (2023)	Grounded Theory/ Social network analysis	Policies	78		Existing policies are lacking in enforceable standards. Policies explicitly outline the data type encompassed within policy, as well as the institutions and related data governance policies.
50	Manco, (2024)	Interview	Researchers	17	Researchers encounter various challenges during research such as: securing funding as a result of political changes within the country. funding disparities within institutions.	
51	Griffin & Janz, (2024)	Questionnaire	Library Professionals	118	Skills required for providing RDM service, such as technical software knowledge, data management principles, data science skills, coding, and programming abilities. Library mostly provided advisory services. Lack of funding, cultural barriers to data exchange, and a shortage of qualified personnel are obstacles to providing RDM services.	
52	Asok et al., (2024)	Review Policy	Publishers	12		Most policies focus in providing repository location for data submission, provide a data availability statement (DAS), and use an open standard for data format. To make data sharing easier, policies from many publications must be harmonized.

Table 2: Description of results (included studies)

Sl.no	Authors	Data collection tools	Target population	Sample size	RDM services Results	RDM policy results
53	Howlett et al., (2024)	Website analysis/	University Libraries	44	Most emerging research support services that libraries provide are assistance in DMP, training and workshops on data management, data visualization tools and software, open data practices, and information on data repositories.	
54	Dessa & Dani, (2024)	Website Analysis	Universities Libraries	10	Libraries do not mention their RDM services on their websites. Most libraries provide only advisory services, not technical services such as data deposit in repositories, data sharing, and policy on RDM.	
55	Cooper & Costanzo, 2024	Questionnaire	Research institutions	92		In response of Canda Tri-Agency policy (2021) Many institutions formed working groups, and started developing RDM strategies. Stakeholders involvement libraries, research office, IT service, ethics board is crucial for effective RDM strategy implementation Major barriers in RDM lack of time, budget, awareness, and supporting resources
56	Njagi et al., (2024)	Questionnaire	Librarians	35	Major challenges in RDM services are IPR, data security, sensitive data confidentiality, and restrictions on using research data.	An explicit policy can handle RDM challenges by addressing legal issues, providing guidelines for data storage and data usage, and including the role and responsibility of different stakeholders.
57	Jose et al., (2024)	Questionnaire	Researchers	652	Need need for RDM training for researchers is due to the following reasons: data is stored without backup and security, lack of know how to process data, and awareness of the legal aspect of data Insufficient infrastructure to facilitate RDM services.	

Table 2: Description of results (included studies)

Sl.no	Authors	Data collection tools	Target population	Sample size	RDM services Results	RDM policy results
58	Donner & Huber, (2024)	Interviews	Researchers	40	Researchers in the early stage of the career avoid RDM due to competitive pressure. limited time and a low level of awareness about RDM Supportive environments of institutions and supervisors can significantly promote RDM among researchers.	

Table 2: Description of results (included studies)

Blind Peer Reviewed
(Previously UGC-CARE)
ISSN: 0972-2750

Manuscript Received on –
28th December, 2025
Primary Reviewed on –
4th January, 2026
Peer Reviewed on –
12th January, 2026
Revised on –
28th January, 2026
Accepted for publication
5th February, 2026

Envisioning the Future of Sustenance of Pattachitra: A Study on the Patua Community of Raghurajpur Village in Puri, Odisha, in their Information Requirement Landscape

Debalina Mukherjee, Arabinda Maity & Biplab Chakrabarti

ABSTRACT

Introduction: *Pattachitra is a traditional art form in India that depicts our rich culture through its paintings, disseminating information on history, mythology, social and cultural events, and current issues.*

Purpose: *Despite its significant value, the Patua of Raghurajpur had been facing a survival threat due to its low socio-economic status. The authors had tried to examine the socio-economic condition of the Patua community through the product commercialisation in the information requirement landscape at Raghurajpur village in Puri, Odisha. This study revealed major socio-economic challenges and specific information requirements for the sustainability of Pattachitra.*

Research problem: *The main problem is the lack of significant literature on the Patua Community, which has not identified its specific information requirements for professional development.*

Objectives: *The main objective of this study is to determine their sustenance in future based on their information requirements for professional development.*

Methodology: *The survey method was used with the help of a questionnaire and interview methods. The observation method was also a part of the data collection. After the data collection procedure, these were tabulated and analysed. On the basis of the analysis, findings and discussion were drawn. To better understand, the Kolmogorov-Smirnov test was used to test different hypotheses, which identified the important factors for product commercialisation.*

Findings: *This study found that the Patua community was information-disadvantaged due to its low socio-economic condition. They had major information requirements regarding marketing of the products, which involved financial assistance, marketing and selling places of the product, training, raw materials, infrastructure and transportation.*

Originality: *This research was conducted on the basis of personal field visits, and primary data were collected by researchers through visiting the houses of the respondents in the village to identify the findings originally.*

Conclusion: *It has been found that information requirements differ across communities, depending on socio-economic conditions. Identification of information needs for an individual community is crucial to provide exact information services to the community. This study clearly demonstrated the information requirements of the Patua Community across various aspects, which are directly linked to product marketing and commercialisation.*

Keywords: *Information needs, Pattachitra, Patua Community, Product Commercialisation*

Mrs. Debalina Mukherjee
PhD Research Scholar
Dept. of Lib. & Inf. Sc.,
University of Calcutta, Kolkata
debalinamukherjee0809@gmail.com

Dr. Arabinda Maity
Professor
Dept. of Lib. & Inf. Sc.,
University of Calcutta, Kolkata
maityarabindacu@gmail.com

Dr. Biplab Chakrabarty
Retired Professor
Dept. of Lib. & Inf. Sc.,
University of Calcutta
bcliscu@gmail.com

1. Introduction:

India has a rich cultural heritage of artwork. The art and artists are the invincible, invisible pillars of the Indian economy. The country's traditional art form tells the story of our history and shows the value of our knowledge. Pattachitra is a traditional art form of India that depicts history, mythology, contemporary issues, and other events in the form of paintings for the common people. Pattachitra is a kind of information dissemination channel. Raghurajpur is a heritage village in Puri district, Odisha, that has been carrying this affluent tradition from generation to generation.

Other places in Odisha where this painting has been made are Paralakhi Mundi, Chikiti, Dana Sahi, and Sonepur. The Patta painting in Odisha is linked with Hindu mythology because of the sacred place and the Jagannath Temple at Puri. The word 'Pata' came from the Sanskrit word 'Patta', which means cloths, and 'Chitra' means paintings. In Raghurajpur, there are three types of Pat paintings: canvas painting (Patta - Chitra), wall painting (Bhitti - Chitra) and palm leaf painting (Pothi - Chitra). Nowadays, the Patuas are using different types of canvases for market demand, like utensils (Das, **2023). Mahalik (2024)** mentioned in his study that Raghurajpur achieved its heritage status as a Heritage village in 2000 by the Indian National Trust for Art and Cultural Heritage (INTACH).

Though this is the oldest traditional art form of India, the socio-economic scenario of the Patuas is indigent. **Mahalik (2024)** said that the Ministry of Textiles reported in 2011 that "the artisan class form a major stakeholder of the non-farm rural economy in India with an estimated 7 million artisans according to official figures (and up to 200 million artisans according to unofficial sources) engaged in craft production to earn a livelihood. The artisan is an important link between the Indian society and culture". Every person in this village is connected with different types of artwork. In the 21st century, information is considered a power to regulate every aspect of life. In every corner of society, information plays a vital role in keeping society alive. Information requirements have been used in different ways, like needs, demands and wants. **Chinnasamy (2016)** explained that according to Kuhlthau (1993), information need "as evolving from a vague awareness of something missing and as culminating in locating information that contributes to understanding and meaning". Bradshaw described in his classic taxonomy of social needs that the meaning of *expressed needs* is when people can identify and seek assistance or services. The meaning of unexpressed *needs* is when people do not identify what is required to solve any problem. Sometimes he feels that the needs are very sensitive in nature V**erdecias, 2021)**. But in the present scenario, if the Patuas want to survive and want to uplift their socio-economic and socio-cultural status, the identification of information needs is very important to address their problems. The authors of this research work have portrayed the present socio-economic condition of Patuas, and they identified their information requirements for their socio-cultural and professional development through product commercialisation.

2. Objectives:

The major objectives are:

I. To portray the socio-economic condition of the Patua community in Raghurajpur village

II. To identify the information needs of the Patua community in various aspects of life

III. To find out the demand for information for the upliftment of their professional sector and product commercialisation.

3. Significance of the study: Patuas paint stories of history, mythology and religious pictures through which they present a rich knowledge base of our country. It showcases Indian Knowledge Systems in ancient India. Pattachitra art is a great repository of information which provides information and knowledge of the past and preserves it for the present and future. Identification of information needs of Patuas is important to reveal the behind-the-scenes struggles for making Pattachitra. The study also depicted how they are surviving in the age of ICT and use the technology for their sustainable development.

4. Literature review:

Das (2023) gave a detailed explanation of Pattachitra painting and its process in her article. Richa (2025) told in her article about the presentation of stories through Pat's painting. The cultural differences between Odisha Pattachitra and West Bengal Pattachitra have been discussed by Majumdar (2025). The process of digitisation of motifs painted on Pattachitra has been discussed by Rath (2023). Kanungo (2019) talked about the paucity of funds for Pattachitra making, and they conducted a SWOT analysis to make a policy framework. Mishra (2024) explained women's entrepreneurship through Pata painting. Chakrabarti (2006) supervised research work on the information-seeking behaviour of the Solapith community in the Burdwan district. Chakraborty et al. (2016) depicted a picture of the information needs of the weaver community of Madhya Sualkuchi of Assam for their professional upliftment. Mallik (2022) talked about the information-seeking patterns of the Lodha community in West Bengal. A study was done to assess the information-seeking behaviour and sources used for information looking for by Raising et al. (2021).

5. Socio-economic and cultural profile of Patua community: Raghurajpur is a village in Odisha, situated in the Puri district on the bank of the Bhargavi River. This village is renowned as the "Art Village of India". The village was recognised by UNESCO Heritage Village, which has been carrying the tradition of Pattachitra painting, Gotipaua dance and Stone and Wood Carvings.

Inside the village, there are two temples, Raghunath Temple (or Krishna Temple) and Maa Bhuasuni Temple. Most people in this village belong to a Hindu family. The main profession of these people in this village is Pata paintings. Besides that, they are often engaged in agriculture and animal husbandry. During the Covid period, the Patuas got an amount of ten thousand per family from the Govt of Odisha to paint the house with rich stories to attract more tourists. The Patuas of this village mainly depend on tourists for marketing their products. They trade their Pata from home and online platforms. Apart from this, they visit different fairs and art exhibitions at national and international levels. Pattachitra painting is the main source of income.

The Kapoteswar Upgraded School is the nearby school for primary and secondary education. For the purpose of higher education, they need to go to Puri, Bhubaneswar, or Cuttack.

6. Hypotheses of the study:

The hypotheses are stated below:

Hypothesis 1: There are no significant differences in the responses given by the respondents on the crucial factor "Fund", which is crucial for product commercialisation.

Hypothesis 2: There are no significant differences in the responses given by the respondents on the major factor "Marketing/selling", which has great usefulness for product commercialisation.

Hypothesis 3: There are no significant differences in the responses given by the respondents on the major factor "Infrastructure" that influences product commercialization

Hypothesis 4: There are no significant differences in the responses given by the respondents on the major factor "Training", which is important for product commercialisation.

Hypothesis 5: There are no significant differences in the responses given by the respondents on the major factor "Raw material", which is a basic need for product commercialisation.

Hypothesis 6: There are no significant differences in the responses given by the respondents on the major factor "Design ", which is needed for product upgradation

Hypothesis 7: There are no significant differences in the responses given by the respondents on the major factor "Transportation", which is needed for product marketing

7. Methodology:

Scope and coverage: Raghurajpur village had 160 families, and the total population was 856. Among 856 people, the total male population was 476, and the female population was 380. Among 856 people, the total number of artisans or Patuas were 615 (Male -372 and female-243). The authors had taken the total population for this study. A sample consisting of 615 respondents participated in this study. Among 615 Patuas, 350 people were available during the time of the survey. So, 350 people were considered as respondents for this survey (N=350). Among 350 people, 200 people were women and 150 people were men.

Method used: The research was conducted on the basis of a literature search to identify the research gap. Then the authors collected the local history from different sources to get a brief idea about the Pattachitra and Patua communities. After that, the scope and coverage were decided for the study. The data was collected by using observation and interview methods. For data collection, a semi-structured questionnaire was used. Then the data had been represented in a table format with percentages. For better understanding, the Kolmogorov-Smirnov test was performed to show the actual necessities of information requirements on major factors.

8. Data analysis and findings:

Total Number of respondents: 350

Sl No.	Category		Responses	Percentage
1.	Sex	Male	150	42.85%
		Female	200	57.14%
2.	Age	18-28	107	30.57%
		29-38	90	25.71%
		39-48	58	16.57%
		49-58	63	18.00%
		58 Above	32	9.14%

Table 8.1: Demographic profile of respondents (N=350)

Analysis: After analysing the survey data, it was identified that the majority of respondents were female (57.14%) and 42.85% were male in the Patua community. The composition of the age group showed that 30.57% Patua belonged to the 18-28 group, which was the greatest among all distributions. 25.71% Patua belonged to the 29-38 age group, which was the second-largest number of Patua. 16.57% respondents fell under the middle age group (39-48).18% respondents belonged to the 49-58 age group, and only 9.14% respondents were above 58 years old. The distributions represented that the young generations had an interest in taking this age-old art form as a profession, which came from generation after generation.

			Responses	Percentage
Qualifications	Academic	<10	212	60.57%
		10	95	27.14%
		12	17	4.85%
		UG	10	2.85%
		PG	-	
		Ph.D.	-	
	Professional	BFA, MFA	-	

Table 8.2: Distribution of respondents according to their educational qualifications(N=350)

Analysis: 60.57% respondents had a qualification below 10th class. 27.14% Patua cleared 10th class, and only 4.85% completed 12th class. 2.85% Patua went to college. These distributions highlighted a low literacy rate among the Patua community, which was a big concern for their sustenance.

	Types of languages	Responses	Percentage
Languages known	Odia	350	100%
	Hindi	245	70%
	English	32	9.14%
	Others (Especially Bengali)	20	5.71%

Table 8.3: Distribution of respondents according to their comprehension of language(N=350)

	Category	Responses	Percentage
Purpose of the information requirement	Day-to-day activity	350	100%
	Professional development	335	95.71%
	Awareness	347	99.14%

Table 8.4: Distribution of respondents according to their purpose of information requirement(N=350)

Analysis: Every respondent gave the preference to gather information from different sources to perform day-to-day activities (100%). 95.71% of respondents said they required information for professional development, and 99.14% of Patua answered that they had a demand for information on awareness, such as health, legal, and financial.

		Category	Responses	Percentage
Sources of information	Formal	Book	43	12.28%
		Newspapers	156	44.57%
		NGO	-	
		Govt office	169	48.28%
	Informal	Coworker	336	96.00%
		Family members/friends	347	99.14%
		Social media	298	85.14%

Table 8.5: Distribution of respondents according to their usage of sources for information requirement(N=350)

Analysis: It was observed that coworkers (96.00%), family members (99.14%) and friends were the most trustworthy sources for gathering information, although they were informal sources. 48.28% Patua sought information from various government offices like Panchayet, Block Development Officer, etc. 85.14% people operate social networks for information seeking in spite of a low literacy rate; the Patua people knew how to handle social media. No non-governmental organisations provided any useful information for Patuas for their occupational and personal development. 44.57% Patua used newspapers for seeking information, and only 12.28% people used books as one of their information sources.

		Responses	Percentage
Usage of the library/information centre	Yes	79	22.57%
	No	271	77.42%

Table 8.6: Distribution of respondents according to their usage of library and information centres (N=350)

Analysis: A majority of respondents (77.42%) did not access any library or information center for seeking their information. Only 22.57% knew about the library for getting information.

		Responses	Percentage
Satisfaction level of the profession	I love it	256	73.14%
	I like it	94	26.85%
	No comments	-	
	I dislike it	-	
	I hate it	-	

Table 8.7: Distribution of respondents according to their professional satisfaction (N=350)

Analysis: It was shown that despite heavy demands on different factors maximum Patuas were satisfied with their occupation (73.14%).

	Category	Responses	Percentage
Reason for choosing another profession	Less income	276	78.85%
	Less scope for professional development	214	61.14%
	Lack of training	294	84.00%
	Lack of motivation	49	14.00%
	Lack of fund	319	91.14%

Table 8.8: Distribution of respondents according to their information requirement for choosing another profession (N=350)

Analysis: 78.85% of Patua responded that lower income was a major factor in choosing another profession. 91.14% Patuas gave preferences that the lack of availability of funds was another big concern. Lack of proper training was another reason 84.00% Patuas. Lack of motivation was another cause of 14% Patuas.

Product price-range	Category	Responses	Percentage
	<500	350	100%
	501-5000	325	92.85%
	5001-10000	285	81.42%
	>10000	149	42.57%

Table 8.9: Distribution of respondents according to their opinion for product price range (N=350)

Analysis: It was discovered that the product selling price was comparatively low, which fell under the lower and mid-price structure. Keeping in mind the economic background of buyers, the Patuas made various types of paintings, such as small bookmarks, paperweights, and different types of cards, which enhanced sales. Almost every Patua made this type of product within a price range of under 500 rupees. 92.85% respondents made products under Rs. 501 to Rs. 5000. 81% of Patuas created Pat under Rs. 500-Rs. 9999. Usually, they make more critical paintings on cloth and paper. Only 42.57% of Patuas, especially seniors and experienced Patuas, created products worth more than Rs. 10,000.

		Category	Responses	Percentage
Product selling place	Market	National	329	94.00%
		International	10	2.85%
	Fair	342	97.71%	
	Home	335	95.71%	

Table 8.10: Distribution of respondents according to their information requirement for product selling place (N=350)

Analysis: Selling of product from home (95.71%) and in a fair (97.71%) were the most common places according to Patuas. 94.00% of Patuas sought information about the national marketplace, which was appropriate for them, and only 2.85% Patuas had interest in international marketing.

	Category	Responses	Percentage
Monthly income	<4999	142	40.57%
	5000-9999	154	44.00%
	10000-24999	41	11.71%
	25000-49999	10	2.85%
	≥50000	3	0.85%

Table 8.11: Distribution of respondents according to their opinion for monthly income (N=350)

Analysis: The survey results revealed that most of the respondents had a low-income profile. Under this category, 40.57% Patuas earned below Rs. 4999 per month, and 44.00% Patuas earned between Rs 5000 and Rs. 9999. 11.71% of Patuas earned Rs. 10000-24999 per month. Only 2.85% Patuas managed to earn within Rs. 25000-49999. A negligible proportion (0.85%) of Patuas earned more than Rs. 50000 per month.

	Category	Responses	Percentage
Places for getting financial assistance	Govt.	160	45.71%
	NGO	125	35.71%
	Own savings	342	97.71%

Table 8.12: Distribution of respondents according to their opinion for getting financial assistance (N=350)

Analysis: This table showed that 97.71% Patuas used their own earnings for the product making. 45.71% Patuas required Government assistance for funds.

Award of appreciation	Category	Responses	Percentage
	Yes	346	98.85%
	No	4	1.14%

Table 8.13: Distribution of respondents based on their information requirement for award of appreciation (N=350)

Analysis: The above table depicts that 98.95% Patuas wanted the information for the award. It helped to encourage them.

Learning to make a product	Category	Responses	Percentage
	Traditionally	345	98.57%
	Own idea	319	91.14%
	Training	43	12.28%
	Art college	-	

Table 8.14: Distribution of respondents according to their opinion for making the product (N=350)

Analysis: Basically, the survey data showed that every artisan carried on this profession traditionally, and a majority of artisans used their own ideas for painting.

Packaging system	Category	Responses	Percentage
	Newspapers wrap	350	100%
	Thermocol box	248	70.85%
	Wooden and pitch board	130	37.14%
	Plastic bags	326	93.14%
	Others	41	11.71%

Table 8.15: Distribution of respondents according to their opinion for packaging system (N=350)

Analysis: Newspaper wrap (100%) and Plastic bags (93.14%) were mostly used products for packaging because of their low cost and ease of getting them. Thermocol box (70.85%) and Wooden and pitch board (37.14%) were also used for the protection of large and fragile items.

Awareness about the government. scheme	Category	Responses	Percentage
	Yes	138	39.42%
	No	212	60.57%

Table 8.16: Distribution of respondents according to their opinion for packaging system (N=350)

Analysis: A majority of respondents, 60.57%, were not aware of any type of Government schemes. Only 39.42% respondents knew about different Government schemes.

9. Hypothesis test of major factors accountable for product commercialisation:

The major factors for product commercialisation are Fund, Marketing or Selling of Products, Infrastructure, Training, Raw material, Design and Transportation. To know their importance for product commercialisation, the Kolmogorov-Smirnov (K.S) test was done based on responses given by Patuas(respondents). The responses were dissected into five categories: Strongly Agreed (SA), Agreed (A), No Opinion (NO), Disagree (DA) and Strongly Disagree (SDA). The Cumulative Observed Proportion (O) was determined based on the observed number in K.S Test. In each case, the Observed Proportions were determined by dividing the number of responses in each category by the total responses in this particular case. The five categories are there, and each category (i.e. 0.20) was assigned as the expected proportion. The Cumulative Expected Proportion (E) was determined based on the Expected Proportion. The Difference (D) between the cumulative observed proportion (O) and the cumulative expected proportion (E) was calculated. The biggest Difference (Dmax) was considered the computed value for the test of the hypothesis.

Purpose of the information requirement	Category	Responses	Percentage
	Day-to-day activity	350	100%
	Professional development	335	95.71%
	Awareness	347	99.14%

9.1: Distribution of responses for different factors responsible for product commercialisation

Information requirement for product Commercialization	Category	Strongly agree	Agree	No	Disagree	Strongly disagree
	Fund	158	175	10	3	4
	Marketing/selling	238	82	13	15	2
	Infrastructure	123	196	16	12	3
	Training	94	127	45	56	28

	Raw material	26	301	9	11	3
	Design	245	84	12	5	4
	Transportation	299	42	2	5	2

9.1: Distribution of responses for different factors responsible for product commercialisation

The major factors for product commercialisation are Fund, Marketing or Selling of Products, Infrastructure, Training, Raw material, Design and Transportation. To know their importance for product commercialisation, the Kolmogorov-Smirnov (K.S) test was done based on responses given by Patuas(respondents). The responses were dissected into five categories: Strongly Agreed (SA), Agreed (A), No Opinion (NO), Disagree (DA) and Strongly Disagree (SDA).

The Cumulative Observed Proportion (O) was determined based on the observed number in K.S Test. In each case, the Observed Proportions were determined by dividing the number of responses in each category by the total responses in this particular case. The five categories are there, and each category (i.e. 0.20) was assigned as the expected proportion. The Cumulative Expected Proportion (E) was determined based on the Expected Proportion. The Difference (D) between the cumulative observed proportion (O) and the cumulative expected proportion (E) was calculated. The biggest Difference (Dmax) was considered the computed value for the test of the hypothesis.

Sl No	Opinion	Observed Number	Observed Proportion	Cumulative Observed Proportion (O)	Expected Proportion	Cumulative Expected Proportion (E)	D= |O-E|	D= O-E
1.	Strongly Agree (SA)	158	0.45142857	0.45142857	0.2	0.2	0.25142857	0.25142857
2.	Agree (A)	175	0.5	0.95142857	0.2	0.4	0.55142857	0.55142857
3.	No Opinion (NO)	10	0.02857142	0.97999999	0.2	0.6	0.37999999	0.37999999
4.	Disagree (DA)	3	0.00857142	0.98857141	0.2	0.8	0.18857141	0.18857141
5.	Strongly Disagree (SDA) (SDA)	4	0.01142857	1	0.2	1	0	0

Table 9.1.1: Calculation of hypothesis 1, which determines the significance of "Fund" in product commercialisation

The null hypothesis states that there is no significant difference in the responses recorded from the Patuas on the factor "Fund". The value given in the table at a 95 per cent confidence level is equal to 1.36 / √350 = 0.003885. Since the calculated value (Dmax = 0.55142857) is bigger than the value given in the table (0.003885), the null hypothesis has been rejected. Hence, there is a significant difference among the responses given by the Patuas on 'Fund'. From the above table, it can be shown that since the actual differences (without the mode value) are positive for most of the cases, the replies are inclined towards 'Strongly Agree'. We can infer that Fund is a crucial factor for Pattachitra product commercialisation.

Sl.No.	Opinion	Observed Number	Observed Proportion	Cumulative Observed Proportion (O)	Expected Proportion	Cumulative Expected Proportion (E)	D= |O-E|	D= O-E
1	Strongly Agree (SA)	238	0.68	0.68	0.2	0.2	0.48	0.48
2	Agree (A)	82	0.23428571	0.91428571	0.2	0.4	0.51428571	0.51428571

3	No Opinion (NO)	13	0.03714285	0.95142856	0.2	0.6	0.35142856	0.35142856
4	Disagree (DA)	15	0.04285714	0.9942857	0.2	0.8	0.1942857	0.1942857
5	Strongly Disagree (SDA)	2	0.00571428	1	0.2	1	0	0

Table 9.1.2: Calculation of hypothesis 2, which determines the significance of "Marketing/Selling" in product commercialisation

The proposed null hypothesis shows that there is no significant difference in the responses given by the respondents on the factor "Marketing/Selling". The value given in the table at a 95 per cent confidence level is equal to 1.36 / √350 = 0.003885. Since the calculated value (Dmax = 0.51428571) is bigger than the value given in the table (0.003885), the null hypothesis has been rejected. Hence, there is a significant difference among the responses provided by the artisans on 'Marketing/Selling'. From the above table, it can be determined that since the actual differences (without the mode value) are positive for most of the cases, the responses are supported towards 'Strongly Agree'. We can conclude that the marketing/Selling of the product is an important factor for the Pattachitra product commercialisation.

Sl.No.	Opinion	Observed Number	Observed Proportion	Cumulative Observed Proportion (O)	Expected Proportion	Cumulative Expected Proportion (E)	D= |O-E|	D= O-E
1	Strongly Agree (SA)	123	0.35142857	0.35142857	0.2	0.2	0.15142857	0.15142857
2	Agree (A)	196	0.56	0.91142857	0.2	0.4	0.51142857	0.51142857
3	No Opinion (NO)	16	0.04571428	0.95714285	0.2	0.6	0.35714285	0.35714285
4	Disagree (DA)	12	0.03428571	0.99142855	0.2	0.8	0.19142855	0.19142855
5	Strongly Disagree (SDA)	3	0.00857142	1	0.2	1	0	0

Table 9.1.3: Calculation of hypothesis 3, which determines the significance of "Infrastructure" in product commercialisation

The null hypothesis shows that there is no significant difference in the responses given by the respondents on the factor "Infrastructure". The table value at a 95 per cent confidence level is equal to 1.36 / √350 = 0.003885. As the calculated value (Dmax = 0.51142857) is bigger than the table value (0.003885), the null hypothesis is rejected. Hence, there is a significant difference among the responses given by the artisans on 'Infrastructure'. From the above table, it can be noticed that since the actual differences (without the mode value) are positive for most of the cases, the responses are inclined towards 'Strongly Agree'. From this analysis, we can say that infrastructure for making the product is a vital factor for Pattachitra product commercialisation.

Sl.No.	Opinion	Observed Number	Observed Proportion	Cumulative Observed Proportion (O)	Expected Proportion	Cumulative Expected Proportion (E)	D= |O-E|	D= O-E
1	Strongly Agree	94	0.26857142	0.26857142	0.2	0.2	0.06857142	0.06857142

	(SA)							
2	Agree (A)	127	0.36285714	0.63142856	0.2	0.4	0.23142856	0.23142856
3	No Opinion (NO)	45	0.12857142	0.75999998	0.2	0.6	0.15999998	0.15999998
4	Disagree (DA)	56	0.16	0.91999998	0.2	0.8	0.11999998	0.11999998
5	Strongly Disagree (SDA)	28	0.08	1	0.2	1	0	0

Table 9.1.4: Detailed calculation of hypothesis 4, which determines the significance of "Training" in product commercialisation

The null hypothesis depicts that there is no significant difference in the responses given by the respondents on the factor "Training". The table value at a 95 per cent confidence level is equal to 1.36 / √350 = 0.003885. As the calculated value (Dmax = 0.23142856) is bigger than the table value (0.003885), the null hypothesis is rejected. Hence, there is a significant difference among the responses given by the artisans on 'Training'. From the above table, it can be determined that since the actual differences (without the mode value) are positive for most of the cases, the responses are prone towards 'Strongly Agree'. We can infer that Training is a crucial factor for Pattachitra product commercialisation.

Sl No	Opinion	Observed Number	Observed Proportion	Cumulative Observed Proportion (O)	Expected Proportion	Cumulative Expected Proportion (E)	D= |O-E|	D= O-E
1.	Strongly Agree (SA)	26	0.07428571	0.07428571	0.2	0.2	0.12571429	0.12571429
2.	Agree (A)	301	0.86	0.93428571	0.2	0.4	0.53428571	0.53428571
3.	No Opinion (NO)	9	0.02571428	0.95999999	0.2	0.6	0.35999999	0.35999999
4.	Disagree (DA)	11	0.03142857	0.99142856	0.2	0.8	0.19142856	0.19142856
5.	Strongly Disagree (SDA)	3	0.00857142	1	0.2	1	0	0

Table 9.1.5: Calculation of hypothesis 5, which determines the importance of "Raw material" in product commercialisation

The null hypothesis reveals that there is no significant difference in the responses collected from the respondents on the factor "Raw material". The value provided in the table at a 95 per cent confidence level is equal to 1.36 / √350 = 0.003885. Since the computed value (Dmax = 0.53428571) is bigger than the value provided in the table (0.003885), the null hypothesis has been rejected. Thus, there lies a significant difference among the responses collected from the artisans on 'Raw material'. From the above table, it has been confirmed that since the actual differences (without the mode value) are positive for most of the cases, the responses are inclined towards 'Strongly Agree'. Thus, it can be inferred that Raw material is an important factor for Pattachitra product commercialisation.

Sl No.	Opinion	Observed Number	Observed Proportion	Cumulative Observed Proportion (O)	Expected Proportion	Cumulative Expected Proportion (E)	D= |O-E|	D= O-E
1.	Strongly Agree (SA)	245	0.7	0.7	0.2	0.2	0.5	0.5

2.	Agree (A)	84	0.24	0.94	0.2	0.4	0.54	0.54
3.	No Opinion (NO)	12	0.03428571	0.97428571	0.2	0.6	0.37428571	0.37428571
4.	Disagree (DA)	5	0.01428571	0.98857142	0.2	0.8	0.18857142	0.18857142
5.	Strongly Disagree (SDA)	4	0.01142857	1	0.2	1	0	0

Table 9.1.6: Calculation of hypothesis 6, which determines the importance of "Design" In product commercialisation

The null hypothesis reveals that there is no significant difference in the responses provided by the respondents on the factor "Design". The value given in the table at a 95 per cent confidence level is equal to 1.36 / √350 = 0.003885. Since the computed value (Dmax = 0.54) is higher compared to the table value (0.003885), the null hypothesis has been rejected. Hence, there is a significant difference among the responses provided by the artisans on 'Design'. From the above table, it is noted that since the actual differences (without the mode value) are positive for most of the cases, the responses are inclined towards 'Strongly Agree'. So, we can say that Design is a major factor for Pattachitra product commercialisation.

Sl No.	Opinion	Observed Number	Observed Proportion	Cumulative Observed Proportion (O)	Expected Proportion	Cumulative Expected Proportion (E)	D= |O-E|	D= O-E
1.	Strongly Agree (SA)	299	0.85428571	0.85428571	0.2	0.2	0.65428571	0.65428571
2.	Agree (A)	42	0.12	0.97428571	0.2	0.4	0.57428571	0.57428571
3.	No Opinion (NO)	2	0.00571428	0.97999999	0.2	0.6	0.37999999	0.37999999
4.	Disagree (DA)	5	0.01428571	0.9942857	0.2	0.8	0.1942857	0.1942857
5.	Strongly Disagree (SDA)	2	0.00571428	1	0.2	1	0	0

Table 9.1.7: Calculation of hypothesis 7, which determines the importance of "Transportation" In product commercialisation

The null hypothesis represents that there is no significant difference in the responses provided by the respondents on the factor "Transportation". The value provided in the table at a 95 per cent confidence level is equal to 1.36 / √350 = 0.003885. As the calculated value (Dmax = 0.654285710.54) is higher than the value given in the table (0.003885), the null hypothesis has been rejected. Hence, there is a significant difference among the responses provided by the artisans on 'Transportation'. From the above table, it can be said that since the actual differences (without the mode value) are positive for maximum cases, the replies are inclined towards 'Strongly Agree'. We can infer that Transportation is a crucial factor for Pattachitra product commercialisation.

10. Discussion of results of hypotheses: Fund: The K-S test result shows that a high Dmax value for fund indicates a substantial deviation from the expected uniform distribution. The majority of responses are Strongly Agree and Agree, and the null hypothesis has been rejected. This result confirms that the availability of funds is a crucial factor for Pattachitra product commercialisation. ***Marketing/Selling:*** The test result depicts a meaningful response in the Strongly Agree category. The null hypothesis is rejected, suggesting that marketing and selling of products play an important role in the commercialisation of Pattachitra products. ***Infrastructure:*** Responses are highly inclined toward agreement, which results in rejection of the null hypothesis. This shows that adequate infrastructure for production and storage is a crucial factor for product commercialisation. ***Training:*** The K-S test result portrays that training for skill development is an important factor for product commercialisation because the responses are strongly in favour. ***Raw Material:*** The result of the hypothesis shows that the null hypothesis has been rejected, and most respondents agree that access to raw materials and knowledge about them can enhance the sales of more products. ***Design:*** For the successful commercialisation of Pattachitra products, design enhancements are needed, which is reflected in hypothesis 6, where the null hypothesis has been rejected, and the responses are inclined to strongly agree. ***Transportation:*** The highest Dmax value for transportation indicates that the null hypothesis has been rejected, and knowledge and facilities related to market access and timely product distribution enhance commercialisation.

11. Major findings:

- It had been found that the Patua community had a significant number of female artisans.
- 25.71% of Patuas were middle-aged artisans who were 29-38.
- The majority of artisans were qualified under class 10. Only 2.85% artisans were under graduates.
- Very few Patuas knew the English language. A majority knew the Hindi language.
- A huge number of Patuas had information requirements for awareness of health, legal and financial. 95.71% of Patuas needed information for professional development.
- 48.28% of Patuas required information from conventional sources like the Government Office, and the majority of them preferred coworkers and family members for collecting the information.
- 73.14% of Patuas were satisfied in this profession, though various challenges are present.
- It had been identified that funds were a key factor for which Patuas were looking for another profession.
- Almost every Patuas made the product under Rs. 500 for better sales so that the common people could purchase their product.
- Fair was the highest rated place for product selling, according to Patuas.
- 44.00% Patuas had a monthly income of Rs. 5000-9999, which reflected their poor economic status.
- They received minimum financial support from the Government (45.71%) and the NGO (35.71%).
- The majority of artisans had information requirements for different awards for professional recognition, which motivated them.
- It had been identified that 98.57% Patuas used their own idea for product making.
- Plastic bags and Newspapers were mostly used materials for product packaging because of their low price.
- Due to a dearth of communication through the proper channel, many respondents were not aware of different Government schemes. For this reason, the Patuas were deprived of various facilities.
- ***Major factors of Information requirement for product Commercialisation: The K.S Test results emphasised*** that information requirements on seven major factors that are fund, marketing and selling area, infrastructure, training, design, raw material and transportation were very crucial for product development and marketing.

12. Conclusion: It was successfully revealed from the study about the true picture of the socio-cultural and socio-economic status of the Patua community at Raghurajpur. The major struggle of this community was the gap between a rich artistic heritage and low income, which created an unstable socio-economic strata. The major factors involved information poverty and a paucity of funds, although social media played a vital role in disseminating information and in market linkage for Patuas. The true sustenance of the community depended on product commercialisation, but multiple factors hindered this. It is suggested that, for the sustainable development of this community, the Patuas should be aware of market and financial literacy, as well as the exploration of information sources.

References:

Chakrabarti, B. (2006). Occupational information: needs of rural artisans. In: Information Strategies for Rural India, edited by G. Maity & S. Sett.Vidyasagar University, Midnapore, 2006, 116-132.

Chakrabarti, B., Maity, A., Chatterjee, A., Dutta, A., Saha, S., Bhattacharya, J., Mukherjee, T., Chatterjee, A. & Dutta, K. (2016) A study on the information need of weavers' community at Sualkuchi, Assam. Calcutta University Journal of Information Studies, 18:1-25.

Chinnasamy, K. (2016). Information need and information seeking behaviour of engineering college students in Madurai -a case study. Qualitative and Quantitative Methods in Libraries. 5, 131-140. Retrieved from: https://www.qqml-journal.net/index.php/qqml/article/view/311/311

Das, M. (2023). A detail study on pattachitra painting: the heritage artwork of Raghurajpur, Odisha. International Journal of Development Research, 13(10), 64015–64018. Retrieved from https://www.journalijdr.com/sites/default/files/issue-pdf/27333.pdf

Kanungo, P., Sethi, N., & Biswal, P. (2025). A policy framework for the creative pattachitra artisans at the crossroad of financial scarcity and changing trends. Creativity Studies, 13(1), 114-135. Retrieved from: https://doi.org/https://doi.org/10.3846/cs.2020.9923

Mahalik, T. (2024). Glimpase of Odisha's folk art of Raghurajpur village. International Journal of Creative Research Thought, 12(4), b152–b156. Retrieved from: https://www.ijcrt.org/papers/IJCRT2404129.pdf

Majumdar, S., & Bandyopadhyay, J. (2025). Patachitra - a cultural heritage of West Bengal and Odisha. Prakriti - The International Multidisciplinary Research Journal, 2(1), 1–7. Retrieved from: https://doi.org/10.48165/pimrj.2025.2.1.1

Mallik, S., & Bhattacharyya Sahu, N. (2021). Information needs and information seeking behaviour of a tribal community- a study on selected blocks in West Bengal. Journal of the Indian Library Association, 57(3), 1-11. Retrieved from: https://www.ilaindia.net/jila/index.php/jila/article/view/683/272

Mishra, M., & Pattanaik, S. (2024). Women entrepreneurship through pattachitra and their sustainability in rural areas with special reference to Raghurajpur. ShodhKosh: Journal of Visual and Performing Arts, 5(4). Retrieved from: https://doi.org/10.29121/shodhkosh.v5.i4.2024.1863

Raising, S., Gaddimani, P., & K.S. Yadav, A. (2021). Information Need and Information-Seeking Behaviour of Tangkhul Naga Community in India. Library Philosophy and Practice. Retrieved from: https://www.researchgate.net/publication/350580177_Information_Need_and_Information_Seeking_Behaviour_of_Tangkhul_Naga_Community_in_India

Rath, A., Padhy, H. M., Mishra, P., & Mishra, B. K. (2023). Digitalization of Motifs Based on Indian Pattachitra painting through CAD andtheir Adaptation on Apparels using Digital Printing Technique. International Journal of Humanities, Social Science and Management, 3(4), 10–13. Retrieved from: https://ijhssm.org/issue_dcp/Digitalization%20of%20Motifs%20Based%20on%20Indian%20Pattachitra%20painting%20through%20CAD%20andtheir%20Adaptation%20on%20Apparels%20using%20Digital%20Printing%20Technique.pdf

Richa, & Jha, V. (2025). Presenting stories through the art of pattachitra. JournalNX- A Multidisciplinary Peer-Reviewed Journal, 11(8),1-7. Retrieved from: https://repo.journalnx.com/index.php/nx/article/view/5596/5313

Verdecias, N., Garg, R., Steensma, J., McQueen, A., Gree, R., & Kreuter, M. W. (2020). Expressed and unexpressed social needs in low-income adults in the U.S. Health Soc Care Community, 29(5), 184–191. Retrieved from: https://doi.org/10.1111/hsc.13265

About Authors

Mrs. Debalina Mukherjee (BLIS, MLIS, M. Phil, NET) is a Ph.D. Research Scholar of DLIS, University of Calcutta.Presently she is working as a Librarian in Loreto College, Kolkata. She has more than 8 years' experience in different libraries in Kolkata. She worked as a Librarian in Apollo Gleneagles Nursing College and as an Assistant Librarian in Scottish Church College. She was a library trainee in Asiatic Society, Kolkata and has also teaching experience in Maharani Kasiswari College as well as Loreto College in the field of LIS.

Prof. Arabinda Maity, B.Sc. (Chemistry), B.Ed., BLIS, M.Lib.Sc., M. Phil. in LIS, Ph.D. in LIS, Professor of DLIS, University of Calcutta has been teaching in the Department of Library & Information Science, University of Calcutta since 2009. 30 M.Phil scholars, 9 Ph.D. Scholars were awarded under his guidance. He has a number of publications to his credit which includes 6 books and more than 70 articles in National and International journals.

Prof. Biplab Chakrabarti, M.A. in Sociology, LL. B, M.Lib.Sc.(Gold Medalist), Ph.D. in LIS, UGC NET, Former Professor of DLIS, University of Calcutta has 39years of professional and teaching experience in LIS field. He has special research interest in Community Information Service Bibliometrics, LIS Education, Library Management, Collection Development.44 Ph.D. Scholars have already been awarded under his guidance. He holds numerous academic and administrative positions in the university and other renowned organisations. He published vast amount of research works in different journals and wrote various books in this field.

Blind Peer Reviewed
(Previously UGC-CARE)
ISSN: 0972-2750

An Altmetric Analysis of Highly Cited Publications on Robotic Surgery

Sandip Majumdar, Kush Sibbu, Ramesh Kuri, & Sanmati Jinendran Jain

Manuscript Received on –
27th December, 2025
Primary Reviewed on –
4th January, 2026
Peer Reviewed on –
18th January, 2026
Revised on –
28th January, 2026
Accepted for publication
5th February, 2026

Mr. Sandip Majumdar
Assistant Professor
Dept. of Lib. & Inf. Sc.,
University of Gour Banga
times_sandip@yahoo.com

Dr. Kush Sibbu
Assistant Librarian
Indian Institute of Technology Goa
kush@iitgoa.ac.in

Dr. Ramesh Kuri
Assistant Professor
Dept. of Lib. & Inf. Sc.,
Rani Channamma University,
rameshkuri.rcu@gmail.com

Mr. Sanmati Jitendra Jain
Assistant Librarian
Central Library, National Institute of Technology Tiruchirappalli
sanmati.nitt@gmail.com

ABSTRACT

Introduction: *Altmetrics are metrics that go beyond traditional citation counts to reflect the frequency and scope of online sharing and discussion of scholarly publications. In other words, it may be viewed as the measure of the projection of influence of a scholarly publication through various platforms.*

Objective: *This study aims to understand the online projection of influence of highly cited publications upon the peer group and beyondby selecting top-cited papers on robotic surgery and studying them through the lens of Almetrics.*

Methodology: *Using citation counts as the cutoff measure, the study examined the top 100 articles on robotic surgery that received many citations in the Web of Science (WoS). The Altmetric Attention Score (AAS) was extracted for every article to understand the extent ofits online engagement. A keyword string associated with robotic surgery from WoS was used to conduct the research. The journals were assessed using the Scimago Journal Rankings' Quartile category and the 2022 Journal Impact Factor provided by the Journal Citation Report (JCR) by Clarivate Analytics. By tracking mentions across platforms such as news outlets, blogs, social media, and policy documents, Altmetric provides a nuanced understanding of research influence.*

Findings: *The findings reveal significant divergence between traditional citation counts and altmetric scores, highlighting the complementary nature of these metrics. Twitter emerged as the predominant platform for discussing robotic surgery research, emphasising the role of social media in science communication. The study underscores the critical role of social media platforms in the dissemination and communication of scientific research, revealing their influence on broadening the reach and impact of academic work*

.

Keywords: *Altmetrics; Altmetric Attention Score; Robotic Surgery; Social Attention; Citations; Research Impact*

Introduction

Robotic surgery has proliferated across specialties, with bibliometric studies mapping its global research landscape, productivity, and thematic evolution in spinal surgery, urogynecology, and orthopaedics; for example, analyses of robot-assisted spinal surgery trends (1997–2025) reveal core research hotspots and highlight increasing integration of precision technologies, navigation systems, and artificial intelligence to improve clinical outcomes and feasibility in complex procedures (Huang et al., 2025). Similarly, comparative bibliometric evaluations in uro-gynaecology and obstetrics &gynaecology illustrate how robotic and laparoscopic approaches are studied at scale, informing debates on efficacy, complication profiles, and methodological reporting standards that influence citation impact (Bilir et al., 2025). The recent comprehensive mapping of global orthopaedic robotic surgery further confirms expanding interdisciplinary contributions and collaborative networks across high-income countries (Yang et al., 2026). In parallel, altmetrics have matured as complementary impact indicators that capture real-time online visibility and engagement beyond traditional citations, encompassing social media, news outlets, blogs, and reference managers, thereby reflecting broader dissemination trajectories of scientific literature ('Altmetrics', 2026). Empirical investigations in clinical medicine and surgery report moderate positive correlations between Altmetric Attention Scores and citation counts, reinforcing the notion that online engagement may serve as a predictive signal for future academic influence, albeit with disciplinary variability (Koh et al., 2025). Recent systematic reviews advocate hybrid evaluation frameworks, arguing that altmetrics enrich bibliometric assessment by revealing social impact and diverse forms of scholarly attention that traditional metrics alone cannot capture (González et al., 2025). Together, these developments justify applying altmetric analysis to highly cited robotic surgery research to uncover patterns of digital engagement, interdisciplinary reach, and potential linkages between online attention and sustained academic influence, thereby situating the study within current research evaluation paradigms that integrate both citation and altmetric perspectives. This paper attempts to perform an altmetric analysis of highly cited publications on robotic surgery to evaluate their wider impact. By utilising altmetric data, the study aims to understand the various platforms through which these works are shared and the diverse audiences they engage, along with the flow of scholarly information. This method is anticipated to shed light on the broad-reaching effects of important research in robotic surgery and emphasise the importance of incorporating altmetric insights into research evaluation and strategic planning in the medical field, particularly robotic surgery. Integrating altmetrics into assessing robotic surgery research may provide a more holistic view of its influence, offering insights that traditional citation metrics may overlook. Understanding the full extent of how influential publications in robotic surgery resonate within and beyond the academic community can help stakeholders better recognise their importance and guide future research efforts.

Review of Related Literature

Robotic surgery has become a ground-breaking technology in minimally invasive procedures, delivering greater precision, shorter recovery times, and better patient outcomes than conventional surgical techniques. Altmetrics, a term coined by Priem et al. (2011), may offer an alternative approach to assessing how a research publication on Surgical Robotics is being accessed, read, discussed, recommended, and cited by measuring research impact through tracking attention across various online platforms, including social media, news outlets, and policy documents. Koh et al. (2025) examined the relationship between Altmetric Attention Scores and traditional citation counts across high-impact clinical journals, finding a moderate positive correlation that suggests online engagement may be a supplementary predictor of scholarly influence, though this varies by speciality, with weaker associations in surgical fields. Shu (2025) investigated whether Altmetric scores could serve as an alternative to bibliometrics in pediatric robotic surgery, concluding that altmetrics did not replace citation metrics and that combined use provides a more comprehensive evaluation. Chaudhary et al. (2024) analysed article characteristics influencing citations and altmetric scores in general surgery literature, revealing that altmetric performance was significantly associated with author country, publication topic, and journal choice, indicating context-specific predictors of impact. Song et al. (2024) conducted a bibliometric and visualisation study on global trends in robotic surgery literature, highlighting thematic evolution and publication dynamics that set the foundation for subsequent impact analyses. Bilir et al. (2025) provided a systematic review and bibliometric analysis contrasting robot-assisted and laparoscopic surgery in urogynecology, illustrating discipline-specific research priorities that may influence both citations and altmetric attention. Yilmaz andPolat (2023) performed a combined bibliometric and altmetric analysis of highly cited vitreoretinal surgery articles, noting weak correlations between altmetric scores and citation counts, thereby foregrounding the nuanced relationship of altmetrics with traditional measures. Abdelwahab et al. (2025) presented a bibliometric analysis of the top 100 most-cited articles in the Journal of Robotic Surgery, identifying key authors, institutions, and thematic clusters, which provides a basis for selecting highly cited sets for altmetric investigation. Recent comprehensive mapping of orthopaedic robotic surgery research by global bibliometric methods underscores collaborative patterns and emerging technical clusters, suggesting the disciplinary contours within which altmetric attention may vary. Luo et al. (2025) offered a systematic review of clinical accuracy in robotic surgery procedures, emphasising evidence translation into practice, an aspect that may drive both scholarly citations and public interest. Broader meta-analyses of altmetrics literature indicate that altmetric indicators capture dimensions of immediate dissemination not reflected by citations, reinforcing the rationale for incorporating altmetric analysis into research on highly cited robotic surgery publications (‘Altmetrics’, 2026). This underscores the importance of considering altmetric data in research evaluation and strategic planning. Despite the growing interest in altmetrics, there is a lack of research specifically focusing on the altmetric analysis of robotic surgery publications. This paper seeks to address this gap by performing an in-depth altmetric analysis of highly cited publications on robotic surgery. By examining the wider impact of these works, the study aims to understand the different channels through which robotic surgery research is shared and engaged across multiple platforms.

Objectives:

1) To analyse the online attention and dissemination of highly cited publications on robotic surgery.
2) To identify the characteristics, publication distribution, and topics of the highly cited and discussed robotic surgery publications based on altmetric data.
3) To examine the correlation between altmetric attention score and traditional citation counts.

Materials and Method

Data source and selection criteria

The data for this study were obtained from the Web of Science (WoS) database, which was used to identify and analyse the top 100 highly cited publications in robotic surgery. Articles primarily focused on robotic surgery were selected. The study focused on the top 100 publications based on citation count, which served as the cut-off criterion. The data collection was conducted on April 15, 2025.

Search strategy

A comprehensive search was conducted using keywords related to robotic surgery across all fields and collections in the WoS database. To find pertinent articles, the following search string was used: ((“Robotic” OR “robot”) AND (“Surgery” OR “Surgical Procedures”)). The search results were sorted by citation count, and the top 100 articles were selected for further analysis.

Altmetric data collection

To assess the selected publications' broader online engagement and impact, data for every 100 articles were obtained using altmetric.com. The Altmetric Attention Score (AAS) was collected for every article, capturing mentions across various online platforms. These platforms include social media, news outlets, blogs, and policy documents, all of which are tracked by the AAS to reflect the publication's online presence and influence.

Journal evaluation

Several key metrics were analysed to assess the impact and reputation of the journals publishing the selected articles on robotic surgery. The average number of citations received by articles in each journal over the previous year was reviewed to measure short-term impact. The five-year impact factor was considered to capture the journal's broader influence within the field. Quartile rankings, as reported by SCImago Journal Rankings (SJR), were also used to determine each journal’s relative standing within its subject area. These metrics collectively provided a detailed perspective into the influence and scholarly reputation of the publication venues associated with highly cited robotic surgery research.

Data Analysis and Interpretation

A comparative analysis was performed to identify patterns and differences between traditional citation counts and Altmetric Attention Scores (AAS). This analysis explored the relationship between the two metrics, focusing on how platforms such as Twitter and blogs contribute to the dissemination of robotic surgery research and its broader impact.

Group A			
Sl. No.	**DOI**	**Total Citations**	**AA Score**
1	10.1146/annurev-bioeng-010510-103409	1541	52
2	10.1038/s41591-018-0316-z	1538	425
3	10.1038/nature25443	1529	878
4	10.1056/NEJMoa1806395	1284	1027
5	10.1126/science.aap8757	1102	1288

Group B			
Sl. No.	**DOI**	**Total Citations**	**AA Score**
6	10.1038/nature09744	1098	51
7	10.1126/science.1197410	1064	33
8	10.1016/j.eururo.2020.03.055	1023	53
9	10.1016/j.cell.2013.03.021	1022	69
10	10.1126/scirobotics.aam6431	1004	278
11	10.1109/TRO.2015.2489500	933	23
12	10.1016/j.eururo.2012.05.045	923	17
13	10.1097/01.sla.0000103020.19595.7d	900	73
14	10.1016/j.metabol.2017.01.011	893	36
15	10.1109/TRO.2010.2073030	857	22
16	10.1001/jama.2017.7219	801	325
17	10.1038/s41578-018-0022-y	791	44
18	10.1016/j.eururo.2009.01.036	776	3
19	10.1016/S0140-6736(15)00046-X	734	38
20	10.1126/scirobotics.aax7329	711	900
21	10.1056/NEJMp1006602	708	146
22	10.1109/TBME.2004.831542	706	9
23	10.1126/scirobotics.aar7650	701	422

24	10.1109/JPROC.2014.2385105	641	70
25	10.1109/TRO.2009.2035740	626	9

Group C			
Sl. No.	**DOI**	**Total Citations**	**AA Score**
26	10.1016/S1470-2045(17)30068-2	591	75
27	10.1016/j.eururo.2012.05.046	566	12
28	10.1109/TRO.2008.2002311	551	6
29	10.1126/science.aau0780	533	364
30	10.1097/01.mlg.0000227184.90514.1a	517	3
31	10.1016/S0140-6736(18)30996-6	515	562
32	10.1016/S1470-2045(08)70074-3	511	4
33	10.1002/14651858.CD003677.pub5	503	414
34	10.1038/sdata.2014.53	498	6
35	10.1097/MOU.0b013e32831a478c	474	31
36	10.1021/ac0352171	469	12
37	10.1109/JSEN.2008.917481	465	6
38	10.1016/S0140-6736(16)30592-X	458	913
39	10.1016/j.annonc.2020.06.011	454	104
40	10.1016/j.eururo.2014.11.043	443	59
41	10.1016/j.jmbbm.2011.04.005	441	1
42	10.1016/j.eururo.2009.10.024	428	10
43	10.1001/jama.2013.186	419	267
44	10.1016/j.arth.2017.07.021	415	33
45	10.1016/j.juro.2009.05.037	412	11
46	10.1016/j.eururo.2012.02.029	410	16
47	10.1007/s00464-008-0298-x	408	3
48	10.1227/NEU.0b013e31827d1161	407	2
49	10.1016/j.amjsung.2004.08.025	401	—
50	10.1016/j.athoracsur.2004.09.040	396	1
51	10.1089/soro.2014.0001	394	6
52	10.1016/j.eururo.2008.07.009	394	—

53	10.1016/j.neuroimage.2010.10.058	383	3
54	10.1016/j.eururo.2012.05.047	381	8
55	10.1007/s00464-018-6079-2	380	48
56	10.1016/j.eururo.2012.05.044	379	6
57	10.1108/01439910410566362	379	—
58	10.1016/j.eururo.2014.10.026	378	311
59	10.1111/j.1464-410X.2011.10827.x	377	16
60	10.1111/j.1464-410X.2007.07359.x	377	3
61	10.1109/TMI.2014.2321777	376	9
62	10.1097/SLA.0000000000003031	367	33
63	10.1016/j.urology.2007.11.034	359	6
64	10.1245/s10434-009-0435-3	355	9
65	10.1177/0278364908104278	353	4
66	10.1177/0018720816644364	349	51
67	10.1002/lary.23294	348	4
68	10.1038/ncomms15894	348	122
69	10.1056/NEJMcp0910041	347	3
70	10.1146/annurev-bioeng-070909-105249	345	3
71	10.1016/j.surg.2009.09.007	340	6
72	10.1097/SLA.0000000000002979	339	44
73	10.1016/j.urology.2004.06.049	339	—
74	10.1016/j.eururo.2014.12.007	338	19
75	10.1016/j.clinph.2010.01.001	336	19
76	10.1126/scitranslmed.aad9398	333	1825
77	10.1001/jamanetworkopen. 2019.18911	332	325
78	10.1016/j.eururo.2011.12.027	325	27
79	10.1016/j.eururo.2013.07.039	324	32
80	10.1016/j.eururo.2015.02.029	323	31
81	10.1002/cncr.23027	323	—
82	10.1016/j.eururo.2016.08.060	323	28
83	10.1016/j.eururo.2021.06.007	322	84
84	10.1109/ToH.2010.19	322	9

85	10.1021/acs.chemrev.0c01234	321	35
86	10.1016/j.juro.2012.09.042	317	5
87	10.1109/TRO.2013.2256313	317	9
88	10.1016/j.juro.2011.09.032	316	20
89	10.1007/s00464-010-1281-x	315	1
90	10.1097/SLA.0b013e3181e6239e	315	41
91	10.1016/j.eururo.2006.06.017	312	6
92	10.1097/AOG.0b013e318231537c	311	12
93	10.1126/scirobotics.aar2629	311	179
94	10.1016/j.eururo.2008.06.063	309	52
95	10.1007/s00464-009-0366-x	307	6
96	10.1007/s00464-009-0825-4	307	–
97	10.1016/j.ygyno.2008.08.022	303	6
98	10.1177/000348940711600104	292	–
99	10.1016/j.eururo.2014.12.028	289	16
100	10.1016/S1470-2045(18)30357-7	288	104

Table-1Comparison of Highly Cited Publications across Groups A, B, and C

Table 1 is based on the citation and AA Score distributions across the three groups, categorised by citation count into Group A [1101-1600], Group B [601-1100], and Group C [101-600]. The ratio of articles with corresponding citations belonging to different groups stands as 1:4:15. Publications in Group A exhibit a wide range of citations and AA Scores, indicating that some highly cited articles also achieved close public attention. In contrast, others with similar citations show moderate attention. Group B has slightly lower citation counts overall, with its AA Scores showing less variability than Group A, suggesting a more consistent, though generally lower, level of online attention. Group C includes articles with lower citation counts and AA Scores overall, reflecting perhaps more specialised or emerging research areas with less immediate impact.

Only five articles have AA Scores exceeding their citation counts, indicating that for most articles, scholarly impact (as measured by citations) surpasses online attention. This suggests that high social presence does not necessarily align with research influence, as most articles do not achieve AA Scores that match their citation counts. Of the five articles with higher AA Scores than citations, three are in Group C, one in Group A, and one in Group B. This distribution suggests that lower-citation groups, particularly Group C, may sometimes achieve closer online attention despite limited research impact. The article with the highest AA Score is in Group C, which has the lowest citation range, underscoring that online attention can be independent of scholarly citation trends. Most of the articles were published before 2012, and they may have accumulated their citations over a longer period. Still, their AA Scores are limited by the fact that altmetrics were introduced later.

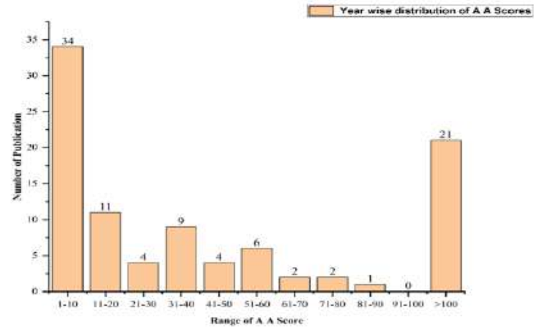


Figure 1: Distribution of Altmetric Attention Scores

Fig. 1 demonstrates the distribution of Altmetric Attention Scores for the highly cited publications on robotic surgery. Most publications fall within the lowest score range (1-10), suggesting limited online engagement. Conversely, many publications have very high AAS (>100), highlighting their substantial impact and attention on online platforms. This pattern underscores the varied nature of online attention that different publications receive, emphasising the importance of considering both highly and less engaged works in altmetric analyses. This actually provides us with the insight that our first objective was intended to achieve.

Journal Analysis

Sl.No	Name of the Journal	Name of the Publisher	Number of mentioned outputs	Total mentions	2023 Journal Impact Factor (JCR)
1	European Urology (EU)	Elsevier	19	524	25.3
2	IEEE Transactions on Robotics (IEEE TR)	Institute of Electrical and Electronics Engineers (IEEE)	5	92	9.4
3	Annals of Surgery (AS)	Lippincott Williams & Wilkins (LWW)	4	209	7.9
4	Science Robotics (SR)	American Association for the Advancement of Science (AAAS)	4	1388	26.1
5	Surgical Endoscopy (SE)	Springer Nature	4	72	2.4
6	New England Journal of Medicine (NEJM)	Massachusetts Medical Society	3	836	96.2
7	The Lancet (TL)	Elsevier	3	1056	98.4
8	Lancet Oncology (LO)	Elsevier	3	107	41.6
9	Science (SCI)	American Association for the Advancement of Science (AAAS)	3	1239	44.7
10	British Journal of Urology (BJU)	Wiley	2	31	9.1

Table 2: Journal Analysis

Table 2 presents the analysis of the top 10 Q1 journals that received significant attention in robotic surgery. Science Robotics has the highest total mentions (1388) despite having only 4 mentioned outputs, indicating a very high level of engagement per publication. It also has a high impact factor of 26.1, placing it in the Q1 quartile. European Urology has 524 total mentions with 19 mentioned outputs. Its impact factor is 25.3 and is also in the Q1 quartile, suggesting it is a leading journal. Annals of Surgery has 209 total mentions across 4 outputs, with an impact factor of 7.9 and a Q1 ranking, indicating strong performance across traditional and altmetric metrics. IEEE Transactions on Robotics has 92 total mentions from 5 outputs, an impact factor of 9.4, and is ranked in Q1, indicating good performance with moderate altmetric engagement. Surgical Endoscopy has the lowest total mentions (72) from 4 outputs, with an impact factor of 2.4. Although it is in the Q1 quartile, its altmetric engagement is relatively low compared to other journals. The table highlights the relationship between traditional impact metrics (impact factor and quartile ranking) and altmetric mentions for various journals on robotic surgery. Journals like Science Robotics and European Urology demonstrate high engagement and impact traditionally and through altmetrics. In contrast, journals like Surgical Endoscopy show high traditional impact but lower altmetric engagement. This indicates that while traditional metrics are important, altmetrics provide an additional dimension of understanding a journal's influence and reach in the broader community. The observation that journals with single-digit Impact Factors often exhibit lower 'total mentions' aligns with existing research. However, it's important to recognise that altmetrics capture a different dimension of scholarly impact compared to traditional citation-based metrics. While Impact Factors reflect the average number of citations to articles published in a journal, altmetrics measure the immediate attention and engagement an article receives across various online platforms.

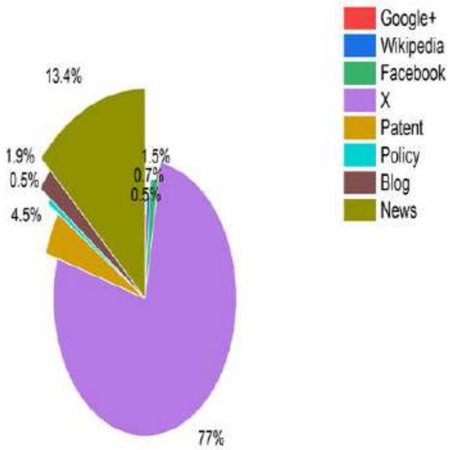


Figure 2: Sources of Online Attention

Twitter (X) is the most significant source of online attention, with 7148 mentions. This suggests that discussions and shares on Twitter are crucial in disseminating information about robotic surgery publications. The second-highest source of online attention is news mentions, totalling 1242. This indicates substantial coverage by news outlets, highlighting the public and media interest in robotic surgery. Facebook mentions (140), Wikipedia mentions (63), Policy mentions (51), and Google+ mentions (47) contribute relatively fewer mentions. While not as prominent, these sources still play a role in overall online engagement. Mendeley readership counts reflect the number of times a research publication has been added to the libraries of Mendeley users. These counts provide a direct measure of researchers' interest in specific publications, as shown in Fig. 2.

Range of Mendeley Readership Counts	No. of Publications	Percentage
1-250	30	31.91
251-500	34	36.17
501-750	16	17.02
751-1000	6	6.38
> 1000	8	8.51

Table 3: Mendeley Readership Distribution

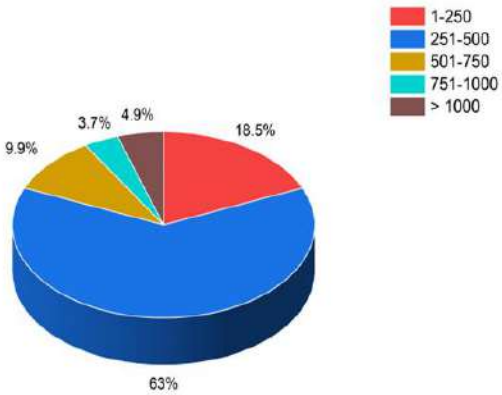
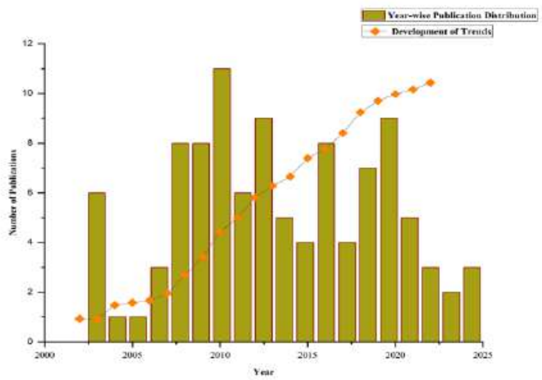


Figure 3: Mendeley Readership Distribution

Mendeley Readership Counts help to identify influential papers shortly after publication, even before citations accumulate. Table 3 shows that most publications on robotic surgery have moderate Mendeley readership counts, with 68.08% of publications falling within the 1-500 range. A smaller percentage of publications (31.92%) have higher readership counts, indicating that while most papers receive moderate attention, a few garner substantial engagement. This distribution highlights the varied levels of interest and use of robotic surgery publications within the Mendeley community, as shown in Fig. 3.



The year-wise distribution of publications on robotic surgery shows a trend of initial growth, fluctuating publication rates in the mid-period, and a sharp decline in recent years, with a slight rebound in 2021, as presented in Fig.4. The peaks in 2010 and 2017 suggest periods of heightened research activity or interest in the field. The overall pattern indicates variability in publication output, which could be influenced by technological advancements, funding availability, and evolving research priorities.

Correlation Matrix

		Citations	Altmetric Attention Score
Citations	Pearson's r	—	
	df	—	
	p-value	—	
Altmetric Attention Score	Pearson's r	0.331 ***	—
	df	98	—
	p-value	< .001	—

Note. * p < .05, ** p < .01, *** p < .001

Table 4: Correlation between Citation Count and AA Score

Table 4 presents the correlation analysis between citations and AA Score, indicating a positive but weak relationship (r=0.331). It also indicates (p=0.001) a statistically significant relationship between the citation count and Altmetric Attention Score. This suggests that articles receiving greater online attention (as measured by the Altmetric Score) are more likely to have higher citation counts. However, the correlation is weak, suggesting that other factors may also influence citation counts. Thus, this addresses our third objective, i.e., correlation between altmetric attention score and traditional citation count.

Discussion

The present study provides a comprehensive altmetric analysis of highly cited publications on robotic surgery, revealing significant results in the dissemination and impact of research beyond traditional citation metrics. The analysis indicates a substantial divergence between traditional citation counts and altmetric attention scores. "Microrobots for Minimally Invasive Medicine" boasts the highest number of citations (1541), but its Altmetric Attention Score (AA Score) is relatively low (52), suggesting limited online engagement. Conversely, "Analysis of shared heritability in common disorders" exhibits both high citation counts (1102) and an exceptionally high AA Score (1288), underscoring its broad appeal and significant online presence. Articles may not have had the same opportunity to gather AA Scores at the time of their initial impact or relevance, which may affect their comparative altmetric attention today.

This disparity highlights the complementary nature of altmetrics and traditional citation metrics. Citations predominantly reflect academic influence, while altmetrics capture broader societal impact, including public and professional engagement (Bornmann, 2014). Citations predominantly reflect academic influence, while altmetrics also capture broader public engagement, social impact, and the immediate online attention an article receives through social media, news outlets, policy documents, and blogs. Studies have shown that altmetric indicators can provide early signals of a publication's impact and are often predictive of future citations (Eysenbach,2011).

The sources of online attention varied significantly, with Twitter (X) emerging as the dominant platform, accounting for 7148 mentions. Twitter has been distinct from other social media platforms for several reasons, particularly its influence on academic and scientific discourse. Twitter's character limit encourages concise communication, enabling researchers to distil complex ideas into key points. This feature allows information to spread quickly. Twitter's real-time interactions facilitate viral spread, with a single post reaching thousands within minutes, often leading to a "burst" of attention. Twitter's format aligns well with cognitive and social behaviours, such as a preference for short, digestible information and immediate social feedback (likes, reposts, replies). This finding aligns with previous research emphasising the role of social media in disseminating scientific knowledge (Haustein et al., 2014). News outlets also played a critical role, with 1242 mentions indicating robust media interest in robotic surgery advancements. The substantial engagement on Twitter suggests that researchers and practitioners in robotic surgery actively use this platform to share findings and engage with a broader audience. This trend underscores the importance of social media in contemporary science communication, as it facilitates rapid dissemination and fosters interactive discussions (Priem et al., 2010).

The year-wise distribution of publications revealed fluctuating research activity, with notable peaks in 2010 and 2017. These peaks may correspond to significant technological advancements or increased funding in robotic surgery during those periods. However, the recent decline in publications from 2018 to 2020 raises concerns about potential challenges in the field, such as funding constraints or shifts in research priorities. The distribution of Mendeley readership counts shows that most publications (68.08%) have a moderate readership (1-500). Only a small fraction (8.51%) achieved readership counts exceeding 1000, indicating exceptional engagement. These findings reflect similar patterns observed in other medical fields, where a few high-impact papers garner disproportionate attention (Costas, Zahedi & Wouters, 2015).

The observed correlation between journals with single-digit Journal Impact Factors and their low altmetric mentions may be partly explained by the influence of human psychology shaped by the dominance of traditional metrics. Academic research institutions have adopted citation-based metrics such as the Journal Impact Factor as the definitive measures of journal and article quality. This relentless emphasis has cultivated a widespread perception among researchers that these metrics represent the standard/benchmark for assessing scholarly impact. This phenomenon reflects the substantial influence of research analytics tools in shaping academic priorities and may contribute to a narrow focus on traditional metrics over alternative measures of research impact.

There is one interesting fact that is evident from Table 2: the journals with a single-digit Journal Impact Factor exhibit poor total mentions. Human psychology could have played a major role in this exhibition. How? Commercial publishing behemoths, along with information intermediaries such as Clarivate Analytics, have relentlessly promoted the supremacy of citation counts and the Journal Impact Factor over other measures (Macdonald, 2025). They have created a large base of believers(Daugherty et al., 2022) of metrics of WoS and Cite Score as gold standards for measuring the quality and influence of an article published in journals meeting their self-defined criteria. Naturally, this large base of researchers and their affiliated institutions may, in turn, psychologically influence new entrants. Eventually, this vicious cycle created an environment in which researchers may, at first, be inclined to judge the merit of an article by the journal that hosts it and by the metrics heavily advertised by intermediaries (Yüksek, 2020). This is a fine example of how a multi-million-dollar publishing industry projects its power and influence, and could be viewed as one of the reasons for the peculiar tunnel vision behaviour observed in Table 2.

Conclusion

This study presents a thorough altmetric analysis of highly cited publications on robotic surgery, revealing critical insights into their broader impact beyond traditional citation metrics. Our findings highlight the complementary role of altmetrics in evaluating the influence and reach of scientific research, particularly in a rapidly evolving field like robotic surgery. The analysis demonstrated a noticeable divergence between traditional citation counts and altmetric scores. The year-wise distribution of publications indicated fluctuating research activity, with notable peaks and recent declines. This pattern underscores the need for sustained investment and support to maintain research momentum in robotic surgery. Most publications had moderate Mendeley readership counts, with a small fraction achieving exceptionally high engagement. Researchers may strategically utilise social media to enhance the visibility and impact of their work. Institutions and funding bodies may consider integrating altmetric data into their research evaluation frameworks. Addressing the recent decline in publication activity requires ongoing support and investment in robotic surgery research. Encouraging interdisciplinary collaborations and securing funding for innovative projects are crucial for sustaining progress in this field. Additionally, the application of Artificial Intelligence (AI) may open new vistas with many promises. The intriguing fact that surfaced in this study is that a few exceptional papers have accumulated substantial citations, but their AA Scores remained low. Future research can be explored on emerging trends and the evolving societal impact of robotic surgery using policy mentions and citation data. Comparative studies across specialities may further reveal discipline-specific engagement patterns.

Conflict Of Interest

The authors declare that there is no conflict of interest among them from the conceptualisation of the idea through to the conduct of the research.

References

Abdelwahab, S. I., Taha, M. M. E., Farasani, A., Moshi, J. M., Assiri, A., Alshahrani, S., Sahli, K. A., Shubaily, H. M., Qadri, M., Khardali, A., & Hassan, W. (2025). A bibliometric analysis of the top 100 most-cited articles published in the Journal of Robotic Surgery. *Journal of Robotic Surgery, 20*(1).

https://doi.org/10.1007/s11701-025-03056-z

Wikipedia contributors. (2026, January 5). *Altmetrics*. Wikipedia. https://en.wikipedia.org/w/index.php?title=Altmetrics&oldid=1331294565

Bilir, E., Derin, X., VetaDarkovski, J., Kaur, M. M., Ackermann, J., Maass, N., Allahqoli, L., Cartwright, R., &Alkatout, I. (2025). A systematic review and bibliometric analysis of robot vs laparoscopic surgery in urogynecology: Current trends and future directions. *Journal of Robotic Surgery, 19*(1), 748. https://doi.org/10.1007/s11701-025-02885-2

Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics, 8*(4), 895–903. doi:10.1016/j.joi.2014.09.005

Chaudhary, D., Acharya, S., Aggarwal, V., Huzaifa, M., Kain, P., Garg, R., Harlalka, K., Kumar, S., &Vasudev, A. (2024). Predictors of citations and altmetric scores in general surgery literature. *Turkish Journal of Surgery*, *40*(2), 145–153. https://doi.org/10.47717/turkjsurg.2024.6201

Costas, R., Zahedi, Z., & Wouters, P. (2015). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology, 66*(10), 2003–2019. doi:10.1002/asi.23309

Daugherty, A., Hegele, R. A., Lu, H. S., Mackman, N., Rader, D. J., & Weber, C. (2022). Web of science's citation median metrics overcome the major constraints of the journal impact factor. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *42*(4), 367–371. doi:10.1161/ATVBAHA.122.317426

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research, 13*(4), e123. doi:10.2196/jmir.2012

González, P., Fors, M., & Torres, A. (2025). Altmetrics in the evaluation of scholarly impact: A systematic and critical literature review. *Frontiers in Research Metrics and Analytics, 10*. https://doi.org/10.3389/frma.2025.1693304

Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology, 65*(4), 656–669. doi:10.1002/asi.23101

Huang, H., Lian, Z., Liu, Y., & Sun, K. (2025). Global research trends in robot-assisted spinal surgery: A visualized bibliometric analysis. *Journal of Robotic Surgery,* ***20***, 4. https://doi.org/10.1007/s11701-025-02954-6

Koh, A., Lewis-Lloyd, C. A., Wong, T., & Lobo, D. N. (2025). Correlation between Altmetric Attention Scores and citation scores across the high-impact-factor journals in Medicine, Surgery, and Anaesthesia. British Journal of Anaesthesia, 134(3), 703–712. https://doi.org/10.1016/j.bja.2024.09.034

Luo, Z., Unkovskiy, A., & Li, J. (2025). Accuracy of robotic computer-assisted implant surgery in clinical studies: A systematic review and meta-analysis. *BMC Oral Health, 25.* https://doi.org/10.1186/s12903-025-05837-2

14.Macdonald, S. (2025). The scandal of academic publishing. *Publishing Research Quarterly*, *41*(3–4), 328–356. doi:10.1007/s12109-025-10042-8

Priem, J., Taraborelli, D., Groth, P., &Neylon, C. (2011, September 1). *Altmetrics: A Manifesto.* https://altmetrics.org/manifesto/

Shu, B. (2025). Could the Altmetric Score be an alternative approach to traditional bibliometrics for evaluating influential articles in pediatric Robotic Surgery? *Medicine Advances, 3*(3), 171-176. https://doi.org/10.1002/med4.70026

Song, M., Liu, Q., Guo, H., Wang, Z., & Zhang, H. (2024). Global trends and hotspots in robotic surgery over the past decade: A bibliometric and visualised analysis. *Journal of Robotic Surgery*, *19*(1), 33.

https://doi.org/10.1007/s11701-024-02203-2

Yang, Z., Chang, X., Fu, G., Wu, X., Fan, J., & Zhou, C. (2026). Global research hotspots and emerging trends in orthopaedic robotic surgery: A comprehensive bibliometric analysis. *Journal of Robotic Surgery, 20*(1), 179. https://doi.org/10.1007/s11701-025-02998-8

Yilmaz, M., &Polat, O. (2023). Altmetric and bibliographic analysis of the most popular articles on vitreoretinal surgery between 2010 and 2020. *Cureus,15*(3), e36465. https://doi.org/10.7759/cureus.36465.

Yüksek, A. (2020). Should the Quality of the Journal Determine the Quality of the Research? In *Judging Research*. https://doi.org/10.3390/books978-3-03928-315-6-7

About Authors

Mr. Sandip Majumdar has been serving as an Assistant Professor in the Department of Library and Information Science at the University of Gour Banga, Malda, West Bengal, since 2016. Before this assignment, he served as a librarian at Sammilani Mahavidyalaya, Kolkata, from 2014 to 2016; worked at the IGNOU Library and Documentation Division, New Delhi, as a Semi-Professional Assistant from 2010 to 2014; and was a trainee at the Birla Industrial & Technological Museum Library, Kolkata. He has published 27 research papers in both national and international peer-reviewed journals, as well as book chapters and conference papers. He received a gold medal for securing the first rank in M.Lib.I.Sc (Master of Library and Information Science) examination at Rabindra Bharati University. His areas of interest include knowledge organisation, scholarly communication, scientometrics, and altmetrics.

Dr. Kush Sibbu is currently serving as Assistant Librarian at the Indian Institute of Technology Goa, where he has been contributing to academic and research support services with a strong focus on digital resources and scholarly communication. He is a versatile Library and Information Science professional with over 15 years of progressive experience across diverse sectors, including infrastructure and project management, information technology, legal institutions, and academia. His professional journey includes significant roles with reputed organisations such as the British Library, Adani Group, Tata Consultancy Services, and the High Court of Gujarat, providing him with broad institutional exposure and a multidisciplinary perspective.

Dr. Ramesh B Kuri is currently working as Assistant Professor in the Department of Library and Information Science at Rani Channamma University, Belagavi. He is passionate about research and has published 86 research papers in national and international peer-reviewed journals, conference proceedings, and edited books. He has been awarded one short-term project funded by ICSSR, New Delhi, and three minor research projects funded by PMEB, Rani Channamma University, Belagavi.His research interests include academic library system, ICT application in libraries, bibliometrics, scientometrics, Altmetrics, information sources and services, library automation, digital library, etc.

Mr. Sanmati Jinendran Jain works as Assistant Librarian (Group 'A' Officer) at the Central Library, National Institute of Technology, Tiruchirappalli since January 2025. Prior to this role, he served as Library Superintendent at the Indian Institute of Technology Goa. His professional journey also includes significant contributions at esteemed academic and corporate institutions, such as Central University of Tamil Nadu, Thiruvarur; Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram; Tata Consultancy Services (TCS), Chennai; and Indian Institute for Human Settlements, Bangalore. He qualified for UGC-NET (Assistant Professor) – Library & Information Science in January 2018 and is pursuing his PhD (Library & Information Science) at the Indira Gandhi National Open University, New Delhi. He holds B.Sc. (Biochemistry) from Vels University, Chennai; M.Lib.I.Sc. (Library & Information Science) from the University of Madras, Chennai, and Post Graduate Diploma in Library Automation and Networking from the University of Hyderabad, Telangana. He has been recognized for his academic excellence, receiving the Prof. Parvataneni Gangadhara Rao Memorial Medal for securing the first rank in the M.Lib.I.Sc. (Library & Information Science) programme and Gold Medal for achieving the first rank in the B.Sc. (Biochemistry) programme. His areas of interest include Research Integrity, Altmetrics, and Open Science.

Blind Peer Reviewed
(Previously UGC-CARE)
ISSN: 0972-2750

Manuscript Received on –
30th December, 2025
Primary Reviewed on –
6th January, 2026
Peer Reviewed on –
24th January, 2026
Revised on –
30th January, 2026
Accepted for publication
7th February, 2026

Designing an Interoperable Metadata Schema for Indigenous Language Archiving: Mapping Customised QDC to OLAC, IMDI and CMDI

Samhati Soor

ABSTRACT:

Introduction: *The archiving of indigenous languages is important because it documents linguistic units of native or lost languages, systematically manages ethnographic information, and preserves traditional knowledge of a particular community. Digital Archiving is an integral part of the future viability of linguistic diversity and cultural heritage, depending upon how well language data is handled. Resource discovery, sharing and sustainability can be accelerated through metadata interoperability and long-term preservation techniques.*

Purpose: *This study aims to implement standard archiving practices by designing a robust, scalable, and flexible metadata schema that documents descriptive, technical, ethical, preservation, and linguistic metadata for indigenous language resources, enabling them in the long run.*

Research Problem: *Although existing metadata standards of the Linguistics domain, such as OLAC (Open Language Archives Community), IMDI (ISLE MetaData Initiative) or CMDI (Component MetaData Infrastructure), offer dominant frameworks in the domain of linguistics, the inconsistent application of these standards in digital archiving results in the creation of barriers to cross-platform interoperability. The research problem addresses the lack of such a metadata framework, which is standardised, interoperable and culturally sensitive.*

Objective: *The objective is to develop a generic, adaptive and interoperable metadata schema for indigenous language archiving, demonstrating the schema's alignment to Qualified Dublin Core (QDC) and establishment of its mappings to existing metadata standards of the linguistic field, supporting sustainable management of multimodal resources.*

Methodology: *The metadata schema is designed by defining semantics and incorporating necessary ethical notes. Each metadata field is first aligned with QDC, with or without custom properties, and then mapped to its equivalents in OLAC, IMDI, and CMDI. Special attention is given to embracing controlled vocabularies, creating persistent identifiers, using ethical metadata, documenting responsibility, field notes, and linguistic encoding, and validating through representative metadata records for different file types to establish authenticity, consistency, and sustainability.*

Findings: *As demonstrated, the schema reliably captures comprehensive metadata of diverse indigenous language resources. It enables cross-archive interoperability, enhances discoverability, supports ethical and culturally informed access and provides indigenous knowledge to the communities and researchers. The schema's modularity and scalability allow its application in small community projects to large institutional archives, ensuring the long-term preservation and responsible sharing of indigenous language materials.*

Ms. Samhati Soor
Librarian
Heramba Chandra College,
Kolkata, West Bengal , India
hccsamhati@gmail.com

Keywords: *Metadata Schema, Qualified Dublin Core, Language Archiving, Indigenous Languages, Metadata Interoperability, Digital Preservation*

1. Introduction

The preservation and documentation of indigenous languages are critical for safeguarding linguistic diversity, cultural heritage and community knowledge. The sustainable means of capturing language data, yet its effectiveness depends on how systematically and accurately these resources are described, structured and maintained, is facilitated through digital archiving procedures. Metadata plays a principal role in discoverability, interoperability, and the long-term usability of archived language materials. Existing linguistic metadata standards, such as the Open Language Archives Community (OLAC), the ISLE Metadata Initiative (ISLE stands for International Standards for Language Engineering), the IMDI, and the Component Metadata Infrastructure (CMDI), provide valuable frameworks for describing linguistic resources. However, irregular and erratic execution across repositories, limited support for complex relationships and insufficient integration of ethical and community-centred practices often block cross-platform interoperability and responsible management of culturally sensitive content. To address these challenges, this study proposes a generic, adaptive, and flexible metadata schema based on Qualified Dublin Core (QDC), with detailed mappings to OLAC, IMDI, and CMDI. The schema accommodates diverse resource types and supports both language-level and item-level metadata, ensuring comprehensive representation of linguistic content as well as cultural context. This study aims to provide a robust, coherent and interoperable approach to metadata design, offering critical insights for archivists, linguists and community practitioners.

2. Review of Literature

The documentation and archiving of indigenous and endangered languages have been widely discussed in scholarly literature over the past two decades, largely in response to language loss and the deterioration of community knowledge systems. Early studies focus on the urgency of systematic documentation and recognising linguistic diversity as a vital component of human cultural heritage (Crystal, 2000). Further developments in language documentation move from data collection to comprehensive documentation integrating linguistic structure, language use, sociocultural context and long-term sustainability (Austin, 2013; McDonnell, Berez-Kroeker, & Holton, 2018). In the meantime, digital archiving emerges to enable continued accessibility, reusability and preservation of language resources beyond individual research projects (Henke & Berez-Kroeker, 2016; Trilsbeek & König, 2014). As digital repositories became relevant to documentation workflows, metadata gained prominence for organising, describing and exchanging linguistic resources. Bird and Simons (2003) defined portability as a foundational principle for sustainable language documentation, linking metadata quality, standardised formats and access mechanisms to long-term usability. General-purpose standards, such as Dublin Core, enable cross-domain interoperability through minimal, extensible metadata (DCMI Usage Board, 2012, 2020), though linguistic researchers argue that indigenous language documentation demands richer, domain-specific metadata capable of representing linguistic complexity and contextual depth.

OLAC and IMDI emerge as dominant metadata standards to satisfy the specific needs of language documentation. Derived from Dublin Core, OLAC extends core elements through language identifiers and linguistically specific resource descriptions, facilitating aggregation and discovery across language archives (Bird & Simons, 2001, 2003; Open Language Archives Community, n.d.), although empirical studies revealed persistent inconsistencies in the application of OLAC vocabularies, undermining semantic interoperability and limiting effective cross-archive integration (Paterson III, 2022, 2023, 2024). Moreover, OLAC's relatively flat structure restricts its capacity to represent complex relationships required for multimodal and session-based documentation. IMDI addresses this limitation by offering a hierarchical metadata model designed to capture detailed information on sessions, projects, participants, recordings and contextual relationships (Broeder & Wittenburg, 2006). Though IMDI proves effective for managing audiovisual corpora within large-scale documentation projects, researchers identify challenges related to schema rigidity, implementational complexity and long-term adaptability, particularly when interoperability with non-linguistic repositories is required (Klassmann et al., 2006; The Language Archive, 2020). Crosswalks between OLAC and IMDI exist; they primarily focus on technical alignment and offer limited support for integrating the cultural and ethical dimensions of language documentation. The introduction of CMDI attempts to address heterogeneity and complexity in linguistic metadata by enabling modular and component-based metadata profiles that support semantic interoperability across infrastructures such as CLARIN or Common Language Resources and Technology Infrastructure (Broeder et al., 2011; Broeder et al., 2012; Witt & Wittenburg, 2009; Van Uytvanck, Stehouwer, & Lampen, 2012; Windhouwer & Goosen, 2022). Studies demonstrate CMDI's potential to bridge discrete schemas (Durco & Windhouwer, 2013); however, its technical complexity presents significant barriers for smaller, community-based or resource-constrained projects (Burke et al., 2022; Hooshiar, 2023).

Alongside these technical developments, scholars increasingly emphasise that metadata is not neutral but actively shapes access, authority and representation of indigenous knowledge (Farnel, 2018). Although ethical principles such as provenance documentation, community consent and access differentiation are widely advocated (Bow & Hepworth, 2019; Booth, 2020), they are rarely embedded structurally within interoperable metadata frameworks. These challenges are further intensified in multilingual contexts, such as India, where multiple scripts, orthographic variation and uneven digitisation practices complicate metadata standardisation, as reflected in national initiatives such as NAM, NDLI, Indian Culture (formerly NVLI), TDIL and Bhararvani. Indian LIS scholarship cautions against the uncritical application of Western metadata models to indigenous and regional materials, underscoring the need for culturally adaptive and participatory approaches (Saxena & Kaur, 2025). Collectively, the literature reveals a clear research gap - the absence of a generic, adaptive and interoperable metadata schema that systematically aligns Qualified Dublin Core and also remains extensible towards OLAC, IMDI and CMDI, embedding ethical and community-centred considerations particularly within linguistically and culturally complex archival environments.

3. Objectives of the Study

This study seeks to address the challenges of documenting and archiving indigenous languages by pursuing several interrelated objectives.

First, it aims to develop a generic, adaptive and robust metadata schema suitable for both community-based and institutional repositories.

Second, it attempts to map schema-aligned Dublin Core qualifiers to the OLAC, IMDI, and CMDI standards to enable interoperability across different archives and platforms.

Third, the study sets out practical guidance for applying structured metadata to multi-lingual and multimodal resources.

Fourth, it strives to improve the discoverability, reusability and long-term sustainability of language resources using controlled vocabularies and persistent identifiers.

Finally, the framework pursues ethically grounded, community-centred metadata principles to ensure the responsible management of linguistic and cultural knowledge.

4. Scope of the Study

This study comprehensively examines the following key aspects and considerations in the design of the interoperable metadata schema for indigenous language archiving.

Generic yet Adaptive Schema: The study offers a flexible metadata schema suitable for archiving indigenous language and culture across diverse archival contexts, including community archives, academic institutions, and government repositories.

Coverage of Multimedia: The schema supports multi-type and multi-formatted resources, including audio, video, text, images, and annotated corpora, across a wide range of file formats.

Dual-Level Support: The schema incorporates both language-level (macro) and item-level (micro) descriptions, providing a rich, detailed representation of linguistic resources.

Ethical Metadata Integration: Culturally sensitive elements, such as rights, ownership, community permissions, and provenance, which are often underrepresented in technical metadata studies, can be contextualised through appropriate metadata fields.

Unified Crosswalk: The study presents one of the first comprehensive frameworks that establishes links between QDC, OLAC, IMDI, and CMDI, specifically for indigenous language archiving, to enhance interoperability and standardisation.

Guidelines for Best Practices: Recommendations for controlled vocabularies, access controls and documentation protocols are thoroughly provided.

Applicability to the global landscape: The study focuses on metadata design in Indian and other multilingual, multicultural contexts, addressing local archival challenges that remain unexplored in existing LIS scholarship supporting UTF-based annotations.

The study does not involve software development or documentation of specific field datasets, but aims to provide a generalizable metadata framework applicable across diverse documentation projects.

5. Metadata Standards in Context

Metadata standards provide structured ways to describe, manage and share digital resources, making them easier to discover, reuse and preserve. The existing standards, such as Dublin Core, OLAC, IMDI and CMDI, differ in scope, complexity and focus, but each contributes consistent and meaningful metadata practices when applied correctly. The purposes, key features, strengths and limitations of these standards are outlined in Table 1.

Metadata Standard	Purpose / Focus	Key Features	Strengths	Limitations
DC (Dublin Core)	Cross- domain, general- purpose metadata	15 core elements - Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights. Refinements with 55 terms, including 40 properties as qualifiers of the elements, encoding schemes and controlled vocabularies through Qualified DC (QDC).	Simple and Customizable. Widely adopted. Interoperable across domains. Easy to implement. Extensive multimedia support. Relationship establishment option available among items.	Lacking granularity for language-specific or culturally sensitive resources without the use of qualifiers and/or customisation, such as the inclusion of custom properties.
OLAC (Open Language Archives Community)	Domain- specific to language resources	ISO 639-3 language codes. Linguistic resource types (lexicon, grammar, primary text, etc.). Contributor roles (speaker, transcriber, annotator). Controlled vocabularies.	Precision for linguistic resources added. Widely used for endangered/minority languages. Consistent terminology.	Limited multimedia support. Less hierarchical detail than IMDI. Mainly focused on metadata for resource description rather than session-level or relational information.
IMDI (ISLE Metadata Initiative)	Multimedia and multimodal linguistic data	Hierarchical structure capturing actors, recording sessions, media specifications, contextual information and relationships between annotations, translations and media.	Very rich and detailed. Ideal for audiovisual corpora. Support for complex documentation workflows.	Complex to implement. Less interoperable with non-linguistic or simple repositories.

				Requires mapping for cross-standard integration.
CMDI (Component Metadata Infrastructure)	Flexible component based metadata for linguistic and cultural resources	Modular and reusable components. Customizable metadata profiles. Upholding multilingual, hierarchical and complex relationships.	Highly flexible. Suitable for complex resources.	Requires higher technical expertise. Challenging implementation and maintenance.

Table- 1: Comparative Overview of Metadata Standards - Dublin Core, OLAC, IMDI and CMDI

Dublin Core offers simplicity and universality, making it useful for cross-domain interoperability. OLAC and IMDI add the necessary linguistic specificity and contextual detail, which is essential for indigenous language documentation. CMDI further extends this flexibility by supporting modular and customizable metadata profiles. Inconsistent application and partial adoption of these standards across repositories create fragmented metadata ecosystems, limiting discoverability, interoperability and long-term preservation. This shows a genuine need for a generic, adaptive and flexible metadata schema capable of aligning with all these standards, preserving both simplicity and richness, and respecting linguistic, cultural and ethical requirements.

6. The Schema: Mapping Customised QDC to OLAC, IMDI and CMDI

The proposed metadata schema includes a set of metadata fields essential for consistent digital archiving, while also addressing ethical, cultural and community considerations in the documentation and sharing of language resources. A clear depiction of how each field, along with its semantics, syntax/vocabulary encoding schemes for value insertion, necessary ethical notes, and alignment with QDC, is provided in Table 2.

Sl. No.	Metadata Field with Semantics, Syntax/Vocabulary Encoding Scheme or Code for Value Insertion and necessary Ethical Note	Alignment to Qualified Dublin Core
1.	**Language** Semantics: Language to archive. Syntax Encoding Scheme: en_US (English United States) and ISO 639-3 (Ethnologue)	dcterms. language
2.	**Responsibility of Language** Semantics: Person(s), institution, organisation, or community that is responsible for the entire archiving of a particular indigenous language Code: en_US or controlled vocabulary, depending on the Project Ethical Note: Explicit permission from the language community. Respecting community authority over linguistic and cultural data.	dcterms.contributor. *responsibilityOfLanguage* [customised]
3.	**Title** Semantics: Title/Name of the Resource. Code: en_US	dcterms.title

4.	**Subject** Semantics: Subject/Topic of the Resource represented by keywords, key phrases or classification codes. Code: en_US and/or Notation of Classification Scheme (LCC/DDC)	dcterms.subject dcterms.subject. *classificationCode* [customized]
5.	**Date** Semantics: Date of creation, modification, formal issuance and availability of the resource and copyright imposed on the resource. Syntax Encoding Scheme: ISO 8601 [YYYY-MM-DD]	dcterms.created dcterms.modified dcterms.issued dcterms.available dcterms.dateCopyrighted
6.	**Location** Semantics: Spatial applicability or relevancy of the resource, such as the location of the fieldwork (GPS reading) or the native place of the informant. Code: en_US and/or Notation of Latitude and Longitude	dcterms.spatial
7.	**Time** Semantics: Temporal information of the resource, such as when data is collected on a particular day. Syntax Encoding Scheme: W3C DTF [Complete date plus hours and minutes: YYYY-MM-DDThh:mmTZD (eg 1997-07-16T19:20+01:00) TZD = time zone designator (Z or +hh:mm or -hh:mm)]	dcterms.temporal
8.	**Resource Person** Semantics: The knowledge expert who has lived experience of the particular indigenous language and contributes directly to the documentation process, such as native speakers, elders, community leaders, storytellers, singers, poets or ritual specialists. Code: en_US Ethical Note: Consent must be obtained before providing linguistic or cultural data. Protecting anonymity if requested. Clear documentation of ethical agreements, if any.	dcterms.contributor. *resourcePerson* [customized]
9.	**Responsible Person for Resource** Semantics: Responsible Person(s) on behalf of the institution, organisation, or community who are responsible for collecting, editing or curating the resource. Code: en_US Ethical Note: Individual authorisation to collect, edit or curate resources. Maintaining accountability as well as respecting confidentiality and community protocols.	dcterms.contributor. *responsiblePersonForResource* [customized]
10.	**Responsible Person for Recording** Semantics: Responsible Person(s) on behalf of the institution, organisation, or community who are responsible for capturing linguistic or cultural data directly	dcterms.contributor. *recordingTeam* [customized]

	from the field and/or informants. Code: en_US Ethical Note: Following ethical guidelines. Obeying community norms and privacy. Informed consent for all recordings.	
11.	**Recording Equipment** Semantics: Equipment used for capturing the resource. Code: en_US Ethical Note: Use of equipment ensuring no inadvertent disclosure of sensitive practices or locations. Choosing devices that minimise intrusion or disruption.	dcterms.description. *recordingEquipment* [customized]
12.	**Size** Semantics: File size of the resource. Code: en_US	dcterms.extent. *size* [customized]
13.	**Duration** Semantics: Duration of playing or execution of the resource. Code: hh:mm:ss	dcterms.extent. *duration* [customised]
14.	**Type** Semantics: File type of the resource, such as text, image, audio, video or others. Vocabulary Encoding Scheme: DCMI Type	dcterms.type
15.	**Format** Semantics: File format of the resource, such as PNG image, GIF image, MPEG audio, WAV audio, MP4 video. Code: Internet Media Type (MIME Type)	dcterms.format
16.	**Content Description** Semantics: Brief description of the content and context of the resource. Code: en_US Ethical Note: Revealing no culturally sensitive content publicly. The use of controlled language was agreed with the community. Maintaining accuracy while protecting restricted knowledge.	dcterms.description. *contentDescription* [customized]
17.	**Field Note** Semantics: Observations made during fieldwork related to the particular resource. Code: en_US Ethical Note: Inclusion of only ethically approved observations. Distinct separation of sensitive cultural information.	dcterms.description. *fieldNote* [customized]
18.	**Team Member** Semantics: Responsible Person(s) of the institution or organisation or community involved in the archiving of a particular indigenous language, other than the Responsible Person for Language, the Responsible Person for Resource and the Responsible Person for Recording	dcterms.contributor. *teamMember* [customized]

	Code: en_US	
19.	**Publisher** Semantics: Person(s) or Organisation or Institution that is authorised to publish the resource. Code: en_US	dcterms.publisher
20.	**Audience** Semantics: For whom the resource is intended or useful. Code: en_US	dcterms.audience
21.	**Provenance** Semantics: Origin of the resource including creation, ownership and any important changes made to it. Code: en_US Ethical Note: Applying access restrictions according to community decisions. Certain data for researchers only. Following CARE/OCAP principles.	dcterms.provenance
22.	**Rights** Semantics: Information about rights held in and over the resource. Code: en_US or URI Ethical Note: Documenting the chain of custody in a way that acknowledges community ownership and protects against overstepping community boundaries. Assurance of transparency and accountability.	dcterms.rights
23.	**Rights Holder** Semantics: A person, organisation, or institution owning or managing rights over the resource. Code: en_US Ethical Note: Attributing rights to individuals or communities with authority over the data. Preventing misrepresentation or exclusion of community authority.	dcterms.rightsHolder
24.	**Acknowledgement** Semantics: Acknowledgement to the speakers, community or others who contributed to archiving the resource. Code: en_US Ethical Note: Protecting privacy and sensitive identities. Seeking consent before public acknowledgement.	dcterms.description. *acknowledgement* [customised]
25.	**Source** Semantics: A Reference to a resource from which this resource is derived. Syntax Encoding Scheme: URI Ethical Note: Attributing derivative resources, acknowledging local ownership and intellectual property.	dcterms.source
26.	**Relation** Semantics: A reference to a related resource or a collection of resources.	dcterms.relation [this is a generic property. others include - isVersionOf, isReplacedBy, replaces, isRequiredBy, requires,

	Syntax Encoding Scheme: URI Ethical Note: Prohibition to expose sensitive connections. Respecting cultural protocols when linking related materials.	isPartOf, hasPart, isReferencedBy, references, isFormatOf, hasFormat, hasVersion]
27.	**Identifier** Semantics: Unambiguous identifier of the resource. Syntax Encoding Scheme: DOI (Digital Object Identifier)	dcterms.identifier
28.	**International Phonetic Alphabet** Semantics: Description of the resource, how it is pronounced using the International Phonetic Alphabet (IPA), applicable to word or sentence data, to represent the pronunciation in a specific indigenous language based on an audible resource, collected directly during fieldwork or extracted through sound editing. Code: IPA	dcterms.description. *internationalPhoneticAlphabet* [customized]
29.	**National Level Scripting** Semantics: Description of the resource, how it is written using a national-level script, such as the Devanagari script for India. Code: National-level script	dcterms.description. *nationalLevelScript* [customized]
30.	**Regional Level Scripting** Semantics: Description of the resource, how it is written using a national-level script, such as the Bengali script for West Bengal, India. Code: Regional level script	dcterms.description. *regionalLevelScript* [customized]
31.	**Morpheme Break** Semantics: Splitting a sentence into its smallest meaningful units or morphemes to show internal word structure Code: Linguistic representational norms	dcterms.description. *morphemeBreak* [customized]
32.	**Gloss** Semantics: Providing a meaning or grammatical label of a sentence for each morpheme after breaking. Code: Linguistic representational norms	dcterms.description. *gloss* [customized]

Table- 2: Proposed Metadata Schema aligned to Qualified Dublin Core

Each metadata field of the schema, along with its QDC equivalent, is mapped to three widely recognised metadata standards of the linguistics field, OLAC, IMDI and CMDI, as embodied in Table 3, accommodating both standard and custom properties to ensure that resources remain interoperable and can be seamlessly shared across archives adhering to these standards.

Sl. No.	Metadata Field with QDC Equivalent	Mapping to OLAC	Mapping to IMDI	Mapping to CMDI
1.	**Language** dcterms.language	olac:language	Content/Contentlanguages/ContentLanguage Actor/ActorLanguages/ActorLanguage	cmdi-language cmdi-language:LanguageName

				cmdi-actorlanguage
2.	**Responsibility of Language** dcterms.contributor. *responsibilityOfLanguage*	olac:role:*extension* [customized]	Actor/Role/*extension* [customized]	[No corresponding CMDI component registry]
3.	**Title** dcterms.title	same as QDC	[no such corresponding]	cmdi-generalinfo:Title
4.	**Subject** dcterms.subject dcterms.subject. *classificationCode*	same as QDC olac:language [language as subject] olac:linguistic-field [linguistic field as subject]	Content/Subject	cmdi-content:Topic
5.	**Date** dcterms.created dcterms.modified dcterms.issued dcterms.available dcterms.dateCopyrighted	same as QDC	Session/RecordingDate	[No corresponding CMDI component registry]
6.	**Location** dcterms.spatial	same as QDC	Session/LocationAddress Session/LocationCountry Session/LocationContinent Session/Regions	cmdi-location cmdi-location:region
7.	**Time** dcterms.temporal	same as QDC	Session/Description/*extension* [customized]	[No corresponding CMDI component registry]
8.	**Resource Person** dcterms.contributor. *resourcePerson*	olac:role:consultant	Actor/Role/*extension* [customized]	cmdi-actor
9.	**Responsible Person for Resource** dcterms.contributor. *responsiblePersonForResource*	olac:role:researcher	Actor/Role/*extension* [customized]	[No corresponding CMDI component registry]
10.	**Responsible Person for**	olac:role:recorde	Actor/Role/*extension*	[No corresponding CMDI component

	Recording dcterms.contributor. *recordingTeam*	r	[customized]	registry]
11.	**Recording Equipment** dcterms.description. *recordingEquipment*	same as QDC [customized]	Resources/Description/ *extension* [customized]	cmdi-description
12.	**Size** dcterms.extent. *size*	[no such corresponding]	Resources/MediaFile/Size	[No corresponding CMDI component registry]
13.	**Duration** dcterms.extent. *duration*	[no such corresponding]	Resources/MediaFile/TimePosition	[No corresponding CMDI component registry]
14.	**Type** dcterms.type	same as QDC olac:linguistic-type olac:discourse-type	Resources/MediaFile/Type	cmdi-mediafile:Type
15.	**Format** dcterms.*format*	same as QDC	Resources/MediaFile/Format	[No corresponding CMDI component registry]
16.	**Content Description** dcterms.description. *contentDescription*	same as QDC [customized]	Content/Description	cmdi-content
17.	**Field Note** dcterms.description. *fieldNote*	same as QDC [customized]	Resources/Description/ *extension* [customized]	cmdi-description
18.	**Team Member** dcterms.contributor. *teamMember*	olac:role:*extension* [customized]	Actor/Role/*extension* [customized]	[No corresponding CMDI component registry]
19.	**Publisher** dcterms.publisher	same as QDC	Catalogue/Publisher [partial]	[No corresponding CMDI component registry]
20.	**Audience** dcterms.audience	[no such corresponding]	[no such corresponding]	[No corresponding CMDI component registry]

21.	**Provenance** dcterms.provenance	[no such corresponding]	[no such corresponding]	[No corresponding CMDI component registry]
22.	**Rights** dcterms.rights	[no such corresponding]	Resources/Access GenericResources/Access Lexicon/Access	cmdi-access
23.	**Rights Holder** dcterms.rightsHolder	[no such corresponding]	[no such corresponding]	[No corresponding CMDI component registry]
24.	**Acknowledgement** dcterms.description. *acknowledgement*	same as QDC [customized]	Resources/Description/ *extension* [customized]	cmdi-description
25.	**Source** dcterms.source	same as QDC	Resources/Source	cmdi-source
26.	**Relation** dcterms.relation	same as QDC	Resources/ResourceLink GenericResources/ResourceLink WritingResources/MediaResourceLink	[No corresponding CMDI component registry]
27.	**Identifier** dcterms.identifier	same as QDC	Resources/Key GenericResources/Key	cmdi-generalinfo:ID
28.	**International Phonetic Alphabet** dcterms.description. *internationalPhoneticAlphabet*	same as QDC [customized]	ResourceData/Annotation/AnnotationType/*extension* [customized]	cmdi-annotationtypes
29.	**National Level Scripting** dcterms.description. *nationalLevelScript*	same as QDC [customized]	ResourceData/Annotation/AnnotationType/*extension* [customized]	cmdi-annotationtypes
30.	**Regional Level Scripting** dcterms.description. *regionalLevelScript*	same as QDC [customized]	ResourceData/Annotation/AnnotationType/*extension* [customized]	cmdi-annotationtypes
31.	**Morpheme Break** dcterms.description.	same as QDC [customized]	ResourceData/Annotation/AnnotationType/*extension*	cmdi-annotationtypes

	morphemeBreak		[customized]	
32.	**Gloss** dcterms.description. *gloss*	same as QDC [customized]	ResourceData/Annotation/AnnotationType/*ex tension* [customized]	cmdi-annotationtypes

Table- 3: Interoperability between Qualified Dublin Core and Linguistic Metadata Standards - OLAC, IMDI and CMDI

It is worth noting that the term "extension" refers to custom additional properties beyond the core property or path in the case of OLAC or IMDI. CLARIN Component Registry is a central repository of reusable metadata components, primarily for building CMDI profiles. Although it also allows metadata components or fields shared by metadata users beyond the profile-based CMDI framework, alignment is made using CMDI's own components only. For some metadata fields in the proposed schema, corresponding components are not available in one or more of the domain-specific standards, such as provenance, audience or rights, consequently, no direct conversion is required to adopt the schema. Their deliberate inclusion in the schema supports future compatibility and adaptability, fulfilling its generic and adaptive design goals.

7. Implementation and Validation

The schema can be implemented using structured formats such as XML, JSON-LD or RDF. To ensure systematic, ethical, and sustainable use, metadata should include persistent identifiers such as DOIs, clearly document collectors and contributors, permissions, and ethical reviews, and provide multilingual support in both English and relevant indigenous languages. Hierarchical links should reflect relationships between related resources, and access controls should protect culturally sensitive materials in line with ethical, community and legal considerations.

Metadata should also be validated against recognised standards to ensure interoperability and adherence to established best practices. To incorporate custom properties, such as dcterms.description.*fieldNote*, into QDC in RDF/XML, an extension namespace must be defined, and each customised element must be declared as an rdf:Property. Each custom property has to be linked to the appropriate term-based property using rdfs:subPropertyOf, e.g., a custom *fieldNote* property as a subproperty of dcterms:description. Model refinements may be done as independent properties with clear semantics and custom properties can then be used normally in metadata records.

Following the implementational measures ensures indigenous language resources are preserved, shared and accessed responsibly. Representative metadata are presented in Table 4 to demonstrate the process of consistently documenting descriptive, technical, ethical, and access-related information to ensure the accuracy, operationalisation, and validation of the schema.

Sl. No.	Metadata Label	Metadata Value for Sample Image Resource	Metadata Value for Sample Audio Resource	Metadata Value for Sample Video Resource
1.	Language	Navajo [en_US], nav [ISO 639-3]	Yoruba [en_US], yor [ISO 639-3]	Kurmali [en_US], kxv [ISO 639-3]
2.	Responsibility of Language	John Smith	Linguistic Unit, University of Ibadan	LM [Project specific code]
3.	Title	Weaving Pattern	A person of good character is telling a story.	Folk Dance Performance
4.	Subject	Indigenous art, Textile symbolism	Oral tradition, Folklore, Sentence Data	Folk dance, Performance tradition, Performing arts

5.	Date	2023-06-10 [created]	2022-09-18 [created] 2022-10-02 [modified] 2022-10-05 [available]	2024-02-05 [created]
6.	Location	Window Rock, Arizona, USA 35.96° N, 109.05° W	Ibadan, Oyo State, Nigeria 7.3775° N, 3.9470° E	Jhalda, Purulia, West Bengal, India 23.33° N, 86.23° E
7.	Time	2023-06-10T14:20-07:00	2022-09-18T17:45+01:00	2024-02-05T09:30+05:30
8.	Resource Person	Navajo weaver (name withheld)	Oba Adeyemi (Elder storyteller) [consented]	Kunti Mahato and group (Kurmali dance troupe lead performers) [All of them consented]
9.	Responsible Person for Resource	Emily R. Johnson	Amina Ndlovu	GK
10.	Responsible Person for Recording	Sarah L. Thompson Daniel K. Rivera	Kwame Mensah	-
11.	Recording Equipment	Nikon D750 DSLR Camera	Zoom H4n Pro Recorder	Sony PXW-Z90 Video Camera
12.	Size	25 MB	850 MB	1.2 GB
13.	Duration	-	00:42:15	00:18:40
14.	Type	Image	Audio	Video
15.	Format	JPEG (image/jpeg)	WAV (audio/x-wav)	MP4 (video/mp4)
16.	Content Description	High-resolution image of a traditional Navajo rug with symbolic motifs	Audio of a sentence while narration of a moral folktale performed in Yoruba	Video recording of the Kurmali folk dance performed during the harvest festival
17.	Field Note	Photograph taken with community approval, sacred symbols excluded	Audience interaction present, background village sounds audible	Certain ritual movements are marked as restricted
18.	Team Member	Jacob A. Miller (Field Recording Specialist), Rebecca Anderson (Documentation Officer)	Fatoumata Diarra	BP SL
19.	Publisher	University of Arizona	Linguistic Unit, University of Ibadan	Registrar, Rabindra Bharati University
20.	Audience	Art historians, the Navajo community	Linguists, Educators	Kurmali community, Researchers
21.	Provenance	Community-commissioned cultural documentation project	Recorded under community-approved folklore initiative	Community-led documentation

22.	Rights	CC-BY-NC	Custom community license (non-commercial)	Rights to the Community, no derivative works
23.	Rights Holder	Navajo Nation	Yoruba Community Council	Kurmali Community
24.	Acknowledgement	Gratitude to the Navajo weavers and elders	Thanks to Yoruba storytellers and hosts	Acknowledgement to the performers and villagers of Jhalda block
25.	Source	https://archive.example.org/navajo_img_050	https://hdl.handle.net/10580/632 [un-cut recording created on 2022-09-18]	-
26.	Relation	isPartOf -> Navajo Textile Collection	hasVersion -> https://hdl.handle.net/10580/732	isPartOf -> Kurmali Ethnographic Data
27.	Identifier	https://doi.org/10.1575/1913/1080	https://hdl.handle.net/10580/659	https://doi.org/10.5227/zenodo.17076311

Table- 4: Representative Metadata Records for Multimodal Indigenous Language Resources - Image, Audio and Video

Metadata for sample sentence data in Mundari, an Indian Indigenous language, is exemplified in Table 5, providing notational pronunciation, transliteration using relevant scripts, word-by-word meanings, and clarification of both lexical and grammatical structure.

Sl. No.	**Metadata Label**	**Metadata Value for Sample Sentence Data**
1.	Title [English Meaning of the Sentence]	They are going to the river.
2.	International Phonetic Alphabet	hoːŋ daɽate geɽi a
3.	National Level Scripting [Transliteration using Devnagari Script]	हों दर-ते गेड़ी अ
4.	Regional Level Scripting [Transliteration using Bengali Script]	হোঁ দর-তে গেড়ি অ
5.	Morpheme Break	hoːŋ daɽa-te ge-ɽi a
6.	Gloss	they-3PL river-to go-PROG COP-PRES

Table- 5: Linguistic Metadata Example for sample sentence data in the Mundari Language with IPA, Script Representation and Morphological Analysis

These examples reflect that the schema can reliably capture descriptive, technical, ethical, preservation and linguistic metadata across diverse resource types, confirming its interoperability and usability in real-world archival contexts.

8. Assessment and Implications

The metadata schema provides a robust framework for the effective and sustainable documentation of indigenous language resources. By enabling mappings between existing standards, it improves cross-archive interoperability, allowing different repositories to share and communicate resources seamlessly. Consistent and structured metadata enhances discoverability, making it easier for researchers, communities and educators to locate, understand and reuse linguistic materials, and enabling long-term preservation and sustainable management of digital resources. The schema's flexibility and scalability allow it to be applied from small community projects to large institutional archives, and to support multilingual or multimodal datasets. Implementing the schema requires careful attention to several challenges to ensure consistency and reliability. Such challenges include uneven updates to controlled vocabularies to reflect linguistic, cultural, and technological changes; stakeholders' lack of expertise; and technical complexity, such as establishing hierarchical relationships within large corpora. These kinds of challenges can be addressed through regular vocabulary updates, ongoing archivist training, modular implementation of metadata components and close collaboration with language communities. The framework allows indigenous language materials to be responsibly documented, preserved and shared as research practices and technologies evolve.

The schema ensures that each resource is accurately described, easily accessible and ethically grounded through structured metadata. Core metadata fields such as language, title, subject, date, location and time provide contextual information, whereas contributor and team details emphasise accountability, attribution and recognition of community knowledge holders. Ethical notes protect rights, consent and sensitive cultural information. Technical fields capture recording details and file properties, accelerating preservation and usability. Descriptive fields, such as content and field notes, offer context for the resource. Rights, rights holders and identifiers enable legal clarity, adherence to intellectual property principles and persistent referencing, protecting both the community and the institution. The metadata fields provenance, date modified and relation record changes to resources, including transcription updates or corrected annotations, capture what changes were made and when, ensuring traceability and transparency in the workflow while establishing version control through clear links between different versions or related resources. Scripting standards provide accurate pronunciation, national readability and regional representation. Morphological information supports interpretation, comparability and computational processing, e.g., in corpus linguistics or NLP applications. Collectively, these make indigenous language archives comprehensive, ethically responsive and scientifically valuable, promoting language preservation, revitalisation, and research.

9. Conclusion

The study demonstrates how systematic, structured and standardised practices can effectively support the documentation and archiving of indigenous languages. By designing a robust and adaptive schema that aligns with the widely used metadata standard Dublin Core and integrates with multiple established frameworks, the study ensures that linguistic resources can be accurately represented, contextually rich, and technically interoperable. This approach facilitates reliable access, interpretation and reuse of resources across diverse archives, benefiting researchers, communities, archivists, linguists and institutions alike. Importantly, the schema embeds ethical and cultural considerations directly into metadata practices, reinforcing respect for community authority and stewardship in effective archiving strategies, confirming the discoverability, accessibility, usability, sustainability, and integrity of endangered language resources, and facilitating linguistic or ethnographic research. Thus, the study contributes significantly to digital preservation, metadata design and indigenous language documentation, advancing both conceptual understanding and real-world application. These contributions make the study methodologically innovative and practically valuable, propelling responsible, sustainable and interoperable archiving of indigenous languages.

References

Austin, P. K. (2013). Language documentation and meta-documentation. *Keeping languages alive: Documentation, pedagogy and revitalization*, 3-15. https://www.academia.edu/resource/work/2098867

Bird, S., & Simons, G. (2001). The OLAC metadata set and controlled vocabularies. *arXiv preprint cs/0105030*. https://arxiv.org/pdf/cs/0105030

Bird, S., & Simons, G. (2003). Seven dimensions of portability for language documentation and description. *Language*, *79*(3), 557-582. https://arxiv.org/pdf/cs/0204020

Booth, S. (2020). *The role of archives in Indigenous language maintenance and resurgence* (Masters thesis, University of Manitoba). https://mspace.lib.umanitoba.ca/server/ap i/core/bitstreams/735c9328-3714-41af-9ef5-b08cb8ceaf46/content

Bow, C., & Hepworth, P. (2019). Observing and respecting diverse knowledge traditions in a digital archive of Indigenous language materials. *Journal of Copyright in Education & Librarianship*, *3*(1). https://journals.ku.edu/jcel/article/download/7485/9207

Broeder, D., & Wittenburg, P. (2006). The IMDI metadata framework, its current application and future direction. *International Journal of Metadata, Semantics and Ontologies*, *1*(2), 119-132. https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_57331

Broeder, D., Schonefeld, O., Trippel, T., Van Uytvanck, D., & Witt, A. (2011). A pragmatic approach to XML interoperability - the Component Metadata Infrastructure (CMDI). In *Proceedings of Balisage: The Markup Conference 2011. Montréal, Canada, August 2-5, 2011.* https://www.balisage.net/Proceedings/vol7/html/Broeder01/BalisageVol7-Broeder01.html

Broeder, D., Windhouwer, M., Van Uytvanck, D., Goosen, T., & Trippel, T. (2012). CMDI: a component metadata infrastructure. In *Describing LRs with metadata: towards flexibility and interoperability in the documentation of LR workshop programme* (Vol. 1, p. 47). https://d-nb.info/1251813941/34

Burke, M., Zavalina, O. L., Phillips, M. E., & Chelliah, S. (2020). Organization of knowledge and information in digital archives of language materials. *Journal of Library Metadata*, *20*(4), 185-217. https://digital.library.unt.edu/ark:/67531/metadc1783013/m2/1/high_res_d/JLM_nonanonymized.pdf

Burke, M., Tarver, H., Phillips, M. E., & Zavalina, O. (2022). Using existing metadata standards and tools for a digital language archive: A balancing act. *The Electronic Library*, *40*(5), 579-593. https://digital.library.unt.edu/ark:/67531/metadc2201587/m2/1/high_res_d/687236_TELproof_accepted_unencrypted.pdf

CLARIN ERIC. (n.d.). *CMDI: Component Metadata Infrastructure*. CLARIN ERIC. https://www.clarin.eu/content/cmdi-component-metadata-infrastructure

Crystal, D. (2000). *Language Death.* Cambridge University Press. https://textos.pucp.edu.pe/pdf/287.pdf

DCMI Usage Board. (2020). *DCMI metadata terms.* http://dublincore.org/specifications/dublin-core/dcmi-terms/2020-01-20

DCMI Usage Board. (2012). *Dublin Core™ Metadata Element Set, Version 1.1: Reference Description.* https://dublincore.org/specifications/dublin-core/dces

Durco, M., & Windhouwer, M. (2013). Semantic mapping in CLARIN component metadata. In *Research Conference on Metadata and Semantic Research* (pp. 163-168). Cham: Springer International Publishing. http://ndl.ethernet.edu.et/bitstream/123456789/31409/1/220.Emmanouel%20Garoufallou.pdf#page=181

Farnel, S. (2018). Metadata as data: Exploring ethical metadata sharing and access for indigenous resources through OCAP principles. In *Proceedings of the Annual Conference of CAIS/Actes du congrès annuel de l'ACSI.* https://journals.library.ualberta.ca/ojs.cais-acsi.ca/index.php/cais-asci/article/download/974/868

Henke, R., & Berez-Kroeker, A. L. (2016). A brief history of archiving in language documentation, with an annotated bibliography. https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/8d02ff23-a1eb-4aae-8b47-4b32debfc5c3/content

Hooshiar, K. (2023). *Facilitating language documentation through incremental development of digital archive infrastructure* (Doctoral dissertation, University of Hawai'i at Manoa). https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/f596d373-7977-4ac3-ba60-3830c7567f4b/content

Klassmann, A., Offenga, F., Broeder, D., & Skiba, R. (2006). IMDI metadata field usage at MPI. *Language Archive Newsletter*, 6-8. https://pure.mpg.de/rest/items/item_60279/component/file_60280/content

McDonnell, B., Berez-Kroeker, A. L., & Holton, G. (2018). 20 Years after Himmelmann 1998. https://s3-ap-southeast-2.amazonaws.com/figshare-production-eu-latrobe-storage9079-ap-southeast-2/40436006/524563_GawneL_2018.pdf?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIARRFKZQ25KW2DIYRU/20251225/ap-southeast-2/s3/aws4_request&X-Amz-Date=20251225T151615Z&X-Amz-Expires=10&X-Amz-SignedHeaders=host&X-Amz-Signature=5af2ea91ae27ea7a5d002d4a9b156cc7009282035c671e2d87fb43476ee4f0e9

Open Language Archives Community. (n.d.). *Open Language Archives Community (OLAC)*. http://www.language-archives.org

Paterson III, H. (2022). An OLAC perspective on services: The forgotten language resources. In *Proceedings of the International Conference on Dublin Core and Metadata Applications*. Dublin Core Metadata Initiative. https://dcpapers.dublincore.org/files/articles/953120677/dcmi-953120677.pdf

Paterson III, H. (2023). Towards making shared metadata interoperable across the Open Language Archives Community. In *Proceedings of the 2nd International Workshop on Digital Language Archives: LangArc 2023* (pp. 33-37). University of North Texas. https://hughandbecky.us/Hugh-CV/publication/2023-olac-making-shared-metadata-interoperable/LangArc-2_Towards_Making_Shared_Metadata_Interoperable_across_OLAC-Preprint.pdf

PATERSON III, H. (2024). Dublin Core's DCMIType 'PhysicalObject' and its use across the Open Language Archives Community. https://www2.archivists.org/sites/all/files/Paterson_Dublin%20Core's%20DCMIType%20PhysicalObject.pdf

Paterson III, H. J. (2022). Where have all the collections gone?: Analysis of OLAC data contributors' use of DCMIType 'Collection'. In *Proceedings of the 15th Annual Society of American Archivists Research Forum* (Vol. 21). https://www2.archivists.org/sites/all/files/Paterson_2021.pdf

Saxena, S., & Kaur, G. (2025). Preservation and revitalisation of Indian languages through digital archiving: A systematic review of Bharatvani. *DESIDOC Journal of Library & Information Technology*, *45*(1), 57–64. https://publications.drdo.gov.in/ojs/index.php/djlit/article/view/20138/8496

The Language Archive. (2020). *IMDI metadata information. The Language Archive Forums*. https://archive.mpi.nl/forums/t/imdi-metadata-information/2933

Trilsbeek, P., & König, A. (2014). Increasing the future usage of endangered language archives. *Language Documentation and Description*, 12. https://www.lddjournal.org/article/id/1159/download/pdf

Van Uytvanck, D., Stehouwer, H., & Lampen, L. (2012). Semantic metadata mapping in practice: the virtual language observatory. In *LREC 2012: 8th International Conference on Language Resources and Evaluation* (pp. 1029-1034). European Language Resources Association (ELRA). https://pure.mpg.de/rest/items/item_1454694_6/component/file_1454693/content

Windhouwer, M., & Goosen, T. (2022). Component metadata infrastructure. *CLARIN: The Infrastructure for Language Resources*, *1*, 191. https://uplopen.com/books/5607/files/ecb697a1-7bf2-4dbb-ac39-4e6208f19585.pdf

Witt, A., & Wittenburg, P. (2009). Metadata infrastructure for language resources and technology. *CLARIN*, *5*, 1. https://www.clarin.eu/sites/default/files/wg2-4-metadata-doc-v5.pdf

Wittenburg, P., Peters, W., & Broeder, D. (2002). Metadata proposals for corpora and lexica. In *The 3rd International Conference on Language Resources and Evaluation (LREC 2002)* (pp. 1321-1326). European Language Resources Association. https://pure.mpg.de/rest/items/item_57859/component/file_57860/content

About Authors

Ms. Samhati Soor has been the Librarian of Heramba Chandra College, affiliated with the University of Calcutta, for the last five years. She actively contributes to the Library Committee, the Internal Quality Assurance Cell, and the Technology Cell of the college, and is responsible for maintaining the college website. Prior to this, she served for three years as Documentation Officer in a UGC-sponsored project on Documentation and Archiving of Indian Indigenous and Endangered Languages at Jadavpur University, where she developed a DOI-enabled digital repository using DSpace and a project website. Earlier, she worked at Tata Consultancy Services, Kolkata and was involved in Koha-based services. She did an M.S. in Library and Information Science from the Indian Statistical Institute. Her areas of expertise include digital archiving, web-based information services and text analytics

Blind Peer Reviewed
(Previously UGC-CARE)
ISSN: 0972-2750

Manuscript Received on –
22nd December, 2025
Primary Reviewed on –
4th January, 2026
Peer Reviewed on –
04th February, 2026
Revised on –
06th February, 2026
Accepted for publication
13th February, 2026

Exploring Latent Dirichlet Allocation (LDA) as a Thematic Structure Exploration of Library and Information Science Blogs

Mithun Das & Rajesh Das

ABSTRACT

Introduction: *The rapid development of web-based communication platforms has had a remarkable impact on changing the process of creating, sharing, and consuming professional knowledge in the profession of Library and Information Science (LIS). Blogs have become one of these spaces and have gained momentum as platforms where professionals (librarians, information professionals, educators, and researchers) share experiences about their profession, discuss emerging issues, and comment on emerging practices. In contrast to classic peer-reviewed journals, LIS blogs provide timely, practice-based, and experiential information that together build the modern professional discourse of the field. Therefore, they are a rich but little-studied source of unstructured textual material on thematic trends and professional interests in LIS.*

Objectives: *The primary objective is to extract topics from the blogs in the library and information science domain. The blogs contain different themes. This study aims to gather a representative sample of blog postings by librarians and Information scientists and to examine and explore the prevailing thematic patterns in LIS blogs.*

Research question: *What latent thematic structures emerge from a corpus of LIS-related blogs, and to what extent is LDA effective at summarising these latent themes?*

Methodology: *This research is set out to analyse the latent thematic structure of LIS blogs using Latent Dirichlet Allocation. In particular, the research aims to create a representative sample of English-language LIS blog posts from 2019 to 2025, preprocess and normalise unstructured blog text for topic modelling, and use LDA to identify underlying thematic topics.*

Scope: *The study is limited to publicly available, English-language LIS blogs published during the given period of time. The discussion focuses on the recognition of thematic tendencies, not on the evaluation of the quality, the scholarly contribution, or the purpose of the individual posts in the blog. The research paper adopts a macro-level, exploratory approach to provide an overview of professional discourses and emerging issues in the LIS blogging community.*

Findings: *The findings reveal overarching themes in information literacy, digital libraries, user engagement, academic librarianship, and open-access advocacy. The work demonstrates that LDA is a valuable tool for analysing large amounts of unstructured text and provides valuable insights to LIS researchers, practitioners, and educators.*

Limitations: *Nevertheless, the study is limited to some extent. The corpus comprises 50 blog postings, and it might not be a complete sample of all that global blogging about LIS practices entails. The difference in the length of the blog, the style of writing and frequency of posting can affect the stability and interpretability of the topics extracted.*

Keywords: *Topic modelling, LDA, LIS blogs, Machine learning model, Library and Information Science*

Mr. Mithun Das
PhD Scholar
Dept. of Lib. & Inf. Sc.
University of Burdwan
02mithun02@gmail.com

Dr. Rajesh Das
Assistant Professor
Dept. of Lib. & Inf. Sc.
University of Burdwan
rajeshdas99@gmail.com

1. Introduction

The rapid development of web-based communication platforms has had a remarkable impact on changing the process of creating, sharing, and consuming professional knowledge in the profession of Library and Information Science (LIS). Blogs have become one of these spaces and have gained momentum as platforms where professionals (librarians, information professionals, educators, and researchers) share experiences about their profession, discuss emerging issues, and comment on emerging practices. In contrast to classic peer-reviewed journals, LIS blogs provide timely, practice-based, and experiential information that together build the modern professional discourse of the field. Therefore, they are a rich but little-studied source of unstructured textual material on thematic trends and professional interests in LIS.

Nevertheless, the growing size, variety, and informality of blog content create significant difficulties in the systematic thematic analysis of the topic through traditional qualitative methods. Manual content analysis, though informative, is time-consuming and susceptible to researcher bias when used to examine large text corpora. Computational text analysis methods, especially topic modelling, provide an efficient and scalable alternative for uncovering the latent thematic structure of large volumes of unstructured text. Latent Dirichlet Allocation (LDA) is a probabilistic topic modelling algorithm that assumes documents consist of several topics, each described by a distribution of words. On the one hand, using LDA allows one to objectively determine the prevailing themes, tendencies, and patterns of professional concern in LIS blog discourse.

It is against this background that this research has been set out to analyse the latent thematic structure of LIS blogs on the Latent Dirichlet Allocation. In particular, the research aims to create a representative sample of English-language LIS blog posts between the years 2019 and 2025, preprocess and normalise unstructured blog text to analyse it through topic modelling, and use LDA to find underlying thematic topics. Moreover, the paper identifies the most appropriate number of topics using coherence-based assessment measures and examines the dominant and peripheral thematic trends in the LIS blogosphere. The research question that guides the study is as follows: What Latent thematic structures can be identified in a corpus of Library and Information Science blogs, and to what extent does Latent Dirichlet Allocation work well at summarising these themes?

The study is limited to publicly available English-language LIS blogs published during the given period. The discussion focuses on the recognition of thematic tendencies, not on the evaluation of the quality, the scholarly contribution, or the purpose of the individual posts in the blog. The research paper adopts a macro-level, exploratory approach to provide an overview of professional discourses and emerging issues in the LIS blogging community.

Nevertheless, the study is limited to some extent. The corpus comprises 50 blog postings, and it might not be a complete sample of all that global blogging about LIS practices entails. The difference in the length of the blog, the style of writing and frequency of posting can affect the stability and interpretability of the topics extracted. Moreover, as an unsupervised statistical model, LDA primarily relies on word co-occurrence information and may not account for contextual nuances and the underlying semantic connotations of professional discourse.

2. Implications of the Study

The results of this paper have significant implications for several stakeholders in the LIS community. The mentioned themes of information literacy, digital libraries, and professional development may be used by LIS educators to shape the curriculum and instructional focus. The practitioners can apply such findings to ensure that library services and outreach programs are aligned with current issues within the professional community, as discussed in blogs. To scholars, the paper reveals the methodological usefulness of LDA as an effective analytical instrument to study informal scholarly communication and monitor the changes in the trends over and above the conventional journal literature. On a larger scale, the analysis is also a contribution to the increasing number of studies that support the adoption of computational text analysis methods in research in Library and Information Science.

3. Literature review

Koh and Fienup (2021) have proposed the increased significance of academic library chat transcript analysis as a way to improve virtual reference services. However, the unstructured character of these data restricts the application of conventional analysis techniques. Topic modelling has turned out to be an effective method of discovery of shared user requirements and service trends in the chat environment. Comparative studies indicate that even more basic probabilistic models like pLSA can provide a librarian with more interpretable topics in comparison with popular models like LDA. The results also indicate that semi-supervised models with preset instructions are not necessarily better than fully unsupervised ones. In general, the inclusion of entire chat transcripts as opposed to initial user queries enhances the content and utility of extracted topics for service improvement.

P. Li et al. (2019) have conducted a bibliometric research in the domain of Library and Information Sciences (LIS), employing citation data on a large scale to represent intellectual structure, knowledge flow and evolution of LIS research. Longitudinal co-citation studies indicate the emergence of a new LIS perspective, moving from organising information to extracting, and adding informational value--and with more focus on users and networked resources. Information retrieval, information behaviour, social media, bibliometrics, and knowledge management have all been core research topics over the last few decades. Visualisation software such as CiteSpace can effectively identify the evolution of research and thematic clusters. The social media mambo: LIS research following in the steps of social media across subject domains. Recent literature indicates that LIS research will be significantly influenced by breakthrough advances in leading-edge social media.

Maier et al. (2021) have focused on prior research using Moral Foundations Theory and the Model of Intuitive Morality, and Exemplars suggest that moral judgements emerge through shared intuitive principles. However, investigations that hoover up moral content from media texts have generally utilised a traditional content analysis approach of dubious reliability. Recent work suggests that both textual features and differences among individual coders have a substantial effect on agreement in moral coding. There is recent computational and crowd-based work that has been proposed to analyse such a scenario or to minimise its impact on measurement reliability.

Madzík and Falát (2022) wrote a research paper regarding the Analytical Hierarchy Process, which has been studied extensively in research. People have examined a number of papers to see how the Analytical Hierarchy Process has evolved. They used computers to analyse thousands of publications from different fields. This has shown that the Analytical Hierarchy Process is now very important in areas such as helping people make decisions, sustainable risk management, and evaluating systems. By using methods to group similar topics together, researchers can see what the main areas of research are and how they have changed over time. The Analytical Hierarchy Process is still a part of many studies. Temporal trend analysis further illustrates changing research interests and the maturity of AHP applications. Collectively, this literature demonstrates the value of computational reviews in mapping long-term methodological development and future research opportunities.

Tsai (2010) has conducted research on blog mining, showing that we need to change the way we do text mining to deal with the special nature of blog data. Blog data differs from documents because it includes features like tags that provide more information about what the blog is saying. Other studies have examined how to use these tags to build topic models that understand what the blog is about. When we use tags with topic models, we can simultaneously identify topics, keywords, and related tags across blog collections. Tests on blog data show that this approach makes it easier to understand the topics and provides more accurate labels for the blogs. Blog mining is important because it helps us make sense of blog data. Blog data is unique. We need to use specialised techniques, such as tag-integrated topic models, to get the most out of it. This literature underscores the value of tag-aware modelling approaches for more effective analysis of blog-based information sources.

Rodriguez & Storer (2019) discussed the points regarding the increasing difficulty in analysing vast, unstructured data of social media through the use of traditional qualitative analysis. Qualitative thematic analysis has extensively served to address sensitive social concerns on social sites, like Twitter, providing abundant, contextual accounts on user accounts. By comparison, quantitative methods such as topic modelling allow identifiable latent themes in large text collections to be identified with more efficiency. Comparative studies show that the two approaches are inductive and data-driven, and seek to explain underlying social phenomena. Nevertheless, topic modelling leans more towards the use of lexical patterns and qualitative analysis towards meaningful units of meaning. Two tools that support this have been used to check the emotional and cognitive dimensions of themes extracted by either method and compare them. These include LIWC and others. On the whole, the literature indicates that topic modelling combined with qualitative analysis offers a robust mixed-methods approach to the study of social media.

Paul and Dredze (2014) have continued to investigate Twitter as a massive data source to trace the trends and conversations of public health. The topic modelling method has been demonstrated to be effective for discovering latent themes in health-related data, even with large volumes of social media text, with minimal human oversight. Research has shown that unsupervised models can detect various ailments and behavioural health issues, rather than focusing on a single disease. Dominion knowledge-based specialised topic models have facilitated relevance and interpretability in medical text analysis. Results suggest that trends extracted from topics may correlate substantially with law enforcement and survey data, including influenza and allergy reports. Topic models are also superior to simple keyword methods, as they capture more co-occurrences and are more expressive of experiential health. In general, the literature indicates that topic modelling is a useful exploratory tool that is used alongside conventional systems of public health surveillance.

4. Research Question

What latent thematic structures emerge from a corpus of LIS-related blogs, and to what extent is LDA effective at summarising these latent themes?

5. Objectives

The objectives of this study are specific, and they are:

i) To gather a representative sample of blog postings by librarians and Information scientists;

ii) to preprocess and clean unstructured blog text to be used in topic modelling;

iii) to use Latent Dirichlet Allocation (LDA) to determine latent topics;

iv) to identify the potential number of topics based on the coherence test;

v) to examine and explore the prevailing thematic patterns in LIS blogs.

6. LDA technique & Blogs:

Latent Dirichlet Allocation (LDA) is a probabilistic topic modelling algorithm that discovers the implicit thematic organisation of large collections of text. It operates on the assumption that every document is a combination of many topics and that every topic is reflected by a distribution of words. LDA makes use of Dirichlet distributions to estimate the likelihood of document topics and word topics. In processing, we can say that the algorithm will assign words to topics iteratively based on their co-occurrence patterns across documents. The topic-word and document-topic distributions become stable over a series of iterations. LDA is a method of shrinking a huge unstructured text into thematic clusters. It is widely applied to text mining, information retrieval, digital libraries, social media analysis, and bibliometric research (Liao et al., 2025).

Blogs are web-based platforms that publish written content updated periodically, known as blog posts. They operate under a content management system in which authors write, modify, and post online content. All posts are presented in reverse-chronological order, with the latest ones appearing first (Ochsner et al., 2025). Blogs enable readers to interact with each other by leaving comments, liking or sharing on social media. They may carry text information as well as pictures, video and hyperlinks to supplement information. Personal expression, education, dissemination of news and sharing of knowledge are popular methods of using blogs. Business, digital marketing, academic communication, and public engagement are also examples of its usage (Finlay et al., 2013).

7. Methodology

7.1 Data Collection

A sample size of 50 blogs in the field of Librarianship and Information Science was selected and filtered as a corpus, comprising distinguished sources such as professional library association blogs, scholarly LIS blogs, and personal practitioner blogs. The selection criteria were:

*Topicality to themes of LIS.

*Publication between 2019 and 2025

* Public accessibility

*English-language content

The blog materials were mined and structured for further processing and analysis.

No.	Blog Name	Author / Organisation	Country	Focus Area	Website Type
1	Librarian by Day	Emily Clasper	USA	Academic librarianship	Personal
2	Librarianshipwreck	Annie Pho	USA	Critical librarianship	Personal
3	The Library Voice	Walt Crawford	USA	Library trends	Personal
4	Hack Library School	LIS Students	USA	LIS education	Community
5	In the Library with the Lead Pipe	LIS Professionals	USA	Library practice	Group
6	Librarian in Black	Sarah Houghton	USA	Library technology	Personal
7	The Distant Librarian	Michael Stephens	USA	Emerging technologies	Personal
8	Library Stuff	Steven M. Cohen	USA	Library issues	Personal
9	The Ubiquitous Librarian	Brian Mathews	USA	Academic libraries	Personal
10	ACRLog	ACRL	USA	Academic librarianship	Association
11	Librarian Avengers	Jessamyn West	USA	Rural libraries	Personal
12	Library Journal Blog	Library Journal	USA	LIS news	Professional
13	InfoDocket	Gary Price	USA	Information policy	Professional
14	DigitalKoans	David Lee King	USA	Digital libraries	Personal
15	Library Hat	Meredith Farkas	USA	Instruction & technology	Personal
16	Designing Better Libraries	Aaron Schmidt	USA	UX & library design	Personal
17	Librarians Matter	LIS Community	UK	Library advocacy	Community
18	UK Web Archive Blog	British Library	UK	Digital preservation	Institutional
19	Open Shelf	Michael Stephens	USA	Library innovation	Personal
20	Librarian in Training	Michelle Reale	USA	Teaching librarianship	Personal
21	LIS News	LIS News Organisation	USA	LIS updates	Professional
22	Cataloging Futures	Library of Congress	USA	Metadata & cataloguing	Government
23	ArchivesNext	Kate Theimer	USA	Archives	Personal
24	Library Technology Reports Blog	American Library Association	USA	Library technology	Association
25	School Library Journal Blog	SLJ	USA	School libraries	Professional
26	Information Wants To Be Free	Meredith Farkas	USA	Open access	Personal
27	LIS Cafe	LIS Professionals	India	LIS education	Community
28	Librarian's Help	LIS Community	India	Exam preparation	Educational
29	E-Library Science	LIS Scholars	India	Research & theory	Academic
30	Library Automation Blog	Indian LIS Authors	India	Library automation	Educational
31	Digital Library Blog	LIS Professionals	India	Digital libraries	Educational
32	LIS Links	LIS Community	India	Learning resources	Educational
33	Librarian Guide	LIS Educators	India	Library practice	Educational
34	Information Management Today	IMT	India	Information management	Professional
35	Library Philosophy and Practice Blog	University of Nebraska	USA	LIS research	Academic
36	Open Access Directory Blog	OAD	Global	Open access	Community
37	Scholarly Kitchen	SSP	USA	Scholarly publishing	Professional
38	LibFocus	Dublin Business School	Ireland	LIS student research	Academic
39	Library Babel Fish	Steven Bell	USA	Academic libraries	Personal
40	The Research Whisperer	Research Whisperer Team	Australia	Research methods	Academic
41	InfoJustice	InfoJustice Project	USA	Information ethics	Academic
42	Library Futures Blog	Library Futures	USA	Copyright & access	Advocacy
43	Metadata Matters	Metadata Librarians	Global	Metadata	Professional
44	Code4Lib Blog	Code4Lib Community	Global	Library coding	Community
45	Library Trends Blog	Johns Hopkins University	USA	LIS research	Academic
46	Library Juice Press Blog	Library Juice Press	USA	LIS publishing	Professional
47	Library Research Blog	LIS Researchers	Global	Research skills	Academic
48	The Academic Librarian	LIS Faculty	USA	Academic librarianship	Personal
49	Information Today Blog	Information Today Inc.	USA	Information science	Professional

No.	Blog Name	Author / Organisation	Country	Focus Area	Website Type
50	Open Knowledge Blog (Libraries)	Open Knowledge Foundation	Global	Open data & libraries	Community

Table-1: Details of 50 Blogs in Library & Information Science (LIS)

7.2 Data Preprocessing

The following preprocessing steps were used to help the textual data improve the quality and to increase the performance of topic modelling:

Tokenisation: Splitting up text into single words.

Lowercasing: Reducing the size of all texts to lowercase.

Stopword Removal: This removes the most common words that are useful (e.g. the, and, is, etc.)

Lemmatisation: shortening of words to their roots (e.g., libraries to library)

Filtering: Eliminating low-frequency words and highly short blog posts.

The text was processed and converted into a document-term matrix for LDA modelling.

7.3 Topic Modelling and Data Analysis.

Latent Dirichlet Allocation was fitted on the Python Gensim library. Several models were trained with different numbers of topics, ranging from 5 to 15. The evaluation of the model performance was based on:

Coherence Score (c v): Topic semantic interpretability.

Perplexity: The statistical goodness of fit of the model.

The most coherent and readable model was chosen. The qualitative interpretation of results was facilitated by visualising topic distributions and inter-topic relationships with pyLDAvis.

8. Results and Discussion

8.1 Optimal Topic Selection

Out of the tested models, the 10-topic LDA model showed the best relationship between coherence and interpretability and was therefore chosen for analysis.

8.2 Extracted Topics

Topic ID	Interpreted Topic Label	High-Probability Keywords
T1	Information Literacy	understandings, learning, students, abilities, and evaluation.
T2	Digital Libraries	access, collections, metadata, repository, digital.
T3	User Engagement & Outreach	community, event, users, engagement, outreach.
T4	Academic Librarianship	faculty, research, support, campus, teaching.
T5	Measuring the Impact of Scholarly Open-Access Publication: Retaining the Full Picture.	access, open, rights, publishing, article.
T6	The library technology and innovation are arranged based on the geographical region the library occupies.	technology, tools, AI, systems, software.
T7	Standards of Cataloguing and Metadata.	records, metadata, schema, cataloguing, MARC.
T8	Professional Development	training, education, webinar, skills, career.
T9	Diversity, Equity and Inclusion (DEI).	diversity, inclusion, equity, culture, training.
T10	Management Leadership Library.	management, strategy, leadership, planning, and budget.

Table 2: Topics in the LIS Blog Corpus were extracted using LDA

Sl. No.	Topic Label	Approximate Proportion (%)
1	Information Literacy	16.8
2	Digital Libraries	14.5
3	Academic Librarianship	12.9

Sl. No.	Topic Label	Approximate Proportion (%)
4	Open Access & Academic OpenKeep	11.4
5	User Engagement & Outreach	10.6
6	Technology and Innovation in the Library	9.2
7	Professional Development	8.1
8	Diversity, Equity, and Inclusion (DEI)	7.3
9	Management and Leadership in the Library	5.7
10	Cataloging & Metadata	3.5

Table 3: Prevalence of Topics in the Corpus of LIS Blog

Sl. No.	Topic Label	Approximate Proportion (%)
1	Information Literacy	16.8
2	Digital Libraries	14.5
3	Academic Librarianship	12.9
4	Open Access & Academic OpenKeep	11.4
5	User Engagement & Outreach	10.6
6	Technology and Innovation in the Library	9.2
7	Professional Development	8.1
8	Diversity, Equity, and Inclusion (DEI)	7.3
9	Management and Leadership in the Library	5.7
10	Cataloging & Metadata	3.5

Table 3: Prevalence of Topics in the Corpus of LIS Blog

The prevalence analysis of the topic shows that the leading theme in LIS blogs is Information Literacy, which reveals that the profession places a lot of stress on the development of user research and critical thinking skills. The high prevalence of Digital Libraries and Academic Librarianship is also an indicator of current changes in digital infrastructure and academic support services. The extensive nature of Open Access and Scholarly Communication highlights increased support in favour of fair use of scholarly products. The themes of User Engagement, Technology and Innovation, and Professional Development reflect the interest of librarians in changing their services and expertise based on the needs of the users. However, Cataloguing and Metadata seem less common and perhaps reflect the trend toward more strategic, user-focused issues in blog discourse that supersedes more conventional technical procedures.

8.3 Topic Distribution Analysis

It was analysed that the most common topics in the corpus were Information Literacy and Digital Libraries. The problems associated with Open Access were often comorbid with Academic Librarianship, meaning that there was a lot of thematic overlap in professional discussions.

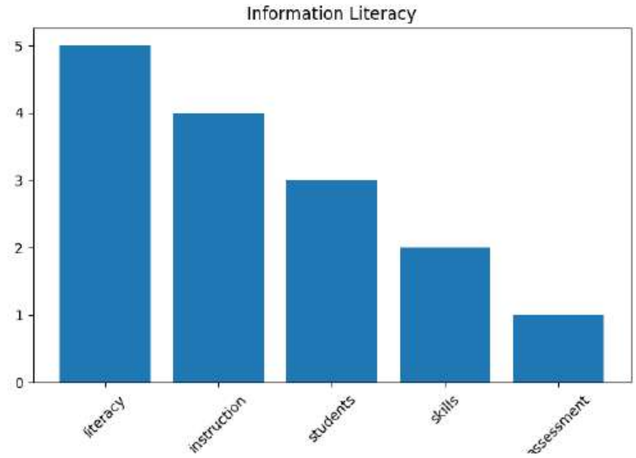


Figure 1: Frequency of the words and subject of Information Literacy

This figure shows the most representative keywords concerning the issue of Information Literacy, as it brings into light the ideas which are associated with the teaching, learning of students and the formation of the research skills within the LIS blogosphere.

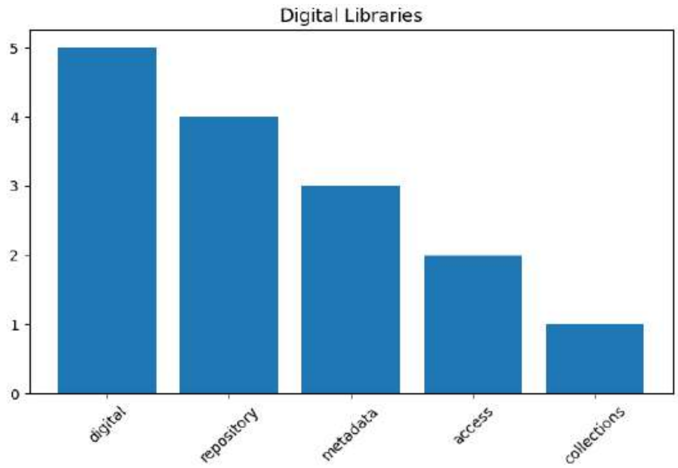


Figure 2: Thematic distribution of the word digital libraries

The bar chart represents the most common terms and words regarded in the repositories, metadata, collections, and discussions of access provided in digital libraries in the material of LIS blogs.

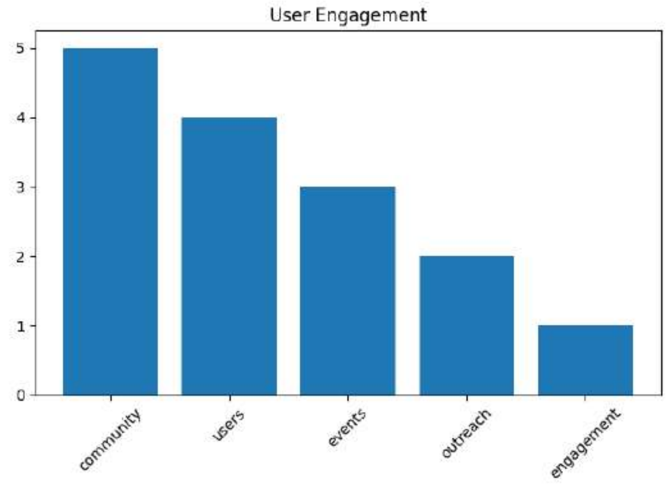


Figure 3: Topic words distribution in the theme

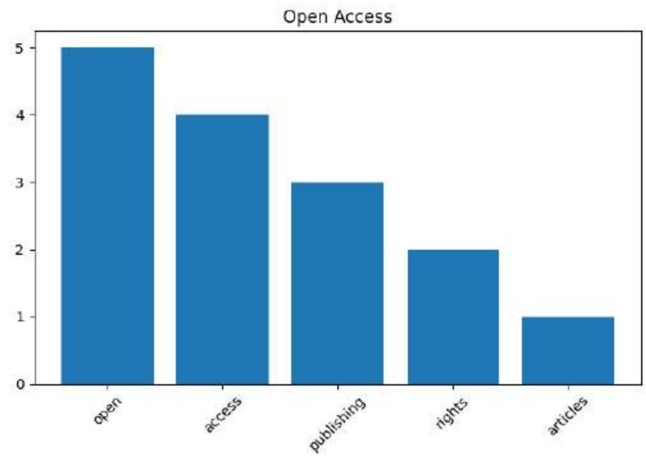


Figure 4: Distribution of topic words in the theme of Open Access and Scholarly Communication

The visualisation displays those terms that denote open-access publishing, intellectual property rights, and other scholarly communication practices that have been widespread in LIS discourse.

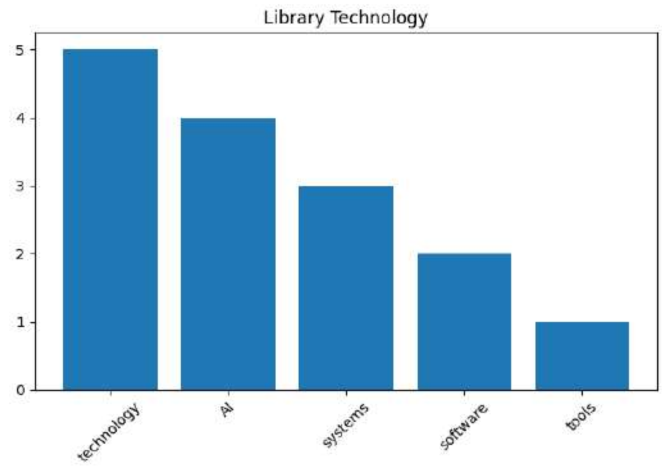


Figure 5: Distribution of topic words of the theme of Library Technology and Innovation

This number reflects the significant technological terminology used, such as digital tools, artificial intelligence, and library systems, which reflects the increased amount of attention paid to technological innovation in LIS blogs.

91. Results and Discussion

91.1 Thematic Insights

The prominence of Information Literacy continuously shares the professional effort to teach students important skills in a time of disinformation. The strong role played by Digital Libraries indicates that there is a high level of interest in digitisation, repositories and online access. Topics related to user engagement reflect the move towards library services that are community-centred. Moreover, Diversity, Equity and Inclusion were identified as important themes, having become part of the overall institutional and professional priorities in LIS.

91.2 Practical Implications

- **LIS educators:** Curriculum development can be harmonised with the predominant performances, such as digital library management and literacy of information. Insights into the prevalent issues

  can be of help to practitioners in framing their services and professional development opportunities.

- **For Researchers:** Topic modelling is more of a versatile approach to the analysis of shifts in a professional discourse.

92. Limitations

This study is constrained by the dataset of only 50 blog posts, which may not give a complete representation of LIS blogging styles. It is also possible that differences in writing style and the length of blogs might have an influence on topic stability.

93. Findings

Latent Dirichlet Allocation identified ten topics of the LIS blog corpus that are thematic. Among them, the strongest theme was Information Literacy, which suggests the profession's consistent interest in educating users, teaching critical thinking, and developing research skills. The use of Digital Libraries and Academic Librarianship was also extremely widespread, as there were some changes in digital infrastructure and academic support services. Articles associated with Open Access and Scholarly Communication put forward powerful support in relation to fair access to scholarly information. Also, there were other themes like the

User Engagement and Outreach, Library Technology and Innovation and the Professional Development, which highlighted how the profession responds to the change in technology and the dynamic needs of its users. On the other hand, cataloguing and metadata were not so high, which implied a relative change in the blog discourse towards strategic, user-based and policy-based issues.

94. Recommendations and suggestions

According to the results, there are recommendations to endorse the use of blog-based discourse as a supplementary source of professional intelligence, which LIS professionals and institutions should actively monitor. Topic modelling is also urged to be complemented by qualitative methods to improve the interpretability and contextual interpretation. Future research might involve using a larger corpus and multilingual blogs to enhance representativeness. Also, further exploration of advanced topic modelling methods, including neural or dynamic topic models, could provide more information on temporal and thematic change in the LIS discourse.

95. Further Scope of the Study

Future studies can build on the existing study, including larger, longitudinal datasets to analyse thematic change over time. Further comparative studies between the content of blogs and peer-reviewed academic materials may also serve to shed light on the difference between informal and formal professional communication. Furthermore, a combination of sentiment analysis or author-network analysis with topic modelling could bring more detailed information about the LIS online discourse dynamics.

96. Conclusion

The study does this to show that Latent Dirichlet Allocation is the best tool for capturing latent semantic structures in unstructured LIS blog content. The findings point out principal areas of professional activities, such as information literacy, digital libraries, user engagement, and open access advocacy. Through the use of topic modelling techniques, one can get useful hints about current trends in the formation of the LIS profession. Future research could enlarge the dataset, involve the temporal analysis, or compare LDA with an advanced neural topic model system. This number reflects the significant technological terminology used, such as digital tools, artificial intelligence, and library systems, which reflects the increased amount of attention paid to technological innovation in LIS blogs.

References:

Ahmed, F. (2025). Addressing the interpretability of big data for social sciences: a contextualization approach to LDA topic models. Quality & Quantity. https://doi.org/10.1007/s11135-025-02491-2

Airoldi, E. M., & Bischof, J. M. (2015). Improving and evaluating topic models and other models of text. Journal of the American Statistical Association, 111(516), 1381–1403. https://doi.org/10.1080/01621459.2015.1051182

Chaney, A., & Blei, D. (2021). Visualizing topic models. Proceedings of the International AAAI Conference on Web and Social Media, 6(1), 419–422. https://doi.org/10.1609/icwsm.v6i1.14321

Eshima, S., Imai, K., & Sasaki, T. (2023). Keyword-Assisted topic models. American Journal of Political Science, 68(2), 730–750. https://doi.org/10.1111/ajps.12779

Finlay, S. C., Hank, C., Sugimoto, C. R., & Johnson, M. (2013). The Structure of the Biblioblogosphere: An examination of the linking practices of institutional and personal library blogs. Journal of Web Librarianship, 7(1), 20–36. https://doi.org/10.1080/19322909.2012.714305

Hu, Y., Boyd-Graber, J., Satinoff, B., & Smith, A. (2013). Interactive topic modelling. Machine Learning, 95(3), 423–469. https://doi.org/10.1007/s10994-013-5413-0

Koh, H., & Fienup, M. (2021). Topic modelling as a tool for analysing library chat transcripts. Information Technology and Libraries, 40(3). https://doi.org/10.6017/ital.v40i3.13333

Li, P., Yang, G., & Wang, C. (2019). Visual topical analysis of library and information science. Scientometrics, 121(3), 1753–1791. https://doi.org/10.1007/s11192-019-03239-0

Liao, L., Wu, Y., Xiong, Z., Wen, Y., & Yang, J. (2025). Insights into construction workers' risk-taking behaviors based on LDA topic modelling. Engineering Construction &

Architectural Management, 1–31. https://doi.org/10.1108/ecam-12-2024-1654

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., & Adam, S. (2021). Applying LDA topic modelling in communication research: Toward a valid and reliable methodology (1st ed). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003082606-2/

Madzík, P., & Falát, L. (2022). State-of-the-art on the analytic hierarchy process in the last 40 years: Literature review based on Latent Dirichlet Allocation topic modelling. PLoS ONE, 17(5) e0268777. https://doi.org/10.1371/journal.pone.0268777

Ochsner, C., Pampel, H., Höfting, J., & Rothfritz, L. (2025). Scholarly blogs: an analysis of infrastructural aspects based on German scholarly blogs. Journal of Documentation, 81(7), 520–544. https://doi.org/10.1108/jd-02-2025-0053

Paul, M. J., & Dredze, M. (2014). Discovering health topics in social media using topic models. *PLoS ONE*, 9(8), e103408. https://doi.org/10.1371/journal.pone.0103408

Rodriguez, M. Y., & Storer, H. (2019). A computational social science perspective on qualitative data exploration: Using topic models for the descriptive analysis of social media data. Journal 8835.2019.1616350

Tsai, F. S. (2010). A tag-topic model for blog mining. Journal of Technology in Human Services, 38(1), 54–86. https://doi.org/10.1080/15228835.2019.1616350

Tsai, F. S. (2010). A tag-topic model for blog mining. Expert Systems with Applications, 38(5) 5330–5335. https://doi.org/10.1016/j.eswa.2010.10.025

About Authors

Sri Mithun Das *(http://orcid.org/0009-0007-6482-6425)*, currently pursuing a PhD (Junior Research Fellow) in DLIS from The University of Burdwan. He has completed B. A in History (Hons), an M.A. in History. He has been awarded an MLIS Degree in DLIS from the University of Burdwan. He has also completed an MPhil in DLIS from the University of Burdwan. He has published many articles in national and international journals and seminars. He has also qualified for WBSET in English and History. Previously, he has worked as a Librarian at Bengal Institute for Educational Studies. His research areas include Information Seeking Behaviour, Machine learning, Linked Open Data, and topic models, etc.

Dr. Rajesh Das *(https://orcid.org/0000-0003-3137-864X)* PhD, MLISc, MCA, RHCE (Redhat Certified Engineer) is currently working as Assistant Professor at Department of Library and Information Science, The University of Burdwan, Burdwan. He had worked at many reputed institutions. Previously, he served as Librarian at the West Bengal Taxation Tribunal, Kolkata; as Junior Computer Engineer at SNBNCBS, Kolkata; and as Archivist at Media Lab, Jadavpur University. He has published more than 30 articles in national and international journals and in seminar and conference volumes. He has delivered many invited lectures as a resource person at national-level seminars and workshops. Presently, he supervises four PhD scholars and one M.Phil scholar at the DLIS, the University of Burdwan. His areas of interest are ICT applications in library and information science domain, open-source technologies, Redhat / Debian Linux administration and networking, semantic web technologies, data science and machine learning.

Blind Peer Reviewed
(Previously UGC-CARE)
ISSN: 0972-2750

Manuscript Received on –
27th December, 2025
Primary Reviewed on –
4th January, 2026
Peer Reviewed on –
04th February, 2026
Revised on –
07th February, 2026
Accepted for publication
15th February, 2026

Information Retrieval Systems: A Patent Driven Analysis of Innovation Trends

Goutam Dinda

ABSTRACT

Introduction: *Information Retrieval (IR) systems play a crucial role in organizing, accessing, and managing digital information in libraries, research institutions, and enterprises. The rapid advancements in AI, machine learning, and semantic search have significantly transformed IR technologies over the past two decades.*

Purpose: *This study aims to analyze global patent trends in IR systems to understand technological innovations and guide strategic adoption in libraries and R&D.*

Research Problem: *Despite the proliferation of IR technologies, there is a limited understanding of patent landscapes, emerging innovations, and their implications for technology planning and knowledge management.*

Objectives: *The study seeks to (i) map IR system patents globally, (ii) identify dominant Players and patent classifications, (iii) analyze citation impact and technological maturity, and (iv) provide insights for libraries, librarians, and R&D stakeholders.*

Methodology: *A bibliometric and patent analytics approach was employed using Lens.org, extracting patent data (3201) on IR systems from 2000 to 2024. Data visualization and analysis were performed using Excel and Lens dashboards to identify trends, leading contributors, and technological patterns.*

Findings: *The results indicate rapid growth in AI-driven and semantic-retrieval patents, with significant contributions from both the corporate and academic sectors. Emerging innovations focus on context-aware search, natural language processing, and intelligent indexing, providing actionable guidance for libraries and research institutions to plan technology adoption, improve user services, and foster innovation in information retrieval.*

Keywords*: Patent Analytics; Information Retrieval Systems; Semantic Retrieval; Technological Innovation; Knowledge Management*

Mr. Goutam Dinda
Research Scholar
Dept. of Lib. & Inf. Sc.
Rabindra Bharati University
dindagoutam60@gmail.com

1. Introduction

Information Retrieval (IR) Systems are central to managing and accessing the exponentially growing volume of digital information. These systems have evolved rapidly due to both incremental improvements and disruptive innovations, shaping how data is stored, indexed, and retrieved across digital platforms (Dewulf & Childs, 2021). Understanding the dynamics of innovation in IR requires systematic methods, and patent analysis has emerged as a crucial approach for mapping technological trends, identifying emerging research directions, and assessing commercialisation potential (Mejia & Kajikawa, 2025). By examining patent data, researchers and organisations can gain insights into technological trajectories, competitive landscapes, and strategic opportunities for research and development in the IR domain (El Boukhari, Matin, Atanasov, & Fatimi, 2025; Chung, Kwon, & Kim, 2023). Previous studies have demonstrated the utility of patent analytics for forecasting emerging technologies, understanding technology life cycles, and informing strategic decision-making for startups and established organisations alike (Moro Visconti, 2025; Abbas, Zhang, & Khan, 2014). Text mining, visualisation, and topic-driven patent analysis techniques have been applied to detect patterns of innovation, inventor contributions, and organisational performance (Tseng, Lin, & Lin, 2007; Abraham & Moitra, 2001; Kim, Suh, & Park, 2008). Patent mapping provides both macro-level insights into industry trends and micro-level analyses of individual or organisational innovation activities, supporting informed R&D investment and intellectual property management decisions (Sampaio et al., 2018; Breitzman & Mogee, 2002; Kim & Bae, 2017). Furthermore, patent-driven approaches such as topic mining and visualisation facilitate the detection of knowledge gaps, innovation clusters, and high-impact inventions, enabling evidence-based strategic planning in technology-intensive domains like IR (Hicks, Breitzman, Olivastro, & Hamilton, 2001; Tang et al., 2012). Despite significant advancements, challenges remain in integrating artificial intelligence, optimising retrieval algorithms, and addressing domain-specific requirements. A patent-driven study provides a structured, evidence-based understanding of global innovation patterns, technological concentration, and emerging opportunities, offering valuable guidance to researchers, technology developers, and libraries seeking to leverage IR innovations effectively (Dewulf & Childs, 2021; Mejia & Kajikawa, 2025).

2. Review of Literature

Patent analysis has become an indispensable tool for understanding technological innovation, mapping competitive landscapes, and guiding strategic decision-making across multiple sectors. Early studies emphasised the use of patent data for strategic planning in technology development, providing organisations with the ability to evaluate technological trajectories and prioritise R&D investments (Liu & Shyu, 1997). Building on this foundation, bibliometric and patent-based approaches were applied to forecast emerging technologies, offering a structured methodology to anticipate technological breakthroughs and commercialisation potential (Daim, Rueda, Martin, & Gerdsri, 2006). Recent advancements in computational methods have expanded the scope and precision of patent analytics. Shomee, Wang, Ravi, and Medya (2025) highlighted the integration of natural language processing (NLP) and multimodal artificial intelligence (AI) in patent analysis, enabling deeper insights from the textual, visual, and metadata components of patent documents. Similarly, Krestel, Chikkamath, Hewel, and Risch (2021) reviewed the application of deep learning techniques in patent datasets, demonstrating their utility in identifying innovation clusters and predicting technology trajectories. The extraction of significant yet rare keywords has been shown to provide additional granularity in understanding technological niches and emerging trends (Li, Wang, & Hong, 2009). Empirical studies across industries illustrate the application of patent analysis to real-world innovation. For example, Jun and Sung Park (2013) examined Apple's technological innovation strategies using patent data, while Chang (2012) proposed patent-based approaches to establish an organisation's technological position. Sector-specific studies have utilised patent analysis to track low-carbon energy technologies (Albino, Ardito, Dangelico, & Petruzzelli, 2014), industrial wastewater treatment innovations (Mao et al., 2022), and green hydrogen technologies (Yang et al., 2023), highlighting the utility of patents in monitoring sustainable and emerging technological developments. Longitudinal analyses have further demonstrated the capacity of patent analytics to reveal trends across countries, institutions, and technological domains, such as in nanoscale science and engineering (Huang et al., 2003) and renewable energy (Jiang, Zou, Qiao, & Huang, 2022). In addition, semantic patent analysis has been applied to evaluate patent infringement risks and support sustainable technology management strategies (Bergmann et al., 2008; Choi, Jun, & Park, 2016). Organisational and technological factors influencing green innovation have also been investigated, demonstrating how patent characteristics reflect firm capabilities and market positioning (Messeni Petruzzelli, Dangelico, Rotolo, & Albino, 2011). Finally, methodological studies have traced evolutionary pathways of technologies, revealing the chronological development and knowledge diffusion of innovations across multiple domains (Zhou, Zhang, Porter, Guo, & Zhu, 2014). Overall, this body of literature underscores the growing importance of patent analytics as a robust tool for innovation intelligence. The integration of AI, bibliometrics, and semantic analysis enables organisations, policymakers, and researchers to systematically monitor technological trends, assess competitive positioning, and make informed strategic decisions in a variety of high-technology sectors.

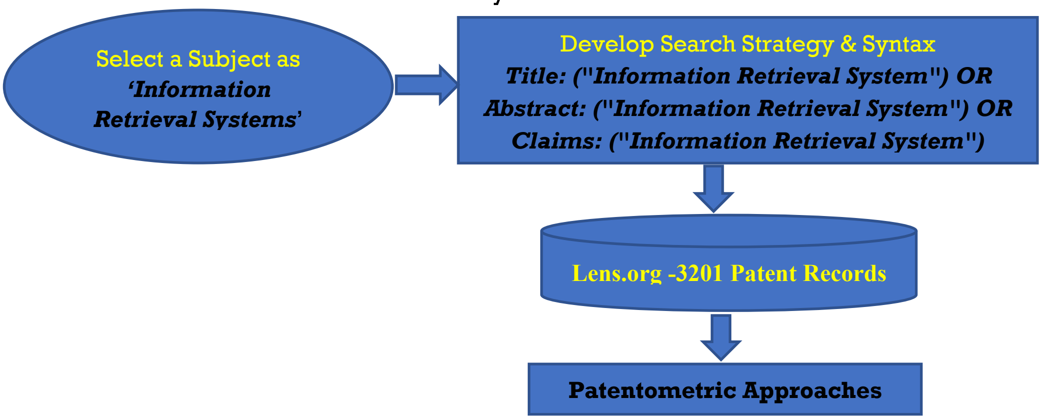
3. Objectives of the Study

1. To analyse global patent trends in Information Retrieval (IR) systems from 2000 to 2024, highlighting patterns in innovation, filing activity, and citation behavior;
2. To identify leading patent owners, inventors, and jurisdictions, thereby providing insights into key contributors and competitive dynamics within the IR domain;
3. To examine technological focus areas through CPC, IPCR, and US classification codes, revealing both emerging and mature innovations in IR systems;
4. To assess patent lifecycle stages, legal status, and citation impact in order to understand technological maturity, knowledge diffusion, and commercialisation potential;
5. To provide actionable insights for libraries, librarians, and R&D professionals to support strategic decision-making, technology adoption, and intellectual property management in the Information Retrieval domain.

4. Methodology (Abraham & Moitra, 2001)

5.1 Research Design, Data Source, and Search Strategy (Framework 1)

The present study adopts a **quantitative patent analytics research design** to examine global innovation trends in **Information Retrieval (IR) systems.** Patent analysis is employed as an objective indicator of technological development, innovation intensity, and knowledge diffusion in the IR domain. Patent data were collected from **Lens.org**, a comprehensive and globally recognised patent database that integrates patent records from major patent offices worldwide. The database enables advanced search, filtering, citation analysis, and visualisation of patent-related metadata. The patent search was conducted using the following structured search syntax: ***Title:*** *("Information Retrieval System")* ***OR Abstract:*** *("Information Retrieval System")* ***OR Claims:*** *("Information Retrieval System").* The search was restricted to **patent documents published between 2000 and 2024.** After applying the time filter and removing irrelevant records, a total of **3,201 patent documents** were retrieved for analysis.



Framework 1: Research Design

5.2 Data Visualisation

To enhance interpretability, **graphs, tables, and charts** were used to visualise trends such as annual patent growth, top applicants, jurisdictional distribution, classification patterns, and legal status. Visualisations were generated using Microsoft Excel and Microsoft Word, while Lens.org's built-in visualisation tools supported interactive patent analytics.

5.3 Ethical Considerations

All data used in this study were obtained from **publicly available patent documents.** No confidential or proprietary information was accessed, and proper citation of data sources was maintained throughout the research process.

6. Results and Discussion

6.1 Patent Trends & Types

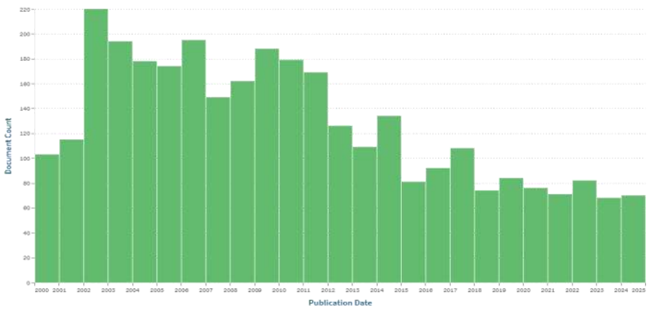


Figure 1: Patents in Information Retrieval (IR) Systems (2000-2024)

Figure 1 illustrates the year-wise distribution of patents in Information Retrieval (IR) Systems from 2000 to 2024, highlighting the changing dynamics of innovation. The Initial phase (2000–2002) shows rapid growth, with patent counts increasing from 103 (2000) and 115 (2001) to a peak of 220 (2002), indicating intensive research and development during the early expansion of web search engines and digital information systems. The subsequent period (2003–2011) reflects relative stability, with moderate fluctuations ranging from 149 (2007) to 195 (2006), suggesting consolidation and incremental advancements in retrieval models, indexing techniques, and relevance ranking. A clear declining trend emerges after 2012, with patents decreasing from 126 (2012) and 109 (2013) to notably lower levels, such as 81 (2015) and 74 (2018). The recent phase (2019–2024) has maintained subdued yet consistent activity, fluctuating between 68 (2023) and 84 (2019), reflecting the integration of IR functionalities into broader AI-driven, data-intensive platforms rather than standalone patenting.

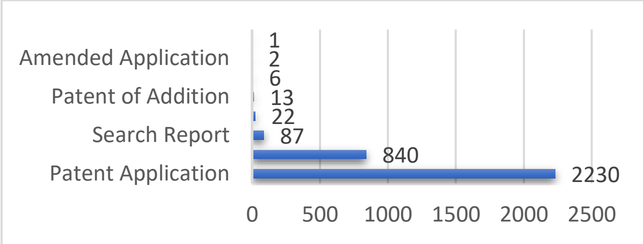


Figure 2: Types of Patent

Figure 2 shows that patenting in Information Retrieval (IR) Systems is dominated by Patent Applications (2230), indicating intense inventive activity and strong protection-seeking behaviour. Granted Patents (840) form a substantial share, reflecting the successful validation and technological maturity of many IR innovations. Search Reports (87) highlight active examination and prior-art assessment within the domain. Limited Figure 2 shows that patenting in Information Retrieval (IR) Systems is dominated by Patent Applications (2230), indicating intense inventive activity and strong protection-seeking behaviour. Granted Patents (840) form a substantial share, reflecting the successful validation and technological maturity of many IR innovations. Search Reports (87) highlight active examination and prior-art assessment within the domain. Limited

6.2 Patent Classification

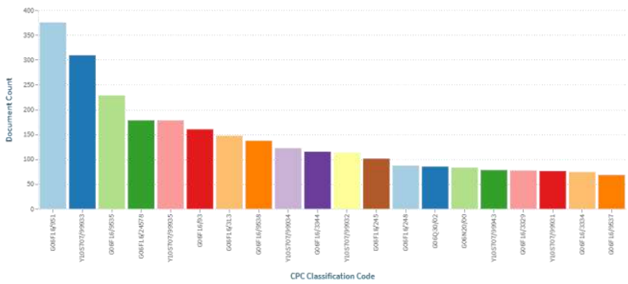


Figure 3: Top CPC classification codes

Figure 3 illustrates the technological concentration of Information Retrieval (IR) system patents across major CPC classification codes. The highest share is observed in G06F16/951 (375), followed by Y10S707/99933 (309) and G06F16/9535 (228), highlighting a strong focus on digital information retrieval, indexing, and database search mechanisms. A significant presence of mid-range classes such as G06F16/24578 and Y10S707/99935 (178 each) and G06F16/93 (160) reflects advancements in structured data retrieval and query processing. Other notable categories, including G06F16/313 (147) and G06F16/9538 (137), indicate optimisation of retrieval architectures and relevance ranking. The inclusion of G06Q30/02 (85) suggests applications in ecommerce and user-oriented retrieval, while G06N20/00 (83) points to growing integration of AI and machine learning. Overall, the distribution underscores a shift toward intelligent, data-driven, and application-specific IR innovations.

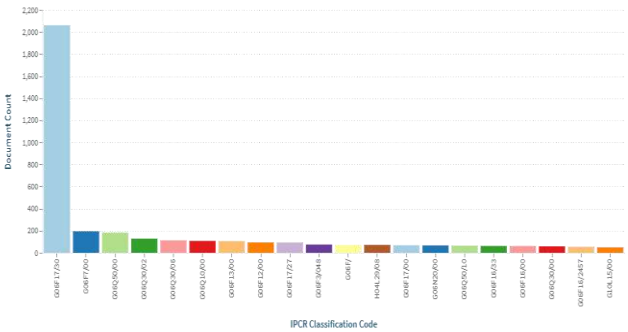


Figure 4: Top IPCR classification codes

Figure 4 presents the **Top IPCR Classification Codes** associated with patents in Information Retrieval (IR) Systems, highlighting the technological core of innovation in this domain. The most dominant class, **G06F17/30 (2061 documents)**, represents digital computing arrangements for information retrieval and database structures, underscoring its foundational role. Other significant codes, such as G06F7/00 (195) and G06Q50/00 (183), indicate strong involvement of data processing and sector-specific information systems. Business and commercial applications are evident through **G06Q30/02 (128), G06Q30/06 (113)**, and **G06Q10/00 (108).** Infrastructure-level support appears in **G06F13/00 (106)** and **G06F12/00 (94).** The inclusion of **G06N20/00 (67)** reflects the growing integration of artificial intelligence, while **H04L29/08 (72)** highlights network-based retrieval, confirming the multidimensional evolution of IR technologies.

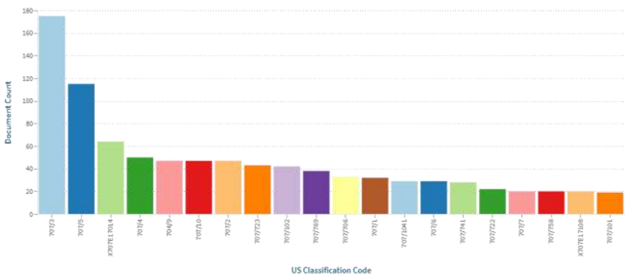


Figure 5: Top **US Classification Codes**

Figure 5 presents the distribution of Information Retrieval (IR) system patents across major **US Classification Codes**, reflecting the structural focus of innovation in this field. The most prominent class, **707/3 (175 documents)**, highlights core database searching and retrieval mechanisms. This is followed by **707/5 (115)** and **X707E17014 (64)**, indicating advancements in data organisation, indexing, and enhanced retrieval processes. Several closely related classes **707/4 (50), 704/9 (47)**, **707/10 (47),** and **707/2 (47)** demonstrate diversification into language processing, query handling, and system-level retrieval functions. Specialised subclasses such as **707/723 (43), 707/102 (42)**, and **707/769 (38)** reflect refinements in structured search and database management. The presence of multiple lower-frequency yet related codes underscores incremental and application-specific innovations, confirming the depth and maturity of IR technologies within the US patent classification framework.

6.3 Patent Jurisdiction

Figure 6 illustrates the global jurisdictional distribution of patents related to Information Retrieval (IR) Systems, highlighting significant geographical concentration. The **United States** leads overwhelmingly with **1418 documents**, underscoring its dominance in IR research, commercial deployment, and intellectual property protection. **Japan (680)** and **China (295)** follow, reflecting strong innovation ecosystems and sustained investments in computing and information technologies. International filings through **WO–WIPO (268)** and **European Patents (203)** indicate applicants' strategic intent to secure broader market protection. Moderate activity in **Korea (88)**, **Canada (78), Australia (65),** and the **United Kingdom (40)** suggests regional innovation hubs with focused contributions. Other jurisdictions, such as Taiwan (22) and Russia (14), show limited yet notable participation, while the remaining countries contribute marginal shares. Overall, the distribution reveals a highly centralised global innovation landscape, dominated by technologically advanced economies with strategic international patenting behaviour.

6.4 Patent Stakeholders

Figure 7: Leading Patent Applicants

Figure 7 highlights the **leading applicants** driving innovation in Information Retrieval (IR) Systems, revealing strong dominance by global technology corporations. **Google Inc. (186 documents)** emerges as the top applicant, reflecting its core business focus on search, indexing, and large-scale data retrieval technologies. This is followed by **IBM (122)** and **Microsoft Corp. (109),** indicating sustained investments in enterprise search, data analytics, and intelligent information systems. Japanese firms such as **NEC Corp. (60), Nippon Telegraph and Telephone (36), Sony Corp. (36), Hitachi Ltd. (35), Ricoh KK (32), Mitsubishi Electric Corp. (29), Fujitsu Ltd. (24),** and **Toshiba Corp. (24)** demonstrate strong participation from Asia, highlighting industrial and applied research strengths. The presence of multiple Google and Microsoft entities reflects strategic portfolio structuring. At the same time, companies like **Xerox, Thomson Reuters**, and **Yahoo** indicate diversification of IR applications across enterprise, media, and knowledge services.

Figure 8: Leading Patent Owners

Figure 8 presents the **leading patent owners** in the domain of Information Retrieval (IR) Systems, highlighting the concentration of ownership among major technology and information service corporations. **Google LLC (140 patents)** ranks first, reflecting its strong proprietary control over search and retrieval technologies. This is closely followed by **Microsoft Technology Licensing LLC (134)** and **Microsoft Corporation (89),** indicating Microsoft's strategic use of multiple entities to manage and protect its IR-related intellectual property. **International Business Machines Corporation (108)** also holds a significant portfolio, emphasising its long-standing role in enterprise information systems and data analytics. The dual presence of **Thomson Reuters** entities (56 and 55) demonstrates focused investment in knowledge management and legal information retrieval solutions. Other notable owners, such as West Services Inc. (44), Oracle entities (34 and 13), Yahoo! Inc. (30), and Xerox Corporation (18), reflect diverse application areas. Overall, the ownership pattern reveals strong corporate dominance and strategic consolidation of IR innovations.

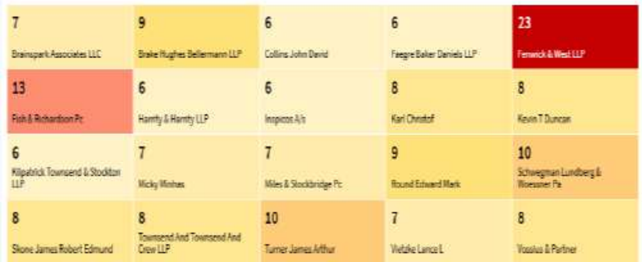


Figure 9: Leading Attorneys & Agents

Figure 9 highlights the **leading attorneys and agents** involved in prosecuting Information Retrieval (IR) System patents, reflecting the legal expertise supporting innovation in this field. **Fenwick & West LLP (23 documents)** emerges as the most active firm, indicating its strong engagement with high-technology and software-related patent filings. This is followed by **Fish & Richardson PC (13)** and **Schwegman Lundberg & Woessner PA (10),** both known for their specialisation in intellectual property for computing and digital technologies. Individual practitioners such as **Turner James Arthur (10), Round Edward Mark (9)**, and **Kevin T. Duncan (8)** also play notable roles, highlighting the contribution of experienced patent professionals. Firms like **Townsend and Townsend and Crew LLP, Vossius & Partner**, and **Kilpatrick Townsend & Stockton LLP** indicate cross-jurisdictional legal involvement. Overall, the distribution reflects a competitive, specialised patent prosecution environment that underpins the advancement of IR technologies.

Figure 10: Leading Inventors

Figure 10 presents the **leading inventors** contributing to innovation in Information Retrieval (IR) Systems, reflecting individual-level research productivity. **Patterson, Anna L. (75 patents),** stands out as the most prolific inventor, indicating sustained and influential contributions to IR-related technologies. This is followed by **Silverbrook, Kia (31)**, and Al-Kofahi, Khalid (27), who highlight their active roles in advancing retrieval algorithms, knowledge systems, and applied information technologies. Inventors such as **Battle, Alexis J. (26), Warner, Douglas K. (26)**, and **Bailey, David R. (25)** further demonstrate consistent inventive output. Mid-tier contributors, including Zelevinsky, Vladimir (23), and Bauer, Walter (21), reflect collaborative, incremental innovation. The appearance of closely related name variants (e.g., *Patterson Anna L* and *Patterson Anna Lynn*) suggests portfolio dispersion across filings. Overall, the inventor distribution reveals a mix of highly prolific individuals and a broad base of contributors shaping the evolution of IR systems.

6.5 Legal Status of Patents

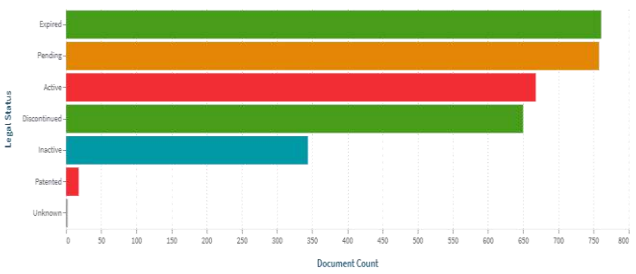


Figure 11: Legal Status of the Patents

Figure 11 illustrates the **legal status distribution** of patents related to Information Retrieval (IR) Systems, reflecting the life-cycle dynamics of innovation in this domain. A large proportion of patents are **Expired (761),** indicating early-stage innovations that have completed their statutory protection period, often corresponding to foundational IR technologies. **Pending patents (758)** form an almost equal share, highlighting sustained research activity and ongoing evaluation of new IR solutions. **Active patents (668)** represent currently enforceable intellectual property, underscoring the continued commercial and technological relevance of IR systems. The notable presence of **Discontinued (650)** and **Inactive (344)** patents suggests strategic abandonment or obsolescence due to rapid technological change and market shifts. Only a small number of patents are categorised as **Patented (18),** reflecting strict grant criteria or dataset classification limits, while **Unknown (2)** documents are negligible.

Overall, the distribution indicates a mature yet continuously evolving innovation ecosystem in IR technologies.

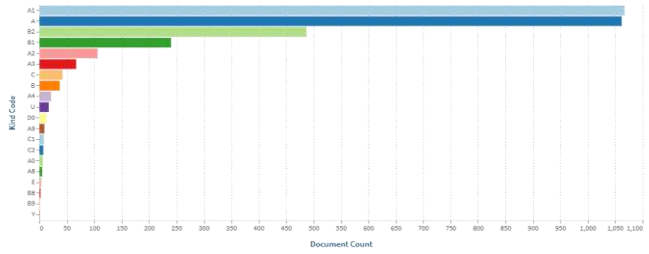


Figure 12: Top Kind Code

Figure 12 depicts the distribution of **Top Kind Codes** for patents related to Information Retrieval (IR) Systems, providing insights into publication and examination stages. **A1 (1067)** and **A (1062)** dominate the dataset, indicating a high volume of published patent applications and early-stage disclosures, which reflects strong innovation activity and frequent knowledge diffusion in IR technologies. Granted patent categories, **B2 (487)** and **B1 (240),** demonstrate that a substantial portion of applications progressed to successful grants, confirming technical validity and commercial relevance. Secondary publication stages such as **A2 (106)** and **A3 (67)** suggest continued amendments and procedural updates during examination. Smaller counts across codes like **C (42), B (37),** and **A4 (21)** reflect jurisdiction-specific or specialised documentation. The minimal presence of utility, design, and correction codes (**U, D0, A9, C1, C2, Y**) indicates limited activity beyond core patent filings. Overall, the kind code distribution highlights an innovation landscape dominated by application-stage activity with selective but significant grant outcomes.

6.6 Patent Citations

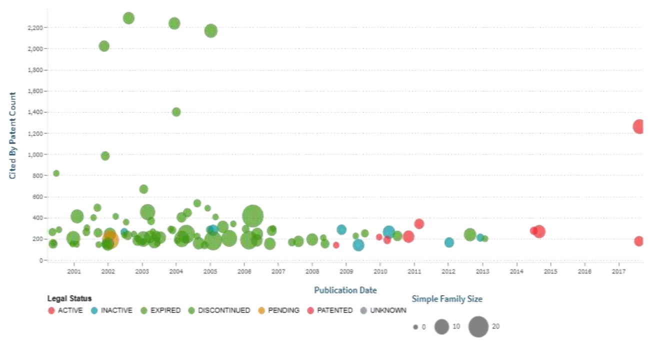


Figure 13: Top Cited Patent

Figure 13 presents the **top-cited patents** in Information Retrieval (IR) Systems, reflecting their technological influence and knowledge diffusion within the patent ecosystem. The most highly cited patents, published between **2000 and 2005,** receive exceptionally high citation counts, led by patents dated **8/13/2002 (2286 citations), 12/16/2003 (2236)**, and **1/11/2005 (2166),** indicating their foundational role in shaping modern IR architectures. The predominance of an **EXPIRED legal status** among highly cited patents suggests that early, pioneering innovations have become standard knowledge, widely reused and built upon. A smaller number of **ACTIVE and PENDING patents,** particularly those published after 2010 (e.g., 2014–2017) demonstrate continuing relevance and commercial value in evolving IR technologies. Larger **simple family sizes** (up to 23) indicate strategic international protection for high-impact inventions. Overall, the citation pattern highlights a strong legacy effect of early IR patents alongside selective, high-value contemporary innovations driving current advancements.

7. Findings

The analysis of patent trends in Information Retrieval (IR) Systems reveals distinct temporal patterns of innovation over the period 2000–2024. As shown in **Figure 1**, patenting activity grew rapidly during the early 2000s, reaching a peak in 2002, followed by a period of relative stability until around 2011. Thereafter, a gradual decline in patent filings is observed, with moderate and consistent activity continuing through 2024. This trend indicates that IR technologies have transitioned from an intensive growth phase to a more mature stage of innovation. The distribution of patent types (**Figure 2**) further supports this observation: patent applications constitute the majority of filings, while a substantial number of granted patents provide technological validation. The presence of search reports, patents of addition, and amended documents reflects ongoing refinement and incremental innovation within the IR domain, even as overall patent volumes stabilise.

Technological focus analysis based on classification systems highlights the core and evolving nature of IR innovations. The dominance of CPC codes related to digital information retrieval, indexing, and database search mechanisms (**Figure 3**) confirms that foundational IR architectures continue to drive innovation. Similarly, IPCR classifications (**Figure 4**) are heavily concentrated in digital computing arrangements for information retrieval, data processing, and sector-specific information systems, reinforcing the centrality of computational retrieval technologies. US classification codes (**Figure 5**) further demonstrate diversification into language processing, query handling, and system-level retrieval functions. Notably, the appearance of AI- and machine-learning-related classes across CPC and IPCR systems indicates increasing technological convergence within IR systems. Jurisdictional analysis (**Figure 6**) reveals a highly centralised global innovation landscape, with the United States dominating patent filings, followed by Japan and China. The presence of WIPO and European Patent filings highlights applicants' strategic intent to secure international protection, while moderate activity in other jurisdictions reflects region-specific innovation hubs.

Analysis of patent stakeholders and lifecycle indicators provides further insights into the structure and maturity of IR innovation. Leading applicants and owners (**Figures 7 and 8**) are predominantly large multinational corporations such as Google, Microsoft, IBM, and Thomson Reuters, indicating strong corporate control over IR-related intellectual property. Attorney and agent data (**Figure 9**) show that patent prosecution is concentrated among a limited number of specialised law firms and professionals, reflecting the complexity of software and information technology patents. Inventor analysis (**Figure 10**) reveals a skewed distribution, where a small group of prolific inventors contributes disproportionately to IR innovation. Legal status distribution (**Figure 11**) shows a balance among expired, active, pending, and discontinued patents, suggesting the simultaneous presence of legacy technologies and ongoing innovation. The dominance of application-stage kind codes (**Figure 12**) highlights sustained disclosure activity. In contrast, citation analysis (**Figure 13)** identifies early IR patents as highly influential, demonstrating their continued role in shaping subsequent technological developments. Overall, the findings indicate that IR systems represent a mature yet evolving technological domain, characterised by corporate dominance, technological convergence, and the enduring influence of foundational innovations.

8. Implications of the Study

a) For Libraries and Librarians: The findings of this study hold significant implications for libraries and librarians engaged in digital information management and knowledge organisation. By analysing patent trends in Information Retrieval (IR) systems, libraries can anticipate technological advancements in search engines, indexing, database architectures, and AI-driven retrieval systems, enabling strategic integration of intelligent search tools, semantic retrieval functionalities, and personalised content delivery. Insights into leading patent owners, inventors, classification codes, and global jurisdictional distributions allow libraries to identify potential software solutions, vendor platforms, and collaborative opportunities aligned with institutional goals. For librarians, understanding dominant patent classifications and lifecycle dynamics, such as granted, pending, or expired patents, facilitates informed decision-making regarding technology adoption, legal considerations, and sustainability. Librarians can leverage this knowledge to enhance search efficiency, support evidence-based services, guide users through complex digital resources, and monitor emerging industry trends. Additionally, the study informs professional development programs focused on technology literacy, AI integration, and patent-informed resource management. Overall, these insights equip libraries and their staff with a strategic framework to adopt high-impact innovations, improve information retrieval capabilities, enhance user satisfaction, and maintain competitiveness in an increasingly data-driven and technologically dynamic knowledge environment.

b) For Research and Development*:* This study offers valuable insights for research and development professionals engaged in IR system innovation. By analysing patent trends, technological focus, and citation impact, R&D teams can identify high-value areas for investment, emerging methodologies, and gaps in existing solutions. The concentration of patents in specific CPC and IPCR classifications highlights technological domains such as AI-assisted retrieval, indexing optimisation, and database management, guiding targeted research initiatives. Understanding global jurisdictional distribution and patent ownership patterns enables organisations to develop competitive strategies, secure intellectual property rights, and explore potential collaborations with leading innovators. Additionally, citation analysis of highly influential patents provides a roadmap for building upon foundational work while avoiding duplication of existing technologies. Insights into patent lifecycle stages and kind codes support efficient project planning, resource allocation, and risk assessment, ensuring that R&D efforts focus on commercially viable and technically validated innovations. The study also emphasises the importance of interdisciplinary approaches that combine computer science, library science, and information management to develop next-generation IR solutions. Overall, the research informs strategic R&D decision-making, accelerates technological advancement, and fosters innovation in both academic and corporate settings.

a. **Strategic Decision Making and Intellectual Property Management in The IR Domain**

a) **Support Evidence-Based Decisions:** Provide data-driven insights from patent trends, classifications, and citation analysis to guide institutional and corporate decision-making.

b) **Identify High-Impact Technologies:** Highlight emerging IR technologies and AI-driven solutions that warrant strategic investment or adoption.

c) **Prioritise Resource Allocation:** Assist libraries, research institutions, and R&D teams in allocating resources efficiently toward high-value IR tools and systems.

d) **Guide Technology Acquisition:** Enable informed selection of commercial or open-source IR platforms aligned with current and future technological needs.

e) **Facilitate Strategic R&D Planning:** Help R&D departments focus on areas with the greatest innovation potential, avoiding duplication and optimising development efforts.

f) **Enhance Intellectual Property (IP) Strategy:** Provide insights into patent ownership, legal status, and international filings to strengthen IP management and competitive positioning.

g) **Support Risk Assessment:** Assess the maturity and commercial viability of IR technologies for informed adoption and integration.

h) **Encourage Collaboration:** Identify potential partners, inventors, and organisations for collaborative innovation in IR systems.

i) **Monitor Global Trends:** Track geographical and jurisdictional patent activity to align technology strategy with international innovation ecosystems.

j) **Promote Knowledge-Driven Innovation:** Ensure adoption of IR technologies is guided by evidence from patent analytics, maximising efficiency, effectiveness, and strategic impact.

9. Conclusion

The present patent-based analysis of Information Retrieval (IR) Systems covering the period from 2000 to 2024 offers a comprehensive understanding of the evolution, concentration, and strategic dynamics shaping innovation in this critical technological domain. The study reveals that IR patenting has progressed through distinct phases, beginning with rapid growth in the early 2000s, driven by intensive research and development associated with the expansion of web search engines, digital libraries, and online information systems. This phase was followed by a period of consolidation and incremental innovation, reflecting the maturation of retrieval models, indexing techniques, and relevance-ranking algorithms. The gradual decline in standalone IR patent filings after 2012, accompanied by stable but subdued activity in recent years, indicates a structural shift in innovation strategies, where IR functionalities are increasingly embedded within broader AI-driven, data-intensive, and platform-based technologies rather than patented as independent systems. Technological analysis using CPC, IPCR, and US classification codes demonstrates a strong and sustained focus on digital computing arrangements, database structures, query processing, and intelligent retrieval mechanisms. The prominence of key classifications such as G06F16/951 and G06F17/30 underscores the foundational role of structured data management and algorithmic retrieval in modern IR systems. The growing presence of artificial intelligence and machine learning related classifications highlights technological convergence and the evolution of IR systems toward semantic search, personalised retrieval, and context-aware information delivery. Jurisdictional analysis further reveals a highly centralised innovation landscape, dominated by the United States, followed by Japan and China. At the same time, international filings through WIPO and European Patent Offices indicate strategic efforts by patent applicants to secure broader market protection. Analysis of patent stakeholders shows strong corporate dominance, with multinational technology firms such as Google, Microsoft, IBM, and Thomson Reuters emerging as leading applicants and patent owners. This reflects strategic intellectual property management and concentrated R&D investment in IR technologies. Inventor-level analysis indicates that innovation is driven by a relatively small group of prolific contributors. At the same time, legal intermediaries and patent attorneys play a critical role in shaping patent prosecution and protection strategies. Citation analysis highlights the enduring technological influence of early IR patents, many of which have entered the public domain yet continue to inform contemporary innovation trajectories. For libraries and librarians, these findings offer valuable insights into emerging retrieval technologies, supporting informed decision-making for system adoption, service enhancement, and user-centric information access. For R&D professionals and policymakers, the study identifies high-impact technological areas, competitive dynamics, and opportunities for strategic collaboration.

During data collection and compilation, certain challenges were encountered that future researchers should carefully consider. The study relied exclusively on patents indexed in Lens.org, which, despite its extensive coverage, may not fully represent all regional or non-digitised patent filings. Variations in the terminology used by patent applicants limit the ability to capture all relevant IR-related innovations, as some technologies are described with broader or alternative terms. Additionally, inconsistencies in inventor names, applicant identities, and legal status reporting across jurisdictions required careful normalisation and verification, increasing the complexity of data processing. Future research may expand this study by incorporating multiple patent databases to enhance coverage and cross-validation. Broader and more inclusive search strategies can be developed to capture emerging and interdisciplinary IR innovations. Comparative analyses between corporate, academic, and open-source IR development models would offer deeper insights into knowledge diffusion and technology transfer. Further studies may also examine patent citation networks, collaboration patterns, and the societal implications of AI-driven retrieval systems,particularly concerning data privacy, sustainability, and accessibility. Aligning technological advancements with practical applications in libraries, academia, and enterprise environments will remain a critical direction for advancing user-centric, ethical, and efficient Information Retrieval systems.

Acknowledgment

The author sincerely acknowledges the support and facilities provided by the Department of Library and Information Science, Rabindra Bharati University, Kolkata, for carrying out the present study. The author expresses sincere gratitude to **Dr. Ziaur Rahman**, PhD Supervisor, for his valuable guidance, continuous support, and constructive suggestions throughout the course of the present study.

References

Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. *World Patent Information*, 37, 3–13. https://doi.org/10.1016/j.wpi.2013.09.003

Abraham, B. P., & Moitra, S. D. (2001). Innovation assessment through patent analysis. Technovation, 21(4), 245–252. https://doi.org/10.1016/S0166-4972(00)00047-7

Albino, V., Ardito, L., Dangelico, R. M., & Petruzzelli, A. M. (2014). Understanding the development trends of low-carbon energy technologies: A patent analysis. *Applied Energy*, 135, 836–854. https://doi.org/10.1016/j.apenergy.2014.08.061

Bergmann, I., Butzke, D., Walter, L., Fuerste, J. P., Moehrle, M. G., & Erdmann, V. A. (2008). Evaluating the risk of patent infringement by means of semantic patent analysis: The case of DNA chips. *R&D Management*, 38(5), 550–562. https://doi.org/10.1111/j.1467-9310.2008.00537.x

Breitzman, A. F., & Mogee, M. E. (2002). The many applications of patent analysis. *Journal of Information Science*, 28(3), 187–205. https://doi.org/10.1177/016555150202800303

Chang, S. B. (2012). Using patent analysis to establish technological position: Two different strategic approaches. *Technological Forecasting and Social Change*, 79(1), 3–15. https://doi.org/10.1016/j.techfore.2011.04.006

Choi, J., Jun, S., & Park, S. (2016). A patent analysis for sustainable technology management. *Sustainability*, 8(7), 688. https://doi.org/10.3390/su8070688

Chung, D., Kwon, Y., & Kim, Z. (2023). Technology life cycle and commercialization readiness of hydrogen production technology using patent analysis. *International Journal of Hydrogen Energy*, 48(33), 12139–12154.

Daim, T. U., Rueda, G., Martin, H., & Gerdsri, P. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting and Social Change*, 73(8), 981–1012.

Dewulf, S., & Childs, P. R. (2021). Patent data driven innovation logic: Textual pattern exploration to identify innovation logic data. *In International TRIZ Future Conference* (pp. 170–181). Springer International Publishing.

El Boukhari, R., Matin, M., Atanasov, A. G., & Fatimi, A. (2025). Patent landscape analysis of antioxidants from Lamiaceae plants: Innovations and applications. *Current Biotechnology*, 14(4), 244–256. https://doi.org/10.2174/012211550138727225 0917121009

Hicks, D., Breitzman, T., Olivastro, D., & Hamilton, K. (2001). The changing composition of innovative activity in the U.S. a portrait based on patent analysis. Research Policy, 30(4), 681–703. https://doi.org/10.1016/S0048-7333(00)00070-0

Huang, Z., Chen, H., Yip, A., Ng, G., Guo, F., Chen, Z. K., & Roco, M. C. (2003). Longitudinal patent analysis for nanoscale science and engineering: Country, institution, and technology field. *Journal of Nanoparticle Research*, 5(3), 333–363. https://doi.org/10.1023/A:1025771614454

Jiang, L., Zou, F., Qiao, Y., & Huang, Y. (2022). Patent analysis for generating the technology landscape and competition situation of renewable energy. *Journal of Cleaner Production*, 378, 134264. https://doi.org/10.1016/j.jclepro.2022.134264

Jun, S., & Sung Park, S. (2013). Examining technological innovation of Apple using patent analysis. *Industrial Management & Data Systems*, 113(6), 890–907. https://doi.org/10.1108/IMDS-02-2013-0076

Kim, G., & Bae, J. (2017). A novel approach to forecast promising technology through patent analysis. *Technological Forecasting and Social Change*, 117, 228–237. https://doi.org/10.1016/j.techfore.2016.10.039

Kim, Y. G., Suh, J. H., & Park, S. C. (2008). Visualization of patent analysis for emerging technology. *Expert Systems with Applications*, 34(3), 1804–1812. https://doi.org/10.1016/j.eswa.2007.01.016

Krestel, R., Chikkamath, R., Hewel, C., & Risch, J. (2021). A survey on deep learning for patent analysis. World Patent Information, 65, 102035. https://doi.org/10.1016/j.wpi.2021.102035

Li, Y. R., Wang, L. H., & Hong, C. F. (2009). Extracting significant-rare keywords for patent analysis. *Expert Systems with Applications*, 36(3), 5200–5204. https://doi.org/10.1016/j.eswa.2008.07.039

Liu, S. J., & Shyu, J. (1997). Strategic planning for technology development with patent analysis. *International Journal of Technology Management*, 13(5-6), 661–680.

Mao, G., Han, Y., Liu, X., Crittenden, J., Huang, N., & Ahmad, U. M. (2022). Technology status and trends of industrial wastewater treatment: A patent analysis. Chemosphere, 288, 132483. https://doi.org/10.1016/j.chemosphere.2021.132483

Mejia, C., & Kajikawa, Y. (2025). Patent research in academic literature: Landscape and trends with a focus on patent analytics. *Frontiers in Research Metrics and Analytics*, 9, 1484685. https://doi.org/10.3389/frma.2024.1484685 Frontiers

Messeni Petruzzelli, A., Maria Dangelico, R., Rotolo, D., & Albino, V. (2011). Organizational factors and technological features in the development of green innovations: Evidence from patent analysis. *Innovation*, 13(3), 291–310. https://doi.org/10.5172/impp.13.3.291

Moro-Visconti, R. (2025). From incremental know-how to patent-driven startups. *In Startup Valuation: From Strategic Business Planning to Digital Networking* (pp. 173–239). Springer Nature Switzerland.

Sampaio, P. G. V., González, M. O. A., De Vasconcelos, R. M., dos Santos, M. A. T., de Toledo, J. C., & Pereira, J. P. P. (2018). Photovoltaic technologies: Mapping from patent analysis. *Renewable and Sustainable Energy Reviews*, 93, 215–224. https://doi.org/10.1016/j.rser.2018.04.068

Shomee, H. H., Wang, Z., Ravi, S. N., & Medya, S. (2025). A survey on patent analysis: From NLP to multimodal AI. *In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (pp. 8545–8561).

Tang, J., Wang, B., Yang, Y., Hu, P., Zhao, Y., Yan, X., … Usadi, A. K. (2012, August). PatentMiner: Topic-driven patent analysis and mining. *In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1366–1374). https://doi.org/10.1145/2339530.2339696

Tseng, Y. H., Lin, C. J., & Lin, Y. I. (2007). Text mining techniques for patent analysis. *Information Processing & Management*, 43(5), 1216–1247. https://doi.org/10.1016/j.ipm.2006.11.017

Yang, D., Lee, J., Song, N. C., Lee, S., Kim, S., Lee, S., & Choi, S. (2023). Patent analysis on green hydrogen technology for future promising technologies. *International Journal of Hydrogen Energy*, 48(83), 32241–32260. https://doi.org/10.1016/j.ijhydene.2023.06.154

Zhou, X., Zhang, Y., Porter, A. L., Guo, Y., & Zhu, D. (2014). A patent analysis method to trace technology evolutionary pathways. *Scientometrics*, 100(3), 705–721. https://doi.org/10.1007/s11192-014-1287-8

About Authors

Goutam Dinda is a Research Scholar in the Department of Library and Information Science, Rabindra Bharati University, West Bengal, India. He obtained his B.Lib.I.Sc. (2021) **and** M.Lib.I.Sc. (2022) from the Department of Library and Information Science, Rabindra Bharati University. He also holds an M.A. in History **(2016)** from the University of Calcutta, a B.Ed. **(2019)** from the West Bengal University of Teachers' Training, Education Planning and Administration (WBUTTEPA), and a B.A. (Honours) in History (2014) from Contai P. K. College under Vidyasagar University. His research interests include research metrics, library classification, library management software, library management, and academic libraries.

Blind Peer Reviewed
(Previously UGC-CARE)
ISSN: 0972-2750

Green Library Initiatives: A Bibliometric Analysis of Global Research Trend

Ningombam Iroshini & Ch. Ibohal Singh

Manuscript Received on –
1st January, 2026
Primary Reviewed on –
6th January, 2026
Peer Reviewed on –
5th February, 2026
Revised on –
09th February, 2026
Accepted for publication
12th February, 2026

ABSTRACT

***Introduction**: With the emergence of climate change, the role of libraries worldwide in tackling this issue must be considered. The green library is one such effort that helps to mitigate climate change.*

***Purpose**: The present study aims to determine the growth trends in publications in the field of green libraries. The dynamics of the scientific literature's development and evolution as a depiction of research have also been assessed in this area.*

***Objectives:** The article's primary contribution is a quantitative analysis of the scientific literature on green libraries, indexed in the Lens.org database, published between 2001 and the present day.*

***Research Problem:** The contribution of global green library research is considered less when compared to other areas of research in the library and information science field.*

***Methodology:** The study uses the bibliometric method to assess green libraries on a global basis.*

***Findings:** It discusses the publication trends, research field, prolific journals, and so on. It, as expected, will help policymakers, administrators, LIS schools and library professionals raise awareness, plan and execute green library initiatives in the near future.*

Keywords: *Green Library, Sustainable Library, Lens.org, Bibliometrics*

Ms. Ningombam Iroshini
Research Scholar
Dept. of Lib. & Inf. Sc.
Manipur University
iroshningombam.2019@gmail.com

Dr. Ch. Ibohal Singh
Assistant Professor
Dept. of Lib. & Inf. Sc.
Manipur University
ibohal68@gmail.com

1. Introduction

The environment is the setting or circumstances in which people, organisations, animals, or plants exist and function. Natural, social, and cultural environments are all possible. "Meeting the needs of the present without compromising the ability of future generations to meet their own needs" is what is meant by Sustainable Development. Sustainability refers to how human progress and development should be supported by the current environment, how environmental constraints might be reduced, and how human quality of life can be improved (Ruggerio, 2021).

Along with other dangers, such as socioeconomic inequality, climate change is endangering the current environment. The goal of sustainable development should be the goal of every organisation. This includes libraries of all sorts, which have the potential to actively and significantly contribute to sustainable development.

The term "Green Library" describes libraries that help preserve the environment and protect the earth, its natural systems, and its resources. The Green Libraries aim to reduce water and energy use through sustainable building design that prioritises natural and renewable resources. According to the Online Dictionary for Library and Information Science, the Green Libraries are "designed to minimise negative impact on the natural environment and to maximise indoor environmental quality by means of careful site selection, the use of natural construction materials like nature and biodegradable products, conservation of resources, and responsible waste disposal" (Rakesh & Namarta, 2021). There is no unambiguous definition of a green library. However, the majority of them have overarching themes that are managed by all of them, such as eliminating negativity and generating positive effects in the community (Hellsten, 2002). Thus, a green and sustainable library is a library which takes into account environmental, economic and social sustainability.

Since the 1990s, the number of articles in professional and scientific journals about green library initiatives aimed at reducing the detrimental effects of human activity on the environment has increased in the field of library and information studies (Kang, 2020). Numerous bibliometric evaluations conducted across various contexts and regions show that the concept of green libraries is gaining scholarly attention. There is an expanding focus on green technological breakthroughs and sustainable practices in library settings, reflecting a larger shift towards environmental concerns in library science (Rifq & Pushpita, 2025). Additionally, the green library prioritised ecology projects & resource efficiency, in line with the global emphasis on sustainability (Simin and Feng, 2023). Green library research is linked to broader sustainability and environmental disciplines, highlighting the field's multidisciplinary nature (Muhmad et al., 2024). In conclusion, a Nation's sustainable development and humanity's involvement in climate change are humanity's involvement in climate change are fundamental issues for society, and libraries are no exception.

2. Green Library Initiatives in combating Climate Change

Environmental sustainability has become a global priority as people become more conscious of the harmful effects of climate change, pollution, and ecological destruction. The goal of global initiatives such as UNESCO's education for sustainable development is to prepare younger generations with the knowledge and skills to address future environmental concerns (Putra, 2025).

As a result of these initiatives, green libraries have become important contributors to the advancement of sustainability through the use of recycled materials, energy efficiency, and environmental friendly architectural designs. These libraries involved communities in environmental projects, as recognised by the IFLA Green Library Award for outstanding contributions to sustainability (IFLA,2024). This emphasises the role of libraries in both social change and the stewardship of the environment and climate change.

The idea of green libraries places a strong emphasis on incorporating sustainable methods into library operations, which can greatly lessen their environmental impact (Fedorowicz-Kruszewska, 2020; Charney and Hauke,2020). Apart from its advantages for the environment, green libraries are essential to achieving the Sustainable Development Goals, especially through public education and carbon emission reduction (U.N.,2020). Green libraries are useful teaching venues for promoting proactive community involvement and environmental awareness. For instance, Vijesh demonstrated how libraries may integrate their operations with the UN's Sustainable Development Goals, strengthening their position as educational platforms for sustainability (Vijesh et al., 2024). Furthermore, bibliometric studies in green libraries have investigated the importance of green leadership and innovative practices in developing green library projects, providing further insights into how green ideas are conveyed and accepted within library contexts. In summary, the analysis acts as a critical tool for tracking the development, thematic focus and research trends in green library studies.

3. Objectives of the Study

The study attempts to achieve the following objectives:

1. To assess the annual productivity of the articles in the field.
2. To examine the author's productivity;
3. To investigate highly cited articles, prolific journals, top contributing countries, and top publishers in the field of green library research.
4. To assess the fields of study under different aspects.

4. Methodology and Scope

This bibliometric analysis has been conducted using the Lens.org database, which contains over 200 million scholarly articles. Lens.org is an effective database for tracking the connections between research and innovation (Penfold,2020). The study is based on a comprehensive bibliometric analysis of green library research literature, including annual research productivity, prolific authors, journals, publishers, the most productive countries, and field-of-study mapping. A carefully constructed search query (Table 1) was created to retrieve pertinent studies, resulting in 313 data points. Publications that met the search parameter were analysed.

Search Query	(Title: ("Green Library," OR ("Sustainable Library," OR ("Eco-friendly Library"))))
Database	Lens.org
Query Performed	12-09-2025

Table 1- Search Query performed in Lens.org

The retrieved data were systematically analysed using the R Statistical package (Biblioshiny) and Microsoft Excel. A descriptive analysis of the dataset is presented in Table 2, which emphasises significant distributions and trends of the literature. Scimago Journal Rank (SJR) was also used to evaluate the quality of the most productive journals. "Datawrapper" was also used to create a map indicating the research distribution of countries within the discipline.

Description	Results
MAIN INFORMATION ABOUT DATA	
Timespan	2001:2025
Sources (Journals, Books, etc.)	203
Documents	313
Annual Growth Rate %	15.97
Document Average Age	7.15
Average citations per doc	3.067
References	1441
DOCUMENT CONTENTS	
Keywords Plus (ID)	209
Author's Keywords (DE)	209
AUTHORS	
Authors	499
Authors of single-authored docs	149
AUTHORS COLLABORATION	
Single-authored docs	163
Co-Authors per Doc	1.82
DOCUMENT TYPES	
Journal article	313

Table 2- Descriptive Analysis of the Retrieved Data

5. Data Analysis

The following sections demonstrate the analysis of the retrieved literature.

5.1. Year-wise publication

Figure 1 represents the annual publication trends in green libraries around the world. The publications in the green library were low at the beginning of the century. With the increase in time, the publication trend increases consistently with a few small spikes in between years. The year 2024 marks the most productive year with 28 publications overall.

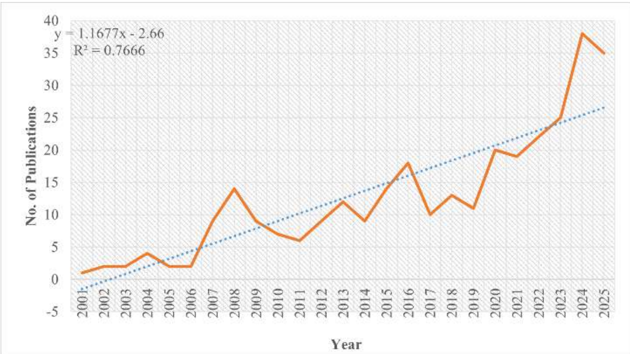


Fig 1- Annual Growth Rate of Green Library Publications

As depicted by the value of R^2, i.e. 0.7666, the publication's growth is expected to be consistent and exponential.

5.2. Prolific Authors

The tables below, 3 and 4, display the most prolific authors in green library research publications. The h-index, g-index, m-index, total citations and total publications, along with the year from which the author's publication starts, are depicted in the table. Considering the number of publications, Fedorowicz-Kruszewska M and Noh Y published the most articles in the research area, while Jankowska Ma published the article with the highest citation count.

Author	h-index	g-index	m-index	TC	NP	PY_start
Fedorowicz-Kruszewska M	3	5	0.273	45	5	2015
Noh Y	3	3	0.214	12	5	2012
Antonelli M	3	4	0.167	70	4	2008
Jankowska Ma	3	3	0.158	106	3	2007
Mwanzu A	3	3	1	31	3	2023
Ahn I	2	2	0.143	6	3	2012
Ismail Mn	2	2	1	4	3	2024
Mulyati T	2	2	0.4	6	3	2021
Yunus N	2	2	1	4	3	2024
Adeniji Ma	2	2	0.143	4	2	2012

Table 3- Top 10 authors as per Publications

Author	h-index	g-index	m-index	TC	NP	PY_start
Jankowska Ma	3	3	0.158	106	3	2007
Marcum Jw	2	2	0.118	101	2	2009
Antonelli M	3	4	0.167	70	4	2008
Fedorowicz-Kruszewska M	3	5	0.273	45	5	2015
Mwanzu A	3	3	1	31	3	2023
Brodie M	2	2	0.143	23	2	2012
Afacan Y	1	1	0.111	19	1	2017
Bosire-Ogechi E	2	2	0.667	17	2	2023
Odero D	2	2	0.667	17	2	2023
Adamatzky A	1	1	0.091	15	1	2015

Table 4- Top 10 authors as per Citations

5.3. Highly Cited Articles

Table 5 represents the top 10 most productive research publications based on citations. Research article highlighting sustainable library authored by Jankowska, MA, published in the year 2010 in the journal College and Research Libraries received a total of 90 citations and topped the most productive research article. Green Library movement, smart libraries, library services to promote sustainable libraries, and the development of green strategies for different libraries are among the various research areas that showcase a wide spectrum of high-impact research publications.

Title	Year	Authors	Journal	TC
Sustainable Challenges for Academic Libraries: Planning for the Future	2010	Jankowska, MA	College & Research Libraries	90
The Green Library Movement: An overview and beyond	2008	Antonelli, M	Electronic Green Journal	58
Green Libraries are more than just buildings	2013	Aulisio, GJ	Electronic Green Journals	39

Smart Libraries	2018	Schopfel, J	Infrastructures	37
More than just a green building: Developing green strategies at the University of Hong Kong Library	2016	Jones, L., & Wong, W	Library Management	30
Challenges in establishing and maintaining functional school libraries: Lessons from Limpopo Province, South Africa	2016	Mojapelo, SM	Journal of Librarianship and Information Science	25
Web 2.0, Library 2.0, Library User 2.0, Librarian 2.0: Innovation Services for Sustainable Libraries	2009	Peltier-Davis, C	Computers in Libraries	25
Green libraries and Green librarianship- Towards conceptualisation	2020	Fedorowicz-Kruszewska, M	Journal of Librarianship and Information Science	23
From the ground Up: Creating a sustainable library outreach program for International students	2015	Langer, C & Kuba, H	Journal of Library Administration	20
Building the sustainable library at Macquarie University	2013	Brodie, M	Australian Academic & Research Libraries	20

TC=Total Citations

Table 5- Top 10 most productive publications on the green library

5.4. Prolific Journals

Table 6 depicts the productive journals on green library research. Electronic Green Journal and Scientific and Technical Libraries produce the highest number of publications, with a total of 10 research articles each and h-Index of 16 and 26, respectively. Considering the number of citations received, Electronic Green Journal also produced the highest number of citations with 117 citations received, followed by Library Management with 82 citations. International Journal of Research in Library Science ranked the highest with 8.076 in SCImago Journal Rank (SJR). SCImago Journal Rank is the metric that evaluates the scientific influence of academic journals.

Journals	**TP**	**% of TP**	**TC**	**% of TC**	**SJR**	**h-index**
Electronic Green Journal	10	3.19	117	12.19	0.184	16
Scientific And Technical Libraries	10	3.19	23	2.40	0.43	26
International Journal Of Research In Library Science	9	2.86	10	1.04	8.076	15
Library Management	8	2.56	82	8.54	0.449	41
IFLA Journal	7	2.24	32	3.33	0.546	29
Business Information Review	5	1.60	4	0.42	0.358	26
IOP Conference Series: Earth And Environmental Science	5	1.60	6	0.62	0.214	58
Journal Of Librarianship And Information Science	5	1.60	55	5.73	0.665	41
Public Library Journal	4	1.28	14	1.46	0.68	24
Australian Academic And Research Libraries	3	0.95	23	2.40	0.640	31

TP=Total Publications; TC=Total Citations; SJR= Scimago Journal Rank

Table 6- Top 10 prolific journals

5.5. Top Contributing Countries

Figure 2 represents the geographical patterns of the top 15 most productive countries in Green Library research all over the world. Out of the total 313 research articles, the United States contributed a total of 27 research articles, followed by China and India contributing 17 and 10 research articles, respectively.

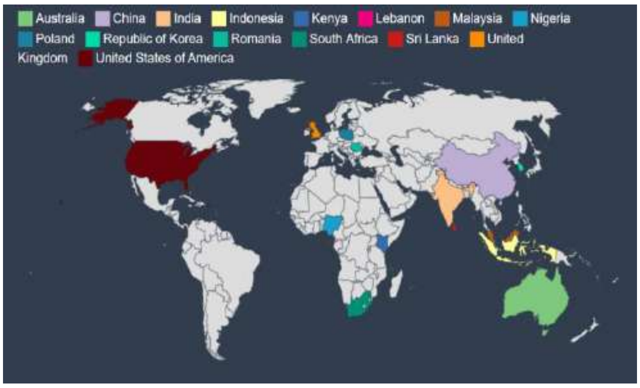


Fig 2- Top contributing Countries

5.6. Top Publishers

Table 7 represents the top publishers in publishing green library research. Informa UK Limited and SAGE Publications published the highest number of publications, i.e. 19 green library research articles contributed by 13 and 6 Journals respectively. It is followed by Emerald with 17 articles contributed by 6 journals.

Publisher	Journal	No. of Article
Informa UK Limited	Public Library Quarterly	3
	Australian Academic & Research Libraries	3
	Journal of Hospital Librarianship	1
	College & Undergraduate Libraries	1
	Journal of Library & Information Services in Distance Learning	1
	Journal of Library Administration	1
	Sustainability: Science, Practice & Policy	1
	History of European Ideas	1
	Technical Services Quarterly	2
	Medical Reference Services Quarterly	1
	Serials Review	1
	The Australian Library Journal	2
	Journal of Urban Affairs	1
SAGE Publications	Business Information Review	1
	SAGE Open	1
	IFLA Journal	7
	Journal of Librarianship & Information Science	5
	Business Information Review	4
	Alexandria: The Journal of National & International Library & Information Issues	1
Emerald	Library Management	8
	Journal of Documentation	1
	Collection & Curation	1
	The Bottom Line	3
	Library Hi Tech News	3
	New Library World	1
State Public Scientific Technical Library	Scientific & Technical Libraries	10
Diva Enterprises Private Limited	Journal of Library & Information Communication Technology	1
	Pearl: A Journal of Library & Information Science	3
	Gyankosh- The Journal of Library & Information Management	1
	International Journal of Information Dissemination & Technology	3
	Library Herald	1
International Journal of Research in Library Science	International Journal of Research in Library Science	9
California Digital Library (CDL)	Electronic Green Journal	7
Elsevier BV	The Journal of Academic Librarianship	2
	SSRN Electronic Journal	3
	Journal of Building Engineering	1
University of Idaho Library	Electronic Green Journal	3
	Library Philosophy & Practice	3
American Library Association	College & Research Libraries News	4
	Library Resources & Technical Services	1

Table 7- Top 10 Publishers in Green Library Research

5.7. Different Fields of Study

Figure 3 displays key terms with more frequency & density in a larger typeface, highlighting visually appealing information. The analysis shows that the key term 'Computer Science (117)' has the highest frequency amongst the keywords, followed by 'Business (101)' and 'Political Science (96)'. The essential terms highlighted in this study have a crucial role in the discussion.

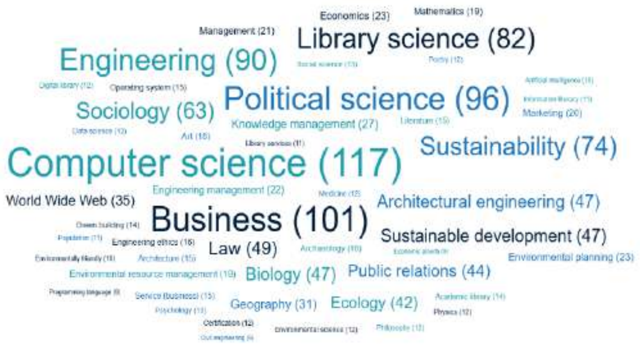


Figure 3- Word Cloud representing Field of Study

6. Result and Discussion

The most significant way that this research adds value to the current field of study is by providing an overview of earlier research conducted worldwide. This article identifies the most prolific authors in the chosen topic, allowing scholars to rapidly discover which country contributes the most to the domain. It also identifies the growth pattern of literature in order to investigate the growing interest in the field of study over time. Publications with the highest citations are also determined to find the most reliable publications. Top publishers and journals are also being evaluated to identify their active participation in the publication of high-quality literature. Finally, the field of study was examined to determine the most active area of research. The study reveals a developing body of literature that combines ecological sustainability, green technology, and creative management methods, demonstrating the growing importance of environmentally conscious approaches in libraries. This study will help to provide a more nuanced picture of how green concepts are investigated, applied and promoted in the library sector around the world.

7. Conclusion

A green library is a modern library that uses eco-friendly building materials, conserves natural resources, reduces waste, and operates efficiently and cost-effectively. The importance of making libraries greener has increased dramatically in recent years. In this context, librarians ought to make some serious choices, put in the required work to make their libraries greener, and take part in the green library movement. Numerous national and international organisations are providing support for improving environmentally friendly libraries. To ensure the success of the green library movement, library patrons, librarians, and the government should collectively take initiatives and actively engage. It is important to remember that libraries are particularly responsible for setting an example and supporting sustainability as knowledge gateways. Indeed, taking little steps towards sustainability can lead to significant long-term benefits.

References

Charney, M. & Hauke, P. (2020). Global action on the urgency of climate change: Academic and Research Libraries' Contribution. *College and Research Libraries News, 81*(3):114. https://doi.org.10.5860/crln.81.3.114

Fedorowicz-Kruszewska, M. (2020). Green Libraries and green librarianship: towards conceptualisation. *Journal of Librarianship and Information Science, 53*(4): 645-654. https://doi.org.10.1177/0961000620980830

Hellsten, I. (2002). From "Burning Library" to "Green Medicine". The Role of Metaphors in Communicating Biodiversity. *Science Communication, 24*(2): 229-245. https://doi.org.10.1177/107554702237848

IFLA. (2024). IFLA Green Library Award. 2024. https://www.ifla.org/g/environment.sustainability-and-libraries/ifla-green-library-award

Kang, Q. (2020). Library directors' concerns and attitudes towards going green and sustainability in China: An unexplored area. *Journal of Librarianship and Information Science, 52*(2): 382-398. https://doi.org.10.1177/0961000618818874

Muhmad, S.N. et al. (2024). Systematic literature review and bibliometric analysis of green finance & renewable energy development. *Sustainable Development, 32*(6): 7342-7355. https://doi.org.10.1002/sd.3093

Parmar, R.D. and Kamdar, N.H. (2021). Green Libraries. *Library Philosophy and Practice*. https://digitalcommons.un/edu/cgi

Penfold, R. (2020). Using the lens database for staff publications. *Journal of the Medical Librray Association,108*(2).

Putra, D.D. (2025). Exploring Green Libraries Innovation toward Environmental Sustainability in Indonesia: A Systematic Literature Review. *IOP Conference Series: Earth and Environmental Science, 1439*.

https://doi.org.10.1088/1755-1315/1439/1/012023

Rifqi, A.N. & Pushpita, A.G. (2025). Systematic mapping of green technology research trends in libraries. IOP Conference Series: Earth & Environmental Science.

Ruggerio, C.A. (2021). Sustainability and Sustainable Development: A review of principles and definitions. *Science of the Total Environment, 786*(1): 1-11. https ://doi.org.10.1016/j.scitotenv.2021.147481

Simin, L. & Feng, Y. (2023). Green library research: A bibliometric analysis. *Public Library Quarterly, 42*(4): 424-443. https://doi.org.10.1080/10616846.2022.2116886

United Nations. (2025). Transforming our world: The 2030 Agenda for Sustainable Development. 2015. https://sdgs.un.org/2030agenda.

Vijesh, P.V., Chopade, V., Joy, V. & Joseph, M.K. (2024). Designing a green library in alignment with the UN's Sustainable Development Goal: A case study of Rajagiri Business School Library, Kerala, India. *Electronic Green Journal, 49*. https://doi.org.10.5070/G314960177

About Authors

Ningombam Iroshini is a UGC-Senior Research Fellow in the Department of Library and Information Science, Manipur University, Imphal, currently pursuing a PhD. Her areas of interest include metric studies in LIS, Research Methodology, Foundation of Librarianship, Information resources and services. She has attended various international, national, and Regional seminars, conferences, and workshops, and has presented papers. She has 10 research articles to her credit, published in reputable journals and as book chapters.

Prof. Ch. Ibohal Singh is currently Head of the Department of Library and Information Science, Manipur University, Imphal. He has more than two decades of teaching, research and administrative experience. He has published around 100 research papers and authored/Edited 12 books. He has guided 3 M.Phils, 15 Ph.Ds, more than 100 MLIS dissertations, conducted more than 10 National-level seminars, conferences and workshops and attended a number of such National and International events and chaired a number of technical sessions held in India and abroad. Besides being a life member of different professional bodies like MALA, ILA, IASLIC, IATLIS, RAS, National Book Trust, etc., he is currently serving as Vice-President of IASLIC, President of MALA and Editor of Library Gee Eehou-MALA International Journal of Library Movement.

Blind Peer Reviewed
(Previously UGC-CARE)
ISSN: 0972-2750

Manuscript Received on –
16th December, 2025
Primary Reviewed on –
1st January, 2026
Peer Reviewed on –
27th January, 2026
Revised on –
10th February, 2026
Accepted for publication
18th February, 2026

Producing a Knowledge Graph from MARC21 Authority Data for Linked Open Data Services

Tapas Roy & Rajesh Das

ABSTRACT

Introduction: *This research aims to generate a suitable knowledge graph from MARC21 personal name authority data for linked data or linked open data services. This research work encompasses several technologies, including RDF Schema, Ontology Vocabularies, Linked Open Data, and Knowledge Graph mapping between RDF elements and MARC 21 authority tags.*

Objective: *The primary objective is to produce a knowledge graph from MARC21 personal name authority data for linked open data services.*

Methodology: *We have collected 500 personal name authority data points as a sample from the Library of Congress database for this research. The personal name authority data has been transformed into linked open data from MARC21 personal name authority using the OpenRefine tool. The linked open data has been produced with different RDF and Ontology schemes.*

Scope: *The knowledge graph is a conceptual network of entities for datasets on a semantic platform. The linked open data service is an important component of the semantic web. In this research study, we have transformed MARC21 personal name authority data into a linked open data knowledge graph. This transforming methodology is applicable to other MARC21 data, such as bibliographic data, subject authority data and other similar data.*

Findings and Results: *We found a linked open dataset of personal name authority data for the services on the semantic web platform. As a result, the dataset contains more than 2,000 RDF triples. This linked open dataset could be published on any semantic web platform for linked open data services.*

Keywords: *Personal Name Authority, Link Open Data, OpenRefine, MARC 21, Knowledge Graph, RDF Schema*

Mr. Tapas Roy
PhD Scholar
Dept. of Lib. & Inf. Sc.
University of Burdwan
tapasroy135013@gmail.com

Dr. Rajesh Das
Assistant Professor
Dept. of Lib. & Inf. Sc.
University of Burdwan
rajeshdas99@gmail.com

Introduction

Currently, we are at the forefront of advanced data technologies across all domains. The academic and educational fields are highly compatible with advanced data management and technologies. We advanced data management and technologies include various concepts such as Data Science, Big Data Analysis, AI/ML (Artificial Intelligence/Machine Learning), Semantic Web, Ontology, linked data management, and technologies/Services, among others. Library and information professionals are well-versed in the above concepts and technologies (Idhalama & Nwachukwu, 2025). The concepts of AI/ML, Data Science, and Big Data Analysis have emerged over the last seven or eight years and are becoming increasingly popular daily. Libraries and information centres are adopting these technologies accordingly.

Before these technologies, the Semantic Web gained popularity in the computer industry for publishing structured data that is both machine-readable and machine-processable on the web. It includes Ontology, HTTP/URI, Linked Data/Linked Open Data, and RDF technologies (Kar and Das, 2020). Various projects have been developed using the semantic web in American and European Countries, such as DBpedia, Wikidata, Geonames, Europeana, and DPLA. They utilise all Semantic Web technologies.

Our research proposes a methodology for generating a knowledge graph from MARC 21 personal name authority data, supporting linked data or linked open data services. This work encompasses several technologies, including RDF Schema, Ontology Vocabularies, Linked Open Data, and Knowledge Graph mapping between RDF elements and MARC 21 authority tags (Kar and Das, 2021). These steps are implemented one by one, as shown in the following sections. As a result of this research work, we found RDF triple-formatted data (also known as Linked Open Data) and visualised it into a knowledge graph (RDF Graph). These generated RDF datasets are serialised into different formats, such as Turtle, N-triple, JSON-LD, and RDF/XML (Cole et al., 2017).

Review of Literature

Saha and Das (2020) have developed a workflow model for transforming library bibliographic data into an RDF triple format as a Linked Open Dataset (LOD), integrating external LOD datasets such as OpenLibrary, VIAF, Wikidata, DBpedia, and GeoNames. The authors also described the methodology and steps involved in conducting this research, including the use of the OpenRefine tool for managing and organising messy data, generating links, reconciling data from external sources, and converting the source format to RDF. In this article, the authors emphasise the significance of semantic web technology in delivering advanced-level services for cataloguing and maintaining bibliographic and authority records in libraries.

Kar and Das (2021) present a workflow model in this article for converting personal name authority data into a LOD dataset, utilising the OpenRefine tool for data refinement, reconciliation with external data sources, and conversion to various RDF formats. In this research work, a methodology is proposed for transforming MARC21 personal name authority metadata into Linked Open Data (LOD) by integrating external LOD datasets, such as the Library of Congress Name Authority File (LCNAF) and the Virtual International Authority File (VIAF), to provide enriched information services to users. After that, they concluded that personal name authority records are important in a library management system and that they should be converted into a LOD dataset to provide further information about personal names used as access points to library bibliographic items.

Kar and Das (2020) have described a model for building a LOD framework to convert and store RDF graphs and RDF triples, demonstrating the process of integrating and configuring Apache Jena Fuseki with DSpace to convert metadata into an RDF triple model and make it available in various RDF formats. This research work highlights a methodology and technical aspects for designing and publishing a structured LOD dataset of bibliographic information from a digital repository containing more than 500 bibliographic datasets developed with DSpace digital library software. This allows other libraries to link their repositories with these LOD, providing additional relevant resources to their end-users.

Alemu et al. (2012) reported that recommendations for transitioning from document-centric to data-centric metadata emphasise the need to adjust current library models, such as RDA and FRBR, to models based on Linked Data principles. In this paper, the authors discuss the importance of re-conceptualising current metadata models in light of Linked Data principles, with implications for metadata sharing, serendipity, identification of Zeitgeist, and emergent metadata, as well as faceted navigation and enriching metadata with links. Their research primarily highlights the limitations of current library metadata models, which are primarily designed for human consumption rather than machine processing, and advocates for a shift toward data-centric models.

Schilling (2017) provides an overview of the current state of the discussion about linked data and its practical implementation in libraries. In this research work, he provides a solid introduction for practitioners who wish to engage in the conversation about linked data and its potential benefits for libraries.

Statement of Research Problem

In the advancement of ICT in every domain, libraries and the information sector are embracing this technology. At present, the library and information centres have developed their libraries with advanced ILMS using Koha, VDLS, LIBSYS, etc.; digital institutional repositories using DSpace, E-prints, GSDL, etc.; and library discovery services using VuFind, BlackLight, EBSCO, SUMMON, etc. Librarians and library professionals continually develop themselves to adopt modern technologies. They also provide their data to a semantic web platform. They convert the library data into RDF data, link open data, and provide it to the semantic web services. In semantic web technologies, there are many components, such as ontologies and RDF graphs. Knowledge graph, Link Open Data, etc. Basically, before producing an RDF graph or Linked Open Data, we need to design a proper knowledge graph. A properly constructed knowledge graph provides better service to end users. In the library and information domain, people are producing RDF files of their library data, but do not follow the knowledge graph. If a proper knowledge graph is not followed, then better RDF data won't be followed. So we can say that our research problem is the lack of a suitable knowledge graph before producing an RDF data or Linked Open Data.

Objectives

The objectives of this research work are to provide personal name authority data into the Linked Open Data using a suitable knowledge graph.

1) To produce a transformation of personal name authority data into the Linked Open Data using a suitable knowledge graph.
2) To convert the MARC 21 authority records file (koha.mrc) into a CSV file (koha.csv) by using the MarcEdit tool.
3) To develop a mapping between MARC 21 tags and the RDF schema.
4) To produce a suitable knowledge graph of the datasets.
5) To produce data filters and data cleaning by using the OpenRefine tool.

Methodology

The main objective of the research work is to transform personal name authority into the linked open data using a suitable knowledge graph. We have collected 500 records of personal name authority data as a sample from the Library of Congress database for the research.

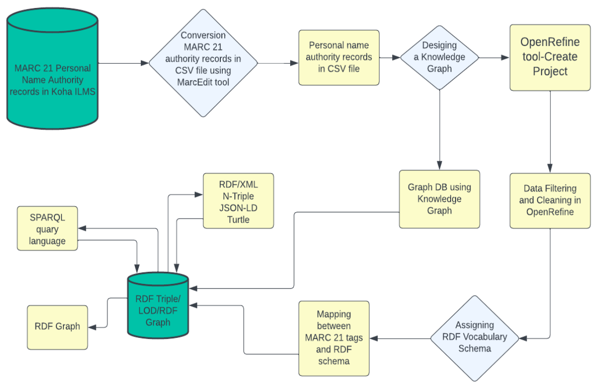


Figure-1: A Framework of Methodology to Transform Personal Name Authority Data into Link Open Data using Knowledge Graph

Our secondary objective is to determine how important a knowledge graph is to produce linked open data from personal name authority data. After studying related work, we found that a suitable knowledge graph can be represented as an RDF triple graph (see Figure 1).

Data Collection

Mapping between personal name authority tags with RDF class/property

The primary aim of our research work is the conversion of MARC records (personal name authority) into linked open data, with the generation of links. Before converting MARC records into LOD, we must define the RDF classes/properties of different schemas and ontologies. We have selected 35 MARC-21 tags for the personal name authority data, as shown in Table 2. Usually, there are two types of metadata: (i) administrative metadata and (ii) descriptive metadata in personal name authority data. We have used several namespaces and ontologies for mapping between MARC-21 personal name authority tags and RDF class/properties for the name authority data (Table 1).

Prefix	Name	URI
ri	Record Info	http://id.loc.gov/ontologies/RecordInfo#
madsrdf	Metadata Authority Description Schema in RDF	http://www.loc.gov/mads/rdf/v1#
foaf	Friend of a Friend Vocabulary	http://xmlns.com/foaf/0.1/
rdf	Resource Description Framework Ontology	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	Resource Description Framework Schema/Ontology	http://www.w3.org/2000/01/rdf-schema#

Table 1: Namespaces and vocabularies for personal name authority data

Details of the namespaces

The 'ri' refers to the RecordInfo vocabulary, which is created by the Library of Congress. It is used for the administrative metadata of MARC records. The namespace 'MADS/RDF' is a standard vocabulary for describing the MARC authority records. The namespace 'FOAF' (Friend of a friend) vocabulary describes people. The people are such things that are interlinked among themselves and perform several activities. The namespace 'RDF' (Resource Description Framework) is a data model for building a knowledge graph / RDF graph. It is a framework for describing e-resources on the web. It is an advanced data model for the semantic web. Basically, the RDF is represented as an RDF Triple (Subject, Predicate and Object/Value). RDF has some elements and properties for describing the e-resources in a data model. The namespace 'RDFS/RDF Schema' is an expression of RDF. It is such a framework for describing application-specific classes and properties. It contains a large number of classes and properties that are based on RDF and OWL ontology.

For this research work, we have adopted the following RDF vocabulary schema. As a dataset, we have exported the authority data in MARC 21 tags from Koha ILMS. Almost 35 different MARC 21 tags have been found in 500 personal name authority data that have been shown in Table 2.

Sl.No	MARC 21 Tags	Associate Fields / Columns	RDF Class / Properties/*vocabulary*	Metadata Type	Description
1	000 (LEADER)	leader	ri: RecordInfo	Administrative metadata	Specifications for Record Structure, Character Sets, and Exchange Media.
2	001 (CONTROL NUMBER)	control_number	ri: RecordID	Administrative metadata	Unique number assigned to a MARC record
3	003 (CONTROL NUMBER IDENTIFIER)	control_number_identifier	ri:recordContentSource	Administrative Metadata	Core related to an Organisation - of the description of the related resource

4	005 (DATE AND TIME OF LATEST TRANSACTION)	date_transaction	ri:recordChangeDate	Administrative Metadata	Transaction date associated with the related resource
5	008 (FIXED-LENGTH DATA ELEMENTS)	fixed_length	ri:recordElement	Administrative metadata	Potentially useful for retrieval and data management purposes
6	010$a (LIBRARY OF CONGRESS CONTROL NUMBER)	lc_control_number	madsrdf:idValue	Administrative metadata	Decribe the LCCN value
7	035$a (SYSTEM CONTROL NUMBER)	system_control_number			
8	040$a (ORIGINAL CATALOGUING AGENCY)	original_cataloging_agency	madsrdf:hasAffiliation	Administrative metadata	Property to associate an individual, such as a foaf:Agent, to a group or organisation with which an individual is or has been affiliated
9	040$b (LANGUAGE OF CATALOGUING)	language_of_cataloging	madsrdf:Language	Descriptive metadata	Describes a resource whose label represents a language.
10	040$e (DESCRIPTION CONVENTIONS)	description_conventions	madsrdf:code	Descriptive metadata	Describes a code as a string of characters associated with an authoritative or deprecated label
11	040$c (TRANSCRIBING AGENCY)	transcribing_agency	foaf:Agent	Administrative metadata	The Agent class is the class of agents; things that do stuff.
12	040$d (MODYFING AGENCY)	modyfing_agency	madsrdf:organization	Administrative metadata	Describes the group or organisation with which an individual is associated
13	046$f (BIRTH DATE)	birth_date	madsrdf:birthDate	Descriptive metadata	Describes the year a person was born. Date of birth may also include the month and day of the person's birth.
14	046$g (DEATH DATE)	death_date	madsrdf:deathDate	Descriptive metadata	Describes the year a person died. Date of death may also include the month and day of the person's death.
15	046$2 (SOURCE OF DATE SCHEME)	source_of_date_scheme	madsrdf:Source	Descriptive metadata	Describes a resource that represents the source of information about another resource
16	053$a (CLASSIFICATION NUMBER ELEMENT-SINGLE NUMBER	classification_number_element_beginning_number	madsrdf:classification	Descriptive metadata	Describes a classification code

	OR BEGINNING NUMBER OF SPAN)				
17	053$b (CALSSIFICATION NUMBER ELEMENT- ENDING NUMBER OF SPAN)	classification_number_element_ending_number	madsrdf:creationdateend	Descriptive metadata	Describes the earliest date (normally the year) associated with a work; that date may be the date the work was created, first published, or released.
18	100$a (PERSONAL NAME)	person_name	foaf:name	Descriptive metadata	Description a resource whose The label represents a personal name.
19	100$b (NUMBERATION)	numberation	madsrdf:PartNumberElement	Descriptive metadata	Describes a number elements
20	368$c (TYPE OF CORPORATE BODY)	type_of_corporate_body	madsrdf:CorporateName	Descriptive metadata	Describes a relationship between a CorporateName Authority and one of the same that is more broadly defined
21	368$2 (SOURCE)	source	madsrdf:Source	Descriptive metadata	Describes a resource that represents the source of information about another resource
22	370$c (ASSOCIATED COUNTRY)	associated_country	madsrdf:country	Descriptive metadata	Describes a Country associated with an address
23	370$2 (SOURCE OF TERM)	source_of_term	madsrdf:Source	Descriptive metadata	Describes a resource that represents the source of information about another resource
24	374$a (OCCUPATION)	occupation	madsrdf:occupation	Descriptive metadata	Occupation description
25	374$2 (SOURCE OF TERM)	source_of_term_occupation	madsrdf:hasSource	Descriptive metadata	Associates a resource description with its Source
26	375$a (GENDER)	gendeassociated_countryr	madsrdf:gender	Descriptive metadata	Gender description
27	375$2 (SOURCE OF TERM)	source_of_term_gender	madsrdf:hasSource	Descriptive metadata	Associates a resource description with its Source
28	377$a (ASSOCIATED LANGUAGE)	language	madsrdf:language	Descriptive metadata	Describes a Language that a person, organisation, or family uses for publication, communication, etc., or in which a work is expressed
29	378$q (FULLER FORM OF	fuller_form_of_person_nam	madsrdf:fullerName	Descriptive metadata	Describes a full form of name needed to distinguish a person from another person with the

	PERSON NAME)	e			same preferred name.
30	400$a (PERSONAL NAME)	person_name _tracing	madsrdf:PersonalName	Descriptive metadata	Describes a resource whose label represents a personal name.
31	400$d (DATE ASSOCIATED WITH A NAME)	date_associat ed_with_nam e	madsrdf:DateNameEle ment	Descriptive metadata	Describes a date name element
32	667$a (NONPUBLIC GENERAL NOTE)	nonpublic_ general_note	madsrdf:note	Descriptive metadata	Describes a nonpublic general note about the resource
33	670$a (SOURCE CITATION)	source_citatio n	Madsrdf:citationSource	Descriptive metadata	A note about the resource
34	670$b (INFORMATION FOUND)	information_fo und	madsrdf:note	Descriptive metadata	Describe the Information about the resource.
35	942$a (AUTHORITY TYPE)	authority_typ e	madsrdf:MADSType	Descriptive metadata	Describe the type of authority

Table 2: Mapping between MARC 21 tags of personal name authority and RDF Class/Properties/vocabulary

From the above table (Table 2), we have designed a knowledge graph of the dataset. A knowledge graph is a conceptual graph that is similar to an RDF graph. It is a pre-implementation stage of the RDF graph. Basically, a knowledge graph has three major components – Nodes, Edges, Levels or Classes, Edges and Property. The classes, edges and properties are some with subject, predicate, and object in the RDF graph.

Knowledge Graph:

Creating linked open data of personal name authority data through the OpenRefine tool

OpenRefine is a powerful tool for working with messy data, for large volumes of data, cleaning and filtering of data, transforming databases from one format to another, assigning RDF classes and properties, web services and external data. It is a tool for converting data into linked data with the generation of links. In this research work, OpenRefine (Version 3.8.2) is a major tool for transforming authority data to linked open data in different formats. Our study deals with two types of authoritative data, viz., personal name authority and subject authority; bibliographic data and DC metadata. This step includes the detailed procedure of transforming authority data, bibliographic data and DC metadata into linked open data with the generation of links (Hofer et al., 2024).

Transforming personal name authority data (MARC21 data) into LOD with links

Before the transformation of personal name authority into LOD, we drew our target knowledge graph based on Table 2.

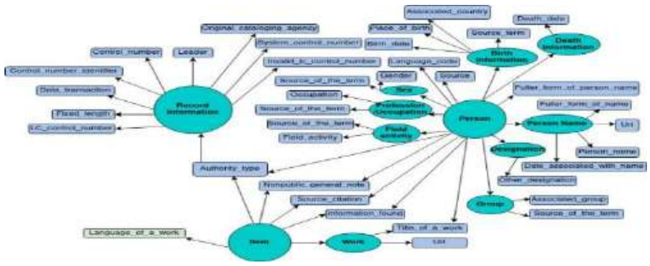


Figure 2: Knowledge Graph for Personal Name Authority Data

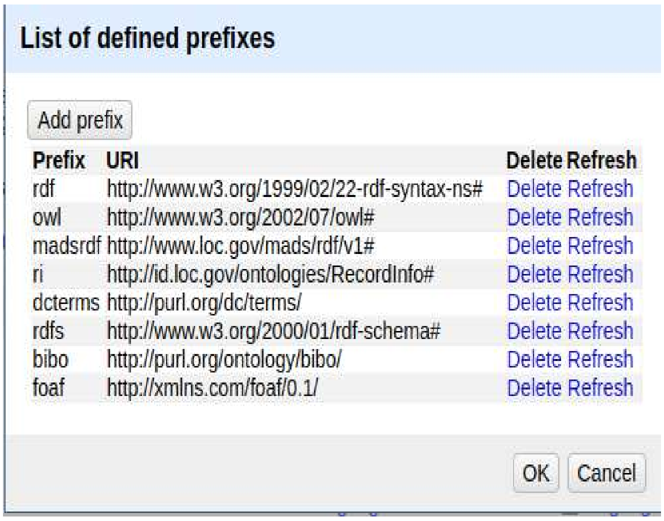
Figure 2 depicts a conceptual knowledge graph for personal name authority linked data. The actual knowledge graph can be visualised after producing the LOD using an RDF graph visualizer supported SPARQL endpoint interface. The LOD has three components – subjects, predicates and objects, commonly referred to as RDF triples. In the graph, entities (subjects) are circles, literals (objects) are rectangles and predicates are represented by arrows.

Figure-3: MARC 21 personal name authority tags in tabular format in OpenRefine (file: Personal name authority file in csv format)

Figure 3 displays the name authority records in the OpenRefine tools with various columns. The OpenRefine tool (3.7.9) is a powerful tool for managing messy data used for transforming RDF Triples. We have set a base URI "http://www.authors.org/perspnal_name/" in place of "http://127.0.0.1:3333/" for purposes of link generation of the personal name authority dataset. ority dataset.

Setting up "Subject-Predicate-Object" through DRF Skeleton

Before setting up 'Subject-Predicate-Object', we need to add prefixes (RDF schema/vocabulary/ontology).



List of defined prefixes

Add prefix

Prefix	URI	Delete Refresh
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#	Delete Refresh
owl	http://www.w3.org/2002/07/owl#	Delete Refresh
madsrdf	http://www.loc.gov/mads/rdf/v1#	Delete Refresh
ri	http://id.loc.gov/ontologies/RecordInfo#	Delete Refresh
dcterms	http://purl.org/dc/terms/	Delete Refresh
rdfs	http://www.w3.org/2000/01/rdf-schema#	Delete Refresh
bibo	http://purl.org/ontology/bibo/	Delete Refresh
foaf	http://xmlns.com/foaf/0.1/	Delete Refresh

OK Cancel

Figure 4: List of prefixes used for mapping personal name authority data

In Figure 4, we can see lots of prefixes like 'rdf', 'owl', 'madsrdf', 'ri', 'dcterms', 'rdfs', 'foaf', and 'bibo'. The primary ontology is 'madsrdfs' for personal name authority. The associate schemas are 'foaf', 'dcterms', 'bibo', 'rdf', 'rdfs', 'owl'. The 'ri' ontology is used for administrative data.

We have constructed an RDF skeleton that includes four root nodes – person_name URI (see figure 5), control_number URI (see figure 6), language URI (see figure 7) and gender URI (see figure 7). The root node person_name URI has formed 18 triples (subject-predicate-object), most of which use 'madsrdfs' prefixes. The second root node control_number URI formed 13 triples (subject-predicate-object), most of which use 'ri' for administrative records. The third root node language URI forms one triple and uses the 'madsrdfs' prefix. The last root node gender URI has formed one triple that uses the 'foaf' prefix.

Figure 5: *Complete RDF skeleton for personal name authority dataset (Snapshot-1)*

Figure 6: Complete RDF skeleton for personal name authority dataset (Snapshot-2)

Figure 7: *Complete RDF skeleton for personal name authority dataset (Snapshot-3)*

In the process of constructing the RDF skeleton, we have used GREL (Google Refine Express Language) for the integration of links for the RDF triples.

After setting up the RDF skeleton, we have exported the records into an RDF XML file of Personal Name Authority records, which is shown in Figure 8.


```
-<rdf:RDF>
 -<rdf:Description rdf:about="http://www.authors.org/personal_name/PrudhommeSully">
   <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
   <foaf:name>Prudhommo Sully</foaf:name>
   <foaf:gender>Males</foaf:gender>
   <foaf:birthday>1839</foaf:birthday>
   <madsrdf:deathDate>1907</madsrdf:deathDate>
   <madsrdf:occupation>Poets</madsrdf:occupation>
   <madsrdf:language>fre</madsrdf:language>
  -<madsrdf:citationSource>
    Enesco, G. Trois mélodies pour voix graves et piano, op. 4, 1956:;LC in RLIN, 10/31/86;Bib. nat.;Lorenz;Dict. des lett. franç., 19e s., 197
    2019;Nobel Foundation, WWW, viewed on June 10, 2019:
   </madsrdf:citationSource>
  -<madsrdf:note>
    t.p. (Sully Prudhomme) p. 1, 2nd group (Sully-Prudhomme);(hdg.: Sully Prudhomme, René François Armand, 1839-1907; usage: Sully Pr
    dit) usage: Sully Prudhomme);(Prudhomme (Sully); b. Paris, 1839);(Prudhomme (René-François-Armand Sully); b. 3/16/1839, Paris; d. 1
    Shilli-Prtidom 1839-1907; pseudonym of a French poet, the first ever Nobel Prize winner in literature (1901), a member of the French A
    Prtidom);The Nobel Prize in Literature 1901 (The Nobel Prize in Literature 1901 was awarded to Sully Prudhomme "in special recognit
    perfection and a rare combination of the qualities of both heart and intellect." Residence at the time of the award: France; Language: Fr
   </madsrdf:note>
   <madsrdf:MADSType>PERSO_NAME</madsrdf:MADSType>
   <ri:RecordID>1</ri:RecordID>
   <dcterms:format>rda</dcterms:format>
   <madsrdf:fullerName>René-François-Armand</madsrdf:fullerName>
  -<rdfs:label>
    Prudhommo, Sully,;Sully Prudhomme, René-François-Armand,;Prudhomme, René-François-Armand Sully,;Сюлли-Прюдом,;Shilli-Prtidor
   </rdfs:label>
   <madsrdf:hasSource>lcdgt</madsrdf:hasSource>
   <madsrdf:country>France</madsrdf:country>
  -<dcterms:date>
    1839-1907;1839-1907;1839-1907;1839-1907;1839-1907;1839-1907;1839-1907
   </dcterms:date>
   <bibo:translator>DLC</bibo:translator>
   <dcterms:modified>OCoLC;IEN;InNd;OrU</dcterms:modified>
  </rdf:Description>
 -<rdf:Description rdf:about="http://www.authors.org/personal_name/1">
   <madsrdf:idValue>n 86048840 </madsrdf:idValue>
   <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#Identifier"/>
   <ri:recordElement>861121n|azannaabn\\\\\\\\\\|b\aaa\\\\\\c</ri:recordElement>
   <ri:RecordInfo>02459cz a2200409n 4500</ri:RecordInfo>
   <ri:recordContentSource>DLC</ri:recordContentSource>
   <ri:recordChangeDate>20230711150145.0</ri:recordChangeDate>
   <madsrdf:hasAffiliation>NjP</madsrdf:hasAffiliation>
   <madsrdf:Language>eng</madsrdf:Language>
```


Figure 8: Exported RDF/XML file of personal name authority records

Subject	Predicate	Object
person_name URI	Foaf:name	person_name cell
person_name URI	foaf:gender	gender URI
person_name URI	foaf:birthday	dob cell
person_name URI	madsrdf:deathDate	dod cell
person_name URI	dcterms:identifier	viaf_uri cell
person_name URI	dcterms:identifier	lcnaf_uri cell
person_name URI	madsrdf:fieldOfActivity	field_activity1 cell
person_name URI	madsrdf:fieldOfActivity	field_activity2 cell
person_name URI	madsrdf:fieldOfActivity	field_activity3 cell
person_name URI	madsrdf:fieldOfActivity	field_activity4 cell
person_name URI	madsrdf:fieldOfActivity	field_activity5 cell
person_name URI	madsrdf:fieldOfActivity	field_activity6 cell
person_name URI	madsrdf:organization	association_group URI
person_name URI	madsrdf:occupation	Occupation1 URI
person_name URI	madsrdf:occupation	Occupation2 URI
person_name URI	madsrdf:occupation	Occupation3 URI
person_name URI	madsrdf:occupation	Occupation4 URI
person_name URI	madsrdf:occupation	Occupation5 URI
person_name URI	madsrdf:occupation	Occupation6 URI
person_name URI	madsrdf:occupation	Occupation7 URI
person_name URI	madsrdf:language	Language URI
person_name URI	madsrdf:citationSource	source_citation cell
person_name URI	madsrdf:note	information_found cell
person_name URI	madsrdf:MADSType	authority_type cell
person_name URI	ri:RecordID	control_number URI
person_name URI	dcterms:format	description_conventions Cell
person_name URI	madsrdf:fullerName	fuller_form_of_person_name Cell

person_name URI	rdfs:label	person_name_tracing Cell
person_name URI	madsrdf:hasSource	source Cell
person_name URI	madsrdf:country	associated_country Cell
person_name URI	dcterms:date	date_associated_with_name Cell
person_name URI	bibo:translator	transcribing_agency Cell
person_name URI	dcterms:modified	modyfing_agency Cell
control_number URI	ri:RecordInfo	leader cell
control_number URI	ri:recordContentSource	contrl_number_identifier cell
control_number URI	ri:recordChangeDate	Date_transaction Cell
control_number URI	ri:recordElement	fixed_length Cell
control_number URI	madsrdf:idValue	lc_control_number Cell
control_number URI	madsrdf:hasAffiliation	original_cataloging_agency Cell
control_number URI	madsrdf:Language	language_of_cataloging Cell
control_number URI	owl:sameAs	system_control_number Cell
control_number URI	madsrdf:definitionNote	nonpublic_ general_note Cell
control_number URI	madsrdf:hasCorporateParentAuthority	type_of_corporate_body Cell
control_number URI	madsrdf:code	source_of_term Cell
control_number URI	madsrdf:code	source_of_term_occupation Cell
control_number URI	madsrdf:code	source_of_term_gender Cell
gender URI	foaf:gender	gender Cell
language URI	madsrdf:language	language Cell

Table 3: RDF triples of the entire personal name authority dataset

Table 3 shows a comprehensive list of RDF triples (subject, predicate and object) of the entire personal name authority dataset. 49 subjects have been assigned with their predicates and objects.

Visualisation and Exploration of Linked Open Data

After the creation of LOD with links, we need to visualise the LOD for data exploration. The visualisation of linked open data is essential for efficiently and effectively handling billions of objects within dynamic datasets. It allows connecting different datasets (or several nodes and their properties within a single dataset) with different visualisation formats in a dynamic way. The visualisation of LOD can be performed either in tabular views or with a graphical representation that shows the RDF triples with networks. One of the most important components of LOD visualisation is the SPARQL endpoint. LOD simplifies, storing, accessing and integrating machine-readable structured data from different sources, and SPARQL makes it easy to select from and analyse the data from the LOD datasets.

We have used the latest version of AllegroGraph (Version 8.4.1) tool that supports 1 trillion triples, along with Gruff (Version 9.3.1) (https://allegrograph.com/) tool to visualise the linked open data in this research article. This tool supports many types of visualisations, such as table view and graph view. We have explored and visualised personal name authority data in LOD. Gruff is a freely downloadable browser-based web application that does not need installation or downloading any applications. It enables users to visualise and explore suitable knowledge graphs (RDF/Triples) that create and display relationships among datasets that are discovered by the users.

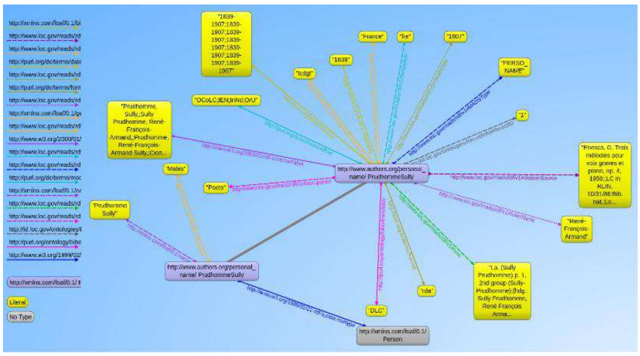


Figure 9: Exploring one RDF triple in a networked graphical view of the personal name authority dataset for a particular subject with URI links

Figure 9 displays a graphical network view of the author "Sully Prudhomme". The subjects, predicates and objects are highlighted in different colours in the networked graphical views of personal name authority. The left side panels display all the RDF schemas and ontologies that are used in the personal name authority dataset.

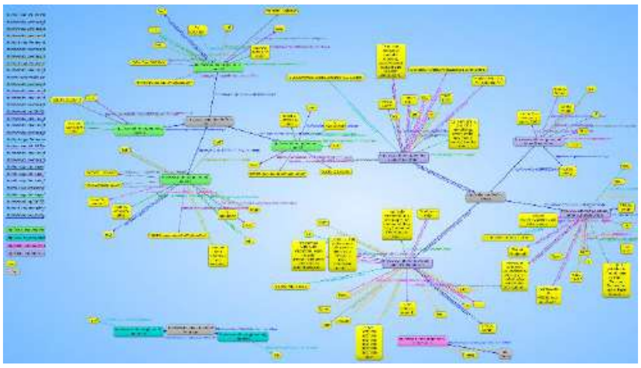


Figure 10: Exploring 100 RDF triples in a networked graphical view of personal name authority datasets

In Figure 10, we have explored the 100 triples in the network with a graphical view of the personal name authority dataset. In this figure, we have shown all nodes of the subject and object in labels and predicates are shown in labels of RDF schemas/ontologies.

Conclusion

This research study is highlighted by the production of a suitable knowledge graph from MARC21 personal name authority records for the purpose of linked open data services. In this context, we have exported the authority records in MARC 21 tags from Koha ILMS datasets. A linked open data service relies on semantic web technologies and its advanced discovery services. This proposed research work, we have performed the conversion of the MARC 21 authority records file (koha.mrc) into a CSV file (koha.csv) by using the MarcEdit tool, and thereafter, we performed the cleaning of the personal name authority records to be used by the OpenRefine tool and also used the Allegrograph and Gruff software for the development of a suitable knowledge graph along with links. Our primary research study aims to utilise a comprehensive methodology that has been identified as a framework for transforming personal name authority into linked open data, along with the creation of relational links. Our research study generated a linked open data (LOD) dataset in RDF format (e.g., RDF/XML, N-triples, Turtle, JSON-LD). However, it generated the LOD files in raw format.

References

Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2012). Linked Data for libraries. *New Library World*, *113*(11/12), 549–570. https://doi.org/10.1108/03074801211282920

Barbera, M. (2013). Linked (open) data at web scale: research, social and engineering challenges in the digital humanities. *DOAJ (DOAJ: Directory of Open Access Journals)*. https://doi.org/10.4403/jlis.it-6333

Berners-Lee, T., Chen, Y., Chilton, L. B., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A., & Sheets, D. A. (2006). *Tabulator: Exploring and Analysing Linked Data on the Semantic Web.* https://api.semanticscholar.org/CorpusID:17241604

Bizer, C., Cyganiak, R., & Heath, T. (2007). *How to Publish Linked Data on the Web.* http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/

Brickley, D., & Miller, L. (2014). FOAF Vocabulary Specification. *XMLNS - Namespace Document.* http://xmlns.com/foaf/spec/

Campbell, L. M., & MacNeill, S. (2010). *The Semantic Web, Linked and Open Data*. https://eclass.aueb.gr/modules/document/file.php/INF180/bibliography/Te_Semantic_Web.pdf

Cannan, J. P., Frank, P., & Hawkins, L. (2019). LC/NACO Authority file in the Library of Congress BIBFRAME Pilots. *Journal of Library Metadata*, *19*(1–2), 39–51. https://doi.org/10.1080/19386389.2019.1589693

Cmcconnell. (2012, September 26). *Transforming Library Metadata into Linked Library Data*. Association for Library Collections & Technical Services (ALCTS). https://www.ala.org/alcts/resources/org/cat/research/linked-data

Cole, T., Han, M., Weathers, W. F., & Joyner, E. (2013). Library Marc Records into Linked Open Data: Challenges and Opportunities. Journal of Library Metadata, 13(2–3), 163–196. https://doi.org/10.1080/19386389.2013.826074

Cole, T. W., Han, M. K., Sarol, M. J., Biel, M., & Maus, D. (2017). Using linked open data to enhance the discoverability, functionality and impact of Emblematica Online. *Library Hi Tech*, *35*(1), 159–178. https://doi.org/10.1108/lht-11-2016-0126

Das, R. (2018). *Linked Open Data (LOD) Management in Digital Library Environment*. Department of Library & Information Science, University of Kalyani.

Downey, M. (2019). Assessing author identifiers: Preparing for a linked data approach to name authority control in an institutional repository context. *Journal of Library Metadata*, *19*(1–2), 117–136. https://doi.org/10.1080/19386389.2019.1590936

Fayyaz, N., Ullah, I., & Khusro, S. (2018). On the Current State of Linked Open Data. *International Journal on Semantic Web and Information Systems*, *14*(4), 110–128. https://doi.org/10.4018/ijswis.2018100106

Frey, J., & Hellmann, S. (2021). FAIR linked data - towards a linked data backbone for users and machines. *Companion Proceedings of the Web Conference 2021*. https://doi.org/10.1145/3442442.3451364

Frosterus, M., Hansson, D., Dadvar, M., Kyriazis, I., & Zapounidou, S. (Eds.). (2021). *Best Practices for Library Linked Open Data (LOD) Publication*. LIBER's Linked Open Data Working Group, LIBER Europe. https://libereurope.eu/article/libers-linked-open-data- working-group-publishes-best-practices-for-library-linked-open-data-lod-publication/

Harris, S., & Seaborne, A. (2013). *SPARQL 1.1 Query Language*. W3C Recommendation. http://www.w3.org/TR/2013/REC-sparql11-query-20130321/

Hofer, M., Obraczka, D., Saeedi, A., Köpcke, H., & Rahm, E. (2024). Construction of Knowledge graphs: current state and challenges. *Information*, *15*(8), 509. https://doi.org/10.3390/info15080509

Idhalama, O. U., & Nwachukwu, P. I. (2025). Assessing the application of artificial intelligence and machine learning technologies in library services in Africa. *Library Management*, *46*(5), 386–400. https://doi.org/10.1108/lm-07-2024-0078

Kar, S., & Das, R. (2020). *Publishing E-resources of Digital Institutional Repository as Linked Open Data: an experimental study*. DigitalCommons@University of Nebraska - Lincoln. https://digitalcommons.unl.edu/libphilprac/4699

Kar, S., & Das, R. (2021). *A Methodology for Transforming MARC21 Personal Name Authority Metadata into Linked Open Data with Integration of VIAF and LCNAF Datasets: an experimental study*. DigitalCommons@University of Nebraska - Lincoln. https://digitalcommons.unl.edu/libphilprac/5458

Saha, B., & Das, R. K. (2020). Producing Linked Open Dataset from Bibliographic Data with Integration of External Data Sources for Academic Libraries. *Library Philosophy & Practice*.https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=8821&context=libphilprac

Schilling,V(2017). Transforming Library Metadata into Linked Library Data: Introduction and

Review of Linked Data for the Library Community, 2003–2011. Retrieved November 01, 2020 from http://www.ala.org/alcts/resources/org/cat/research/linked-data

About Authors

Sri Tapas Roy *(https://orcid.org/0009-0003-0610-5057)*, currently pursuing a PhD (Junior Research Fellow) in DLIS from The University of Burdwan. He has completed B.A (Hons), M.A., and BLIS from the University of Burdwan. He has been awarded an MLIS Degree in DLIS from Kalyani University. He has also completed an MPhil in DLIS from the University of Burdwan. He has published many articles in national and international journals and seminars. He has also qualified for UGC-NET JRF and WBSET. Previously, he has worked as a Librarian at Jaladhar Dey Institute of Education. His research areas include Folksonomy, Subject indexing, Machine learning, Linked Open Data, and Information Seeking Behaviour.

Dr. Rajesh Das ***(**https://orcid.org/0000-0003-3137-864X**)*** PhD, MLISc, MCA, RHCE (Redhat Certified Engineer) is currently working as Assistant Professor at Department of Library and Information Science, The University of Burdwan, Burdwan. He had worked at many reputed institutions. Previously, he served as Librarian at the West Bengal Taxation Tribunal, Kolkata; as Junior Computer Engineer at SNBNCBS, Kolkata; and as Archivist at the Media Lab, Jadavpur University. He has published more than 30 articles in national and international journals and in seminar and conference volumes. He has delivered many invited lectures as a resource person at national-level seminars and workshops. Presently, he supervises four PhD scholars and one M.Phil scholar at the DLIS, the University of Burdwan. His areas of interest are ICT applications in the library and information science domain, open-source technologies, Red Hat/Debian Linux administration and networking, semantic web technologies, and data science and machine learning.

Blind Peer Reviewed
(Previously UGC-CARE)
ISSN: 0972-2750

Manuscript Received on –
07th December, 2025
Primary Reviewed on –
1st January, 2026
Peer Reviewed on –
27th January, 2026
Revised on –
10th February, 2026
Accepted for publication
17th February, 2026

LIS Professionals in the Digital Age: Competencies in Collaborative, Mobile, and AI Technologies

Parbati Pandey & Prof. Margam Madhusudhan

ABSTRACT

Purpose: *This study examines the skills and competencies required by Library and Information Science (LIS) professionals to adopt and utilise emerging and cutting-edge technologies, including collaborative tools, ubiquitous computing applications, and artificial intelligence (AI), within university libraries in Nepal. The aim is to understand the current technological readiness of LIS professionals and identify areas for further professional development to enhance research and academic services.*

Design/methodology/approach: *A survey research design was employed, utilising purposive sampling, to select university libraries across Nepal. A total of 38 library staff members (both professional and semi-professional) were included through census sampling across 12 university libraries: six central universities, one private university, and five deemed academic institutions. The questionnaire was administered through personal visits and email. The collected data were analysed using descriptive statistical techniques in MS Excel.*

Findings: *The results indicate that most LIS professionals demonstrated proficiency in using social networking sites, social media tools, academic social networks, mobile information literacy programs, mobile OPAC, mobile SMS alerts, mobile websites, and AI tools such as ChatGPT. However, competencies were limited in the use of social bookmarking sites, blogs, wikis, QR codes, mobile databases, mobile reference services, mobile library tours, mobile floor plans, and augmented reality applications. Skills related to several emerging technologies were also found to be inadequate. Continuous professional development is necessary to enhance competencies in emerging technologies, including mobile-based and collaborative ones such as augmented reality, podcasts/vodcasts, and RSS feeds. The study highlights the importance of ongoing training and recommends further research employing broader methods and larger populations.*

Dr. Parbati Pandey
Assistant Professor
Dept. of Lib. & Inf. Sc.
Tribhuvan University
parbati.pandey@cdlis.tu.edu.np.in

Dr. Margam Madhusudhan
Professor
Dept. of Lib. & Inf. Sc.
University of Delhi
mmadhusudhan@libinfosci.du.ac.in

Keywords: *Skills and competencies; Mobile technology; Collaborative tools; Emerging technologies; LIS professionals; University libraries; Nepal*

1. Introduction

Libraries must perform various functions to meet the needs of their users. With the advancement of the Internet and technology, the world has become interconnected, prompting libraries to adapt their roles. The rapid development of web-based collaborative ICT applications, mobile-based applications, and AI-integrated applications enhances users' ability to communicate, share, engage, explore, and connect globally. These tools, collectively known as innovative internet-based tools, play a crucial role in libraries. In the context of collaborative tools, Web 3.0 tools have emerged, offering excellent services; however, Web 2.0 tools continue to support libraries significantly by facilitating two-way communication. Web 2.0 applications provide equal opportunities for users to share, retrieve, tag, interact, blog, and self-publish **(Sing, 2021)**. To fully leverage the benefits of Web 2.0, LIS professionals must develop competencies in using social networking sites (SNS) to engage with users, create and organise knowledge, and facilitate interactions through blogs, group postings, and message boards. For example, libraries can use Facebook to promote events and share resources, Twitter for reminders and updates, and SNS features to start discussions and engage with users. Additionally, LIS professionals must master using Academic Social Networking Sites (ASNSs), such as Academia, ResearchGate, Epernicus, and ResearcherID, to connect with the academic community, share interests, and exchange ideas.

Another emerging and widely adopted wireless technology, particularly in mobile devices such as smartphones, tablets, and notepads, has made significant strides. These devices integrate various functionalities, including cameras, scanners, calculators, speech recorders, storage devices, and more, reaching users worldwide. The success of mobile technology can be attributed to its benefits, notably mobility and the ability to provide users with ubiquitous and immersive services. Many software applications are now designed and developed to be compatible with mobile and computer systems. Mobile devices play a crucial role in fostering technology literacy, keeping individuals informed about their surroundings, and bridging the digital divide in society. Many sectors, including libraries, have developed mobile applications to enhance service delivery. Libraries utilise mobile-compatible social media platforms and dedicated library applications to provide a range of services to their users. A note by Duncan notes, "Mobile technology enables libraries to provide cutting-edge services in a form increasingly demanded by users" **(Duncan, 2021)**. Therefore, libraries cannot afford to regress by reducing the adoption of such transformative technology.

Emerging technologies, the latest and most popular, are characterised by novelty, innovation, and advanced capabilities, with significant potential for prototype applications in experimental stages. These technologies offer higher productivity but entail substantial technical and financial risks during initial implementation due to a lack of skills and knowledge, high costs, and technical instability during the development phase. Academic libraries play a pivotal role in adopting effective information technologies to enhance and integrate their specific resources and services. Recently, AI-integrated technologies have gained popularity in the education sector and libraries, sparking discussions among academic circles. These three categorised technologies support libraries, library professionals, and library users across all expected areas of information organisation, sharing, and retrieval; however, LIS professionals must have the technological skills to adopt these applications on library premises. A significant positive relationship exists between librarians' ICT competencies and the use of emerging technologies **(Omehia et al., 2021)**.

2. Review of Literature

In the technological landscape, various AI-powered tools are available to meet users' needs, and library staff must be prepared to provide technology-driven services. **Lateef and Mairaj (2024)** explored Pakistani university librarians' perceptions of virtual reference services (VRS). Web 2.0 tools, which are largely cloud-based, are primarily used to provide resources and services to users. **Kokab et al. (2023)** found that performance, expectancy, effort expectancy, and social influence significantly impacted librarians' behaviour and intention to use social media. Librarians were familiar with the WhatsApp, Facebook, Twitter, and Instagram platforms for personal use, but primarily used Facebook and blogs for professional purposes, as found in the study **(Mogale & Bopape, 2023)**. **Odocha et al. (2023)** examined the use of social networking sites to promote services in Nigeria and found that Facebook, Twitter, WhatsApp, and

Instagram are used to market resources and programs. **Akwang (2022)** found that most librarians are aware of and register with ASNS; however, their usage frequency and satisfaction levels remain low. Insufficient supervisor support, inadequate infrastructure, and low user awareness are the challenges librarians face. However, librarians perceive social media as a valuable and user-friendly tool for service delivery **(Chiparausha et al., 2022)**. **Akwang (2021)** recommendations include adequate ICT resources, organising training programs, and raising awareness about Web 2.0 tools among users and librarians. Mobile OPACs, QR codes, databases, WhatsApp, SMS, and RSS feeds are key tools for enhancing library collections and services. LIS professionals must be skilled in engaging with professional development programs **(Singh & Madhusudhan, 2023)**. **Pandey and Madhusudhan (2023)** discussed mobile-centric library services in India. **Liu and Lewis (2023)** also reviewed the characteristics of mobile users and their service expectations, recommending that libraries adopt user-centric mobile services. **Bajpai and Madhusudhan (2022)** assessed the awareness, usage, and satisfaction levels of LIS professionals with mobile library services. They found that many professionals lacked the necessary skills to utilise mobile technologies effectively. **Dei (2020)** recommended training to enhance the skills needed to handle mobile technology. **Acheampong and Dei (2020)** found a lack of knowledge and resources, recommending staff training to adopt this technology effectively.

Ali et al. (2021) found that Pakistani university librarians are optimistic about AI's ability to enhance information retrieval, automate tasks, and improve the user experience; however, they identified training challenges as a barrier to AI adoption. **Lo (2024)** discusses the impact of AI, particularly GPT-5, on academic librarianship, offering librarians hands-on experience and training in AI. **Anumula et al. (2024)** emphasise that librarians require both professional and technical skills to implement AI tools effectively. **Shahzad et al. (2024)** highlighted the importance of continuous professional development for LIS staff to enhance their AI adoption skills. More advanced AI applications, such as robotics, chatbots, and AI-driven cataloguing tools, remain limited for various reasons, including insufficient training **(Emiri, 2023)**. **Owolabi et al. (2022)** found that most academic librarians need more awareness and knowledge of robotic technology. Emerging technologies offer significant benefits, and librarians must invest in the skills and resources required to adopt them successfully **(Madhusudhan & Pandey, 2023)**. Technical expertise is one of the key obstacles to adopting IoT **(Asim et al., 2022)**. Librarians in Nigerian academic libraries possess moderate ICT skills, and it is recommended that they undergo professional development to enhance their skills and competency as LIS professionals **(Okwu & Nsirim, 2021)**.

Above these several studies, no study has been found in the context of the adoption of these collaborative, mobile-based, and emerging technologies, as well as the skills and competencies required in universities in Nepal.

3. Statement of the Problem

University libraries serve as vital centres of knowledge for researchers, academics, and information seekers. They have long maintained and managed extensive information resources through the purchase and subscription of various resources. These libraries play a crucial role in supporting teaching, learning, and research for both faculty and students. However, the rapid advancement of information and communication technologies (ICT), including collaborative tools, mobile-based applications, and AI-integrated systems, has significantly redefined the responsibilities of library and information science (LIS) professionals. To remain effective in this technology-driven landscape, LIS professionals must develop a broad range of ICT skills and competencies to deliver high-quality, technology-enabled services. In today's "information-on-the-go" era, the success of university libraries largely depends on the ICT proficiency of their LIS staff. In Nepal, university libraries have begun expanding their services to include electronic resources in an effort to meet the growing expectations of their academic communities.

Nevertheless, there remains a pressing need to evaluate how effectively LIS professionals utilise ICT tools and manage emerging technologies, such as mobile platforms and collaborative systems. Despite the increasing importance of ICT, research assessing its use by LIS professionals in Nepal has been limited. This gap underscores a significant disconnect between the availability of advanced ICT services in libraries and the ability of LIS professionals to use them effectively.

4. Objectives of the Study

(i) To assess the skill levels of Library and Information Science (LIS) professionals in using collaborative technologies at select universities in Nepal.

(ii) To evaluate the proficiency of LIS professionals in utilising mobile-based technologies for library services.

(iii) To identify the emerging technology skills exhibited by LIS professionals for enhancing library services.

5. Methodology of the Study

This survey study adopted purposive sampling to select university libraries in Nepal. Based on their ICT-based services and qualified LIS staff twelve universities named as Tribhuvan university, Pokhara University, Mid-West University, Lumbini Buddhist University, Nepal Open University, Rajarshi Janak University, BP Koirala Institute of Health Science, National Academy of Medical Science, Patan Academy of Health Science, Karnali Academy of Health Science, Rapti Academy of Health Science, and Kathmandu University were chosen including six central, five deemed (academy/institute), and one private. Further, a census sampling method was used to select respondents, resulting in 38 LIS professionals participating in this study. The research instruments were administered via email and in-person visits between January and June 2024, resulting in a 100% response rate. Collected data were coded, entered, and analysed using descriptive statistics with MS Excel. The respondents in this study were LIS professionals, semi-professionals, and staff who do not hold professional posts but hold professional academic degrees, such as BLISc and MLISc.

6. Findings of the Study

This section presents the demographic data and proficiency in collaborative, mobile, and emerging technologies, using tables and descriptive text.

6.1 Demographic Information

Demographic data provide the foundation of any research that can influence the core study area. The data is categorised into several key areas, including gender distribution, staff designation, age distribution, appointment status, professional qualifications, and work experience (Table 1).

Sl. No.	Particulars	No. of Responses	Percentage
Gender Distribution			
i.	Male	21	55.00
ii.	Female	17	45.00
Staff Designation			
i.	Assistant Librarian/Library Officer	14	36.84
ii.	Dy. Librarian	10	26.32
iii.	University Librarian	02	05.26
iv.	Asso. Librarian	01	02.63
v.	Other (Miscellaneous)	11	28.95
Age-wise Distribution			
i.	≤ 25 years	00	00
ii.	25 to 35 years	06	15.79
iii.	36 to 45 years	13	34.21
iv.	46 to 55 years	17	44.74
v.	≥55 years	02	05.26
Appointment status of LIS professionals			
i.	Permanent	36	94.74
ii.	On-Deputation	02	05.26
iii.	On Adhoc	00	00
Professional Qualifications			
i.	MLISc	33	86.84
ii.	BLISc.	04	10.53
iii.	PhD	00	00
iv.	MPhil	00	00
v.	Other (MTech)	01	02.63
Work Experience			
i.	21 to 30 years	12	31.58
ii.	10 to 20 years	11	28.95
iii.	≤10 Years	13	24.21
iv.	31 to 40 years	02	05.26

Table 1: Demographic Information of the Respondents (n=38)

Table 1 data reveal that the LIS workforce is predominantly male (55%) with a significant female representation (45%). Most professionals hold mid-level positions, with Assistant Librarians/Library Officers (36.84%) and Deputy Librarians (26.32%) being the most common, while only 5.26% hold the role of University Librarian. The majority (44.74%) are aged 46–55 years, followed by 34.21% in the 36–45 age group, indicating an ageing workforce with limited younger professionals (≤25 years: 0%). Job security is high, with 94.74% in permanent positions. Educational qualifications are primarily at the MLISc level (86.84%), with no PhD or MPhil holders, highlighting a gap in advanced academic credentials. Regarding experience, 31.58% have 21–30 years in the field, while 28.95% have 10–20 years, indicating a well-established workforce but with fewer very senior professionals (5.26% with 31–40 years of experience). These insights suggest the need to attract younger professionals, encourage higher research qualifications, and prepare for leadership succession in university libraries.

6.2 Collaborative skills: Web 2.0/Library 2.0 Tools

Collaborative technology skills refer to the ability to effectively utilise web-based tools for real-time communication, document sharing, and teamwork. These skills enable LIS professionals to enhance service delivery, support virtual collaboration and foster knowledge sharing among users and colleagues. Web 2.0 tools allow interactive communication and improve information sharing in libraries. LIS professionals need to stay updated to use these tools effectively. Respondents rated their skills with Web 2.0/Library 2.0 applications on a 5-point scale, as shown in Table 2.

Sl. No.	Web 2.0/Library 2.0 Tools	Proficiency Rating by LIS Professionals				
		Excellent (%)	Good (%)	Average (%)	Poor (%)	Very Poor (%)
(i)	Social Networking Sites (Facebook, Twitter/LinkedIn)	12 (31.58)	19 (50.00)	05 (13.16)	00	02 (5.26)
(ii)	Social Media Tools (WhatsApp/Viber)	13 (34.21)	17 (44.74)	07 (18.42)	00	01 (2.63)
(iii)	Academic Social Networking Sites (ResearchGate, Academia)	08 (21.05)	17 (44.75)	08 (21.05)	03 (7.89)	02 (5.26)
(iv)	Audio/Video sharing/ Webcasting	07 (18.42)	16 (42.11)	10 (26.32)	02 (5.26)	03 (7.89)
(v)	Social Bookmarking Sites	03 (7.89)	08 (21.05)	15 (39.48)	06 (15.79)	06 (15.79)
(vi)	Blogs	02 (5.26)	14 (36.85)	12 (31.58)	08 (21.05)	02 (5.26)
(vii)	Wikis	01 (2.63)	10 (26.32)	16 (42.11)	07 (18.42)	04 (10.52)
(viii)	Podcast/Vodcast	01 (2.63)	06 (15.79)	08 (21.05)	18 (47.37)	05 (13.16)
(ix)	RSS Feeds	00	09 (23.69)	07 (18.42)	15 (39.47)	07 (18.42)

Table 2: Web 2.0/Library 2.0 Tools (n=38)

Table 2 illustrates that a significant proportion of respondents (81.58%) rated their skills as "Good" or "Excellent" in using social networking sites such as Facebook, Twitter, and LinkedIn. Similarly, 78.95% reported a high level of proficiency with social media tools like WhatsApp and Viber. Moderate proficiency was observed in the use of academic social networking sites and audio/video sharing or webcasting tools, with 65.8% and 60.53% of respondents, respectively, marking themselves as "Excellent" or "Good." However, proficiency noticeably declined for tools such as blogs, wikis, social bookmarking sites, podcasts/vodcasts, and RSS feeds. In particular, 60.53% of respondents rated their skills in podcast/vodcast usage as poor or very poor, while 57.89% reported similarly low confidence in using RSS feeds.

6.3 Mobile-Based Library Services and Skills

The rise of smartphones and integrated applications has reshaped daily activities and transformed library services by enabling access to resources via personal devices (Pandey & Madhusudhan, 2023). Libraries can adopt mobile technology through dedicated apps or common platforms, depending on the ICT skills of their staff. To assess the proficiency of LIS professionals in Nepal regarding mobile-based library application skills, respondents rated their competency on an eleven-point, five-point scale. The results are summarised in Table 3.

Sl. No.	Mobile-based Library Applications	Proficiency Rating by LIS Professionals				
		Excellent (%)	Good (%)	Average (%)	Poor (%)	Very Poor(%)
(i)	Mobile Information Literacy Programmes	08 (21.05)	06 (15.79)	16 (42.11)	05 (13.16)	03 (7.89)
(ii)	Mobile OPAC	07 (18.42)	16 (42.11)	08 (21.05)	04 (10.53)	03 (7.89)
(iii)	Mobile SMS Alerts	06 (15.79)	18 (47.37)	08 (21.05)	05 (13.16)	01 (2.63)
(iv)	Mobile Website	04 (10.53)	18 (47.37)	10 (26.32)	02 (5.26)	04 (10.53)
(v)	Quick Response (QR) Codes	04 (10.53)	16 (42.11)	08 (21.05)	06 (15.78)	04 (10.53)
(vi)	Library App	04 (10.53)	13 (34.21)	12 (31.58)	07 (18.42)	02 (5.26)
(vii)	Mobile Database	04 (10.53)	12 (31.58)	14 (36.83)	04 (10.53)	04 (10.53)
(viii)	Mobile Reference Services	02 (5.26)	16 (42.11)	15 (39.48)	02 (5.26)	03 (7.89)
(ix)	Mobile Library Tours	01 (2.63)	10 (26.32)	14 (36.84)	09 (23.68)	4 (10.53)
(x)	Mobile Floor Plans	00	07 (18.42)	13 (34.21)	13 (34.21)	05 (13.16)
(xi)	Augmented Reality App	00	02 (5.26)	11 (28.95)	16 (42.11)	09 (23.68)

Table 3: Mobile-based Application for Library Services (n=38)

Table 3 shows that the highest self-rated proficiency was observed in Mobile SMS Alerts and Mobile Websites, with 47.37% of respondents rating themselves as "Good." Similarly, Mobile OPAC also showed relatively strong ratings, with 42.11% selecting "Good" and 18.42% rating themselves as "Excellent." In contrast, more advanced tools, such as Augmented Reality Apps, received the lowest proficiency ratings, with no respondents selecting "Excellent" and only 5.26% selecting "Good." In comparison, a significant portion rated themselves as "Poor" (42.11%) or "Very Poor" (23.68%). Likewise, Mobile Floor Plans and Mobile Library Tours reflected limited proficiency, with low "Excellent" and "Good" responses and higher ratings in the "Poor" and "Very Poor" categories. Moderate competence was observed in applications such as Mobile Information Literacy Programmes, Library Apps, and Mobile Databases. However, a large portion of respondents rated themselves as "Average" or below, suggesting a need for further training. For example, 42.11% rated themselves as "Average" in Mobile Information Literacy Programmes, and 36.83% did so in Mobile Databases.

6.4 Emerging Technology and Skills

Emerging technologies, though often experimental and risky, offer great potential, while cutting-edge tools are more stable and productive. In libraries, AI boosts service speed, accuracy, and efficiency. Tools such as ChatGPT, IoT, drones, and blockchain are increasingly utilised for research, data security, and automation. Their success depends on skilled human resources. A survey assessed the proficiency of LIS professionals in Nepal using ten AI-integrated tools, employing a five-point scale. The results are presented in Table 4.

Sl. No.	Emerging Technology Tools	Proficiency Rating by LIS Professionals				
		Excellent (%)	Good (%)	Average (%)	Poor (%)	Very Poor (%)
(i)	ChatGPT	02 (5.26)	20 (52.64)	10 (26.32)	03 (7.89)	03 (7.89)
(ii)	Chatbot	01 (2.63)	06 (15.79)	14 (36.84)	09 (23.68)	08 (21.06)
(iii)	Internet of Things	01 (2.63)	02 (5.26)	14 (36.84)	12 (31.59)	09 (23.68)
(iv)	Bing Chat	00	04 (10.53)	13 (34.21)	14 (36.84)	07 (18.42)
(v)	GPT5	00	04 (10.53)	08 (21.05)	17 (44.74)	09 (23.68)
(vi)	Robotics	00	01 (2.63)	06 (15.79)	14 (36.84)	17 (44.74)

(vii)	Big data applications	00	00	06 (15.79)	15 (39.47)	17 (44.74)
(viii)	Drones	00	00	06 (15.79)	13 (34.21)	19 (50.00)
(ix)	Expert Systems	00	00	05 (13.16)	13 (34.21)	20 (52.63)
(x)	Blockchain	00	00	03 (7.89)	10 (26.32)	25 (65.79)

Table 4: Emerging Technology Tools (n=38)

Table 4 reveals a generally low proficiency among LIS professionals in Nepal regarding emerging technology tools. Among the tools listed, ChatGPT shows the highest familiarity, with 57.9% of respondents rating their proficiency as excellent or good. In contrast, blockchain had the lowest ratings, with 65.79% of professionals marking their skills as very poor and none rating themselves as excellent or good. Similarly, expert systems, drones, big data applications, and robotics received notably low scores, with a majority rating their proficiency as poor or very poor. Tools like GPT-5, Bing Chat, and IoT also showed limited familiarity, with over 60% of respondents rating themselves as poor or very poor. The data indicate a significant need for capacity building and training among LIS professionals in Nepal to effectively adopt and utilise technologies in library settings.

7. Discussion

For library professionals, Web 2.0 skills encompass the proficient use of social media platforms, blogs, wikis, and collaborative tools such as Google Docs and cloud storage services. These competencies enable librarians to engage with patrons, promote library services, and enhance information sharing and collaboration. Library 2.0 adapts the principles of Web 2.0 to the library setting, emphasising user-centred services and participatory technologies. Essential skills in this domain include integrating social media into library operations, managing digital content, utilising online catalogues and databases, and employing interactive tools to foster community engagement.

The findings from Table 2 reveal that LIS professionals in Nepal exhibit strong self-rated proficiency in general Web 2.0 tools, such as social networking sites (81.58%) and social media platforms (78.95%). However, significant skill gaps exist in specialised tools such as RSS feeds, podcasts/vodcasts, social bookmarking, and academic networking platforms. Over half of the respondents rated their skills in podcasts, vodcasts, and RSS feeds as poor or very poor, indicating a need for targeted training. These results align with previous studies by **Bajpai and Madhusudhan (2019)**, **Pandey and Madhusudhan (2023)**, and **Lateef and Miraj (2024)** that observed limited awareness and usage of such tools despite high engagement with social platforms. A lack of institutional policy, time constraints, and staff shortages may further hinder the adoption of tools **(Onuoha et al., 2021; Lavoie, 2022)**. While platforms like Facebook and wikis have seen reasonable adoption, tools such as LinkedIn, expert systems, and blockchain remain underutilised. To bridge this gap, libraries must implement structured training and promote institutional support for collaborative technologies. Strengthening digital competencies in these areas will improve service delivery, foster research collaboration, and enhance user engagement.

Mobile devices play a crucial role in facilitating technology literacy, keeping individuals updated with their surroundings, and bridging the digital divide in society. The findings in Table 3 reveal varying levels of proficiency among LIS professionals in Nepal regarding mobile-based library applications. Most respondents rated themselves as average or good in key areas such as Mobile Information Literacy Programmes (42.11% average, 15.79% good), Mobile OPAC (42.11% good), Mobile SMS Alerts (47.37% good), Mobile Websites (47.37% good), QR Codes (42.11% good), and Mobile Reference Services (42.11% good). However, average ratings dominated responses in Mobile Databases, Library Apps, Mobile Library Tours, and Mobile Floor Plans, indicating moderate familiarity with these topics. Notably, Augmented Reality (AR) apps showed the lowest proficiency, with 42.11% rating their skills as poor and 28.95% as average.

These results align with those of **Bajpai and Madhusudhan (2019)**, who found strong usage of mobile OPACs, library apps, and reference services, but limited familiarity with newer tools such as AR, mobile floor plans, and tours. **Dalili Saleh et al. (2022)** similarly noted challenges in skill and readiness for AR adoption in academic settings.

Despite growing demand for mobile services, the lack of formal planning, policy support, and staff training in Nepalese university libraries has led to inconsistent adoption. While some LIS professionals plan to implement tools like mobile SMS alerts or OPAC soon, these remain underutilised. This points to a critical need for targeted capacity-building initiatives. As per Ranganathan's Fourth Law, "Save the time of the reader", it is imperative that LIS professionals become proficient in mobile technologies to enhance library accessibility and efficiency. Therefore, while basic mobile services, such as SMS alerts and OPAC, are relatively well understood by LIS professionals, more complex or emerging technologies, such as augmented reality and mobile floor plans, reveal significant gaps in proficiency. This indicates a pressing need for targeted capacity-building initiatives to enhance digital literacy and mobile application skills among library staff in Nepal.

Academic libraries play a pivotal role in adopting effective information technologies to enhance and integrate their specific resources and services. Recently, AI-integrated technologies have gained popularity in the education sector and libraries, sparking discussions among academic circles. The findings in Table 4 highlight the varied proficiency levels of LIS professionals in Nepal in relation to emerging technologies. ChatGPT emerged as the most proficiently used tool, with 52.64% of respondents rating their skills as good and 26.32% as average. However, tools like Chatbot and BingChat saw a significant number of respondents (23.68% and 34.21%, respectively) rating their skills as poor, indicating gaps in familiarity. Competency in IoT was mostly average (36.84%), but 31.59% reported poor skills. GPT-4 had 44.74% of respondents rating their skills as poor, and only 10.53% considered themselves proficient. Alarming gaps were found in areas such as robotics, Big Data, Drones, and Expert Systems, where nearly half of the respondents rated their skills as very poor. Blockchain had the highest reported need for improvement, with 65.79% identifying a lack of competency.

These findings align with global studies. **Yoon et al. (2022)** reported that 68% of librarians lacked skills or interest in AI training, and **Emiri (2023)** noted that limited AI adoption in Nigerian libraries stemmed from a lack of skilled personnel. Similarly, **Ali et al. (2021)** found that while basic technologies, such as biometric systems, were in use, advanced AI tools remained underutilised due to a skills gap. The results underscore the urgent need for structured training programs and institutional support to bridge these knowledge gaps. As emerging technologies evolve, LIS professionals must receive continuous, targeted training to remain effective and relevant in the digital age **(Pandey & Madhusudhan, 2023; Bajpai & Madhusudhan, 2021).**

8. Conclusion

University libraries have a vital responsibility to offer innovative, user-centred resources and services. To remain relevant in the digital age, it is essential to maintain two-way communication with users through collaborative and mobile-friendly technologies. Services must be adaptable across various smart devices to meet users' evolving needs. Adopting emerging technologies is crucial for managing and delivering library services more effectively, efficiently, and innovatively. However, the successful integration of these technologies heavily depends on the skills and competencies of LIS professionals.

This study revealed that LIS professionals in Nepal demonstrate strong proficiency in various tools, including social networking sites, social media platforms, academic social networking sites, mobile information literacy programs, mobile OPACs, mobile alerts, mobile websites, QR codes, and ChatGPT. However, proficiency levels varied considerably for other collaborative, mobile-based, and emerging technologies.

To bridge this gap, continuous professional development through targeted training and capacity-building programs is essential. Library authorities and relevant stakeholders must take proactive steps to organise these initiatives and foster an enabling environment that supports the adoption of innovative technologies. Strengthening the technological competencies of LIS professionals is not just a necessity; it is a strategic move toward enhancing the quality and reach of modern library services.

References

Acheampong, E. & Dei, D.-G. J. (2020). Management preparedness towards the implementation of mobile technology library services in academic libraries. *Library Philosophy and Practice (e-Journal)*. Retrieved December 4, 2024, from https://digitalcommons.unl.edu/libphilprac/3967.

Akwang, N. E. (2021). A study of librarians' perceptions and adoption of Web 2.0 technologies in academic libraries in Akwa Ibom State, Nigeria. *Journal of Academic Librarianship*, 47: 102299. https://doi.org/10.1016/j.acalib.2020.102299

Akwang, N. E. (2022). A study on the use of academic social networking sites (ANNSs) by professional librarians in public universities in Akwa Ibom state, *Nigeria. Library Philosophy and Practice (e-Journal)*. Retrieved December 24, 2024, from https://digitalcommons.unl.edu/libphilprac/6962

Ali, M. Y., Naeem, S. Bin & Bhatti, R. (2021). Artificial Intelligence (AI) in Pakistani university library services. *Library Hi Tech News*, 38: 12–15. https://doi.org/10.1108/LHTN-10-2021-0065

Anumula, V. R. B., Anumula, K. R. V. S. S., Anumula, M., & Damerla, S. (2024). Artificial intelligence using library information science professionals. *International Journal of Progressive Research in Engineering Management and Science*, 4: 591-594.

Asim, M., Arif, M. & Rafiq, M. (2022). Applications of internet of things in university libraries of Pakistan: an empirical investigation. *Journal of Academic Librarianship*, 48: 102613. https://doi.org/10.1016/j.acalib.2022.102613

Bajpai, V. K. & Madhusudhan, M. (2019). ICT skills and competencies of library and information science professionals working in college libraries, University of Delhi: A study, *Library Philosophy and Practice (e-journal).* Retrieved December 24, 2024, from https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=5687&context=libphilprac

Bajpai, V. K. & Madhusudhan, M. (2021). Continuous professional development of LIS professionals: an assessment from seventy-four college libraries in Delhi. *Journal of Indian Library Association*, 57: 105–118.

Bajpai, V. K. & Madhusudhan, M. (2022), Mobile-based library service skills and competencies among library and information science professionals working in the college libraries in Delhi. *College Libraries*, 37:1-8.

Chiparausha, B., Onyancha, O. B. & Ezema, I. J. (2022). Factors influencing the use of social media by academic librarians in Zimbabwe: A UTAUT model analysis, *Global Knowledge, Memory and Communication,* 73:142-160. https://doi.org/10.1108/GKMC-09-2021-0151

Dalili Saleh, M., Salami, M., Soheili, F. & Ziaei, S. (2022). Augmented reality technology in the libraries of universities of medical sciences: Identifying the application, advantages and challenges and presenting a model. *Library Hi Tech*, 40: 1782–1795. https://doi.org/10.1108/LHT-01-2021-0033

Dei, D.-G. J. (2020). Assessing the adoption and implementation of mobile technology-based library services in academic libraries. *International Journal of Innovative Technology and Exploring Engineering*, 9:1669–1677. https://doi.org/10.35940/ijitee.C8305.019320

Duncan, A. S. P. (2021). A library at the touch of a finger-tip: An analysis of mobile library services at the University of the West Indies, Mona campus. *Library Hi Tech News*, 38: 14–17. https://doi.org/10.1108/LHTN-06-2020-0059

Emiri, O. T. (2023). Adoption and utilization of artificial intelligence by librarians in university libraries in southern Nigeria. *Library Philosophy and Practice (e-Journal).* Retrieved December 24, 2024, from https://digitalcommons.unl.edu/libphilprac/7570

Kokab, S., Arif, M. and Qaisar, N. (2023). Factors affecting the university librarians' use of social media technologies in Pakistan: A structural equation modelling approach. *Journal of Academic Librarianship*, 49:102719. https://doi.org/10.1016/j.acalib.2023.102719

Lateef, S. & Mairaj, M. I. (2024). Librarians' Perceptions toward virtual reference services in Pakistani university libraries, *Internet Reference Services Quarterly*, 28:55-76.

Lavoie, B. (2022). Library collaboration as a strategic choice: Evaluating options for acquiring capacity. OCLC Research Report. *OCLC Online Computer Library Center*. Retrieved December 24, 2024 from https://www.oclc.org/research/publications/2022/strategic-collaboration/strategic-collaboration-report.html

Liu, Y. Q. & Lewis, S. (2023). Services to mobile users. *Information Technology and Libraries*, 42: 1-25. https://doi.org/10.6017/ital.v42i1.15143

Lo, L. S. (2024). Transforming academic librarianship through AI reskilling: Insights from the GPT-4 exploration program. *The Journal of Academic Librarianship*, 50:102883. https://doi.org/10.1016/j.acalib.2024.102883.

Madhusudhan, M. & Pandey, P. (2023). Innovative horizons: A systematic review of emerging technologies transforming library services. *World Digital Libraries*, 16:119-134.

Mogale, M. G. & Bopape, S. (2023). Trends and issues in the use of social media in academic libraries: experiences from the University of Limpopo subject librarians. South *African Journal of Information Management*, 25: 1580. https://doi.org/10.4102/sajim.v25i1.1580

Odocha, C. P., Udo-Anyanwu, A. J., Opara, G., & Okereke, A. U. T. (2023). Social networking: a marketing strategy for effective service delivery in academic libraries in Imo state. *Library Philosophy and Practice (e-Journal)*. https://digitalcommons.unl.edu/libphilprac/7481

Okwu, E. & Nsirim, O. (2021). Librarians' ICT competencies and utilization of emerging technologies in academic libraries in Rivers State, Nigeria. *Library Philosophy and Practice (e-Journal)*. Retrieved December 4, 2024, from https://digitalcommons.unl.edu/libphilprac/ 5410

Omehia, A., Okwu, E. & Nsirim, O. (2021). Librarians' ICT competencies and utilisation of emerging technologies in academic libraries in Rivers State, Nigeria. Library Philosophy and Practice (e-Journal). Retrieved December 24, 2024, from https://digitalcommons.unl.edu/libphilprac/5410

Onuoha, C., Adeoye, A. A., Obiano, D. C. & Motunrayo, O. (2021). Social media as a collaborative tool for research among library and information science (LIS) professionals in Nigeria. *Library Philosophy and Practice (e-Journal).* https://digitalcommons.unl.edu/libphilprac/5137

Owolabi, K. A., Okorie, N. C., Yemi-Peters, O. E., Oyetola, S. O., Bello, T. O. & Oladokun, B. D. (2022). Readiness of academic librarians towards the use of robotic technologies in Nigerian university libraries. *Library Management*, 43: 296–305. https://doi.org/10.1108/LM-11-2021-0104

Pandey, P. & Madhusudhan, M. (2023). Information and communication technologies influence on librarian's skills in Nepal. *LIS Today*, 9: 1-17.

Pandey, P. & Madhusudhan, M. (2023). Transforming Indian libraries: the mobile and ubiquitous data revolution. In Tyagi, S. and Jha, A. A. (Eds.), *Digital Innovative Practices & Services in Knowledge Management* (13-25). Swami Vivekanand Subharti University, Meerut and Society for Information and Research Studies.

Shahzad, K., Khan, S. A., Iqbal, A. & Javeed, A. M. D. (2024). Identifying university librarians' readiness to adopt artificial intelligence (AI) for innovative learning experiences and smart library services: An empirical investigation. *Global Knowledge, Memory and Communication*: 2514-9342 DOI 10.1108/GKMC-12-2023-0496

Singh, B. P. & Madhusudhan, M. (2023). Mobile apps–based applications in libraries and information centers: a systematic review of the literature and future research agendas. *International Journal of Librarianship*, 8: 83-102.

Singh, G. (2021). Knowledge, attitude, and practices of Web 2.0 applications among engineering students: Implications for academic libraries. *International Journal of Information Studies & Libraries*, 6: 43-50.

Yoon, J. W., Andrews, J. E., & Ward, H. L. (2022). Perceptions on adopting artificial intelligence and related technologies in libraries: public and academic librarians in North America. *Library Hi Tech*, 40: 1893–1915. https://doi.org/10.1108/LHT-07-2021-0229.

About Authors

Dr. Parbati Pandey is an Assistant Professor in the Central Department of Library and Information Science at the Tribhuvan University, Nepal. She has authored two books and twenty research articles. Her areas of interest include systematic review of literature (PRISMA), library automation, knowledge organization, content management, federated search, ICT tools and techniques, and e-resources. She may be contacted at parbati.pandey@cdlis.tu.edu.np.

Dr. Margam Madhusudhan is a Professor of Library and Information Science at the University of Delhi, India. He has held prominent positions such as Deputy Dean of Academics, University Librarian (I/C) and Member of the Academic Council at the University of Delhi. With 26 years of experience in teaching, administration, and research at the university level, his expertise includes designing and evaluating websites, Web-OPACs, ICT in libraries, social networking sites, e-resources, mobile-based library services, and text mining. As the corresponding author, he can be reached at mmadhusudhan@libinfosci.du.ac.in.

Blind Peer Reviewed
(Previously UGC-CARE)
ISSN: 0972-2750

Manuscript Received on –
31st December, 2025
Primary Reviewed on –
7th January, 2026
Peer Reviewed on –
05th February, 2026
Revised on –
11th February, 2026
Accepted for publication
18th February, 2026

Assessing User Perceptions and Satisfaction with Automated Library Services at Rani Lakshmi Bai Central Agricultural University (RLBCAU), Jhansi: A Study

Santosh Kumar Kannaujia, Dr. Madhu Patel, Anubhaw Kumar Suman & Soma Tanti

ABSTRACT:

Introduction: *Technological advances have made automation essential to modern academic libraries, improving efficiency, accessibility, and user satisfaction. RLBCAU has implemented automated library services to enhance operations and user convenience. This study evaluates users' awareness and satisfaction with these services to assess their effectiveness, identify strengths and weaknesses, and offer recommendations for improvement, ensuring the library continues to support education and research effectively.*

Purpose: *This study assesses user satisfaction and awareness of automated library services (Koha, OPAC, e-resources, and RFID self-checkout) and their impact on academics, researchers, and students. Surveys and interviews examine digital literacy, accessibility, and service quality. While users value automation, gaps in training, navigation, and support reduce satisfaction. Recommendations include better training, system upgrades, and stronger technical support to enhance automation benefits.*

Research Problem: *The research addresses gaps in user awareness, satisfaction, and effective use of automated library services at RLBCAU despite their availability.*

Objectives: *The study aims to analyse the use of automated library services, identify limitations faced by users, assess the role of digital literacy and training in user engagement, and evaluate user satisfaction with the accessibility, efficiency, accuracy, and ease of automated library systems.*

Scope: *This survey explores user awareness and satisfaction with automated library services at Rani Lakshmi Bai Central Agricultural University (RLBCAU), Jhansi, Uttar Pradesh.*

Methodology: *This study used the primary data from the users of the Rani Lakshmi Bai Central Agricultural University (RLBCAU), Uttar Pradesh. For this study, researchers use the survey method to collect data through a questionnaire. For this study, 100 questionnaires were distributed among users, out of which only 70 responses were received. To analyse the collected data, researchers use Microsoft Excel.*

Findings: *The study shows that most of the users were undergraduates or postgraduates. Library visits varied in frequency. Library automation awareness was high, with many reporting full automation and positive ratings. OPAC usage was moderate, while a significant number were aware of e-resources, mainly for academic or research purposes. Satisfaction levels were mostly positive.*

Keywords: *Automated Library, Digital Literacy, KOHA, OPAC, RFID Security, User Insights*

Mr. Santosh Kumar Kannaujia
Research Scholar (PhD)
Dept. of Lib. & Inf. Sc.
Mahatma Gandhi Central University
santoshkumarkannaujia1990@gmail.com

Dr. Madhu Patel
Assistant Professor
Dept. of Lib. & Inf. Sc.
Mahatma Gandhi Central University
madhupatel@mgcub.ac.in

Anubhab Kumar Suman
Assistant Librarian
University of Lucknow
sumanubhaw@gmail.com

Ms. Soma Tanti
Research Scholar (PhD)
Dept. of Lib. & Inf. Sc.
Mahatma Gandhi Central University
soma.tanti@gmail.com

Introduction

Technological advancements have revolutionised library services, making automation a key feature in modern academic institutions. Automated systems improve efficiency, accessibility, and user satisfaction by streamlining cataloguing, circulation, and digital resource management. Rani Lakshmi Bai Central Agricultural University (RLBCAU) in Jhansi is a leading agricultural university in India, supporting education and research through its library. To stay current with technology, RLBCAU has adopted automated library services to enhance operations and user convenience. Assessing user awareness and satisfaction with these services is vital for evaluating their effectiveness. Awareness measures how well users understand and utilise automated features, while satisfaction reflects their experience with service quality and ease of use. This study examines these factors at RLBCAU to identify strengths and areas for improvement. By analysing user feedback, challenges, and expectations, the research aims to provide actionable recommendations for optimising automated services. The findings will help refine library automation strategies, ensuring the library remains a central resource for academic and research support.

Statement of the Problem

The swift progress in information and communication technology has led university libraries to switch from manual to automated methods, improving user experience, accessibility, and efficiency. To assist students, researchers, and teachers, Rani Lakshmi Bai Central Agricultural University (RLBCAU) in Jhansi has launched several automated library services. However, user awareness, acceptance, and happiness are key factors that determine how beneficial these services are. Despite the availability of automated systems, challenges such as inadequate user training, limited digital literacy, and technical issues may hinder their optimal utilisation. Additionally, there is a need to assess how well these services meet the expectations and research needs of the academic community at RLBCAU.

Review of Related Literature

Hossain (2024) revealed the immense potential of AI to transform information management and user perceptions in libraries. By automating cataloguing processes, personalising user services, enhancing information retrieval, and utilising predictive analytics, libraries can improve their operations and better meet the needs of their communities. **Moghe, Nagarkar, and Pradhan (2024) examined various** aspects of KOHA implementation, including cost considerations, user experience, technical support, integration capabilities, and customisation options. KOHA is well regarded for its cost-effectiveness, open-source accessibility, usability, and cloud-based hosting. The most commonly utilised Union Catalogue features are the Circulation and OPAC modules. **Suman, Patel and Kishore (2024)** investigated the web presence of state universities of Bihar. The study suggested that website developers must include valuable content for users on their university websites, which will boost the web impact factor by increasing inlinks/backlinks. **Kuhar and Mercun (2022)** investigated the user experience in digital libraries using eye tracking information and discussed that the research on the user experience of digital libraries is lacking, particularly when it comes to the feelings, perceptions, and excitement that are elicited during contact. **Abayomi et al. (2021) used a survey design combining** quantitative and qualitative methods. The study's population consisted of 80 academic librarians from 8 carefully selected university libraries across the nation. **Aparna (2021)** examines the agriculture knowledge management programs created by ICAR, which aim to identify best practices to enhance the production system and increase food production. **Matonkar and Kumar (2021)** explored the computing abilities and library automation of Goa library professionals. **Olubiyo and Awoyemi (2021)** stated that automation is defined in the article as the use of technology to generate and provide goods and services with the least amount of human involvement. **Araya and Mengsteab (2020)** explored an endless amount of digital knowledge will be produced and made available to the globe simultaneously with the aid of the Web-based Digital Library Management System (DLMS). **Basiru and Adebayo (2017)** aimed to discuss the implementation, experience, and usage of the Koha 3.06 library management system (LMS) by Redemption University library personnel and patrons. **Bhoi (2017)** essentially described several technologies and how they are used in library operations. The usage of technology such as RemoteXs, RFID, QR codes, and others to prompt library services has been discussed in the study. **Ranawela and Rajapaksha (2017)** examined how the library's services and resources are used and how they satisfy users. The General Sir John Kotelawela Defence University main library surveyed first- and third-year undergraduate students from the engineering and management faculties, as well as from the social sciences and humanities. **Ojebisi and Choji (2016)** analysed the shortcomings of a library's manual acquisition, cataloguing, and circulation processes and suggested an ILM software to help with the automated processing of these tasks. Visual Basic 6.0 was used to construct the software, while the MSACCESS Relational Database Management System was used to design the database. **Shivakumaraswamy and Narendra (2016)** studied the automation procedure at BG Nagara's BGS Institute of Technology, which is the main subject of the study. For engineering college libraries, library automation is a crucial task.

Objectives of the Study

i. To examine the usage patterns of automated library services, including OPAC, e-resources, self-checkout systems, and KOHA-based library management.

ii. To determine the key limitations users experience in making full use of automated library resources.

iii. To analyse the impact of digital literacy and training programs on user engagement with automated library systems.

iv. To find out user satisfaction with different aspects of library automation, such as accessibility, efficiency, accuracy, and ease of access.

Scope and Methodology

This survey explores user awareness and satisfaction with automated library services at Rani Lakshmi Bai Central Agricultural University (RLBCAU), Jhansi, Uttar Pradesh, India. This study used the primary data from the users of the Rani Lakshmi Bai Central Agricultural University (RLBCAU), Jhansi, Uttar Pradesh. provided data for this study using the survey method to find out about users' Awareness and satisfaction with automated Library Services, a questionnaire was created with the goals in mind. For this survey, 100 questionnaires were distributed among users, out of which only 70 responses were received. Tables and charts were created using Microsoft Word and Excel to analyse and interpret the collected data.

Data Analysis and Interpretation

This particular section consists of data analysis with interpretation:

Gender of the Respondents	Number of Respondens	Percentage of Respondents (%)	X	F	FX	Arithmetic Mean ($\bar{X}$)	Total
1. Male	42	60.00	42	1	42	35	70 (100%)
2. Female	28	40.00	28	1	28		

Table 1: Gender Wise Distribution of Respondents

Table 1 shows that there were 42(60%) male and 28(40%) female respondents who responded to the questionnaire, and an average of 35 users responded to the questionnaire out of the total 100 distributed questionnaires.

Age of the Respondents	Number of Respondents	Percentage of Respondents (%)	X	F	FX	Arithmetic Mean ($\bar{X}$)	Total
1. 18-25 yrs.	67	95.71	67	1	67	35	70 (100%)
2. 26-30 yrs.	3	4.28	3	1	3		

Table 2: Age-Wise Category of the Respondents

Table 2 represents the age-wise category of the respondents. This graph reveals that the maximum number of respondents was between 26 and 30 yrs. 67(95.71%) and only 3(4.28%) respondents were between 18-25 yrs. who responses the questionnaire which I received.

Educational qualifications of the respondents	Number of the Respondents	Percentage of Respondents (%)	X	F	FX	Arithmetic Mean ($\bar{X}$)	Total
1. UG	37	52.85	37	1	37	23.33	70 (100%)
2. PG	28	40.00	28	1	28		
3. Research Scholars	5	7.14	5	1	5		

Table 3: Educational Qualifications of the Respondents

There were 37(52.85%) users who were undergraduate, followed by 28(40%) users who were postgraduate, and 5(7.14%) users who were research scholars, out of the total responses received through distributed questionnaires.

Users Poll	Number of the Respondents	Percentage of Respondents (%)	X	F	FX	Arithmetic Mean ($\bar{X}$)	Total
Yes	26	37.14	26	1	26		
No	8	11.42	8	1	8		**70 (100%)**
Sometimes	21	30.00	21	1	21	14	
Never	1	14.28	1	1	1		
When needed	14	20.00	14	1	14		

Table 4: Visiting the Library by the Users

Table 4 is explained that 26(37.14%) respondents responded Yes that we are visiting the library thereafter 8(11.42%) users responded No we are not visiting the library followed by 21(30.00%) users responded that they visit the library sometimes, 1(14.28%) responded that they never visit the library and 14(20.00%) users said that they visit the library when needed as per the received responses.

Frequency of the users	Numbers of the Respondents	Percentage of Respondents (%)	X	F	FX	Arithmetic Mean ($\bar{X}$)	Total
Per day/Daily	21	30.00	21	1	21		
Twice in a week	12	17.14	12	1	12		
Thrice in a week	18	25.71	18	1	18	14	**70 (100%)**
Once a week	3	4.28	3	1	3		
Whenever needed	16	22.86	16	1	16		

Table 5: Frequency of Users Visiting the Central Library

Table 5 revealed the frequency of visiting the library by the users. There are 21(30.00%) users per day who visit the library; thereafter, 12(17.14%) users visit the library twice in a week, followed by 18(25.71%) users who visit the library thrice in a week. 3(4.28%) users visit the library only once a week, and 16(22.86%) responded that they visit the library when they need to, according to the received responses.

Users Opinion	Numbers of the Respondents	Percentage of Respondents (%)	X	F	FX	Arithmetic Mean ($\bar{X}$)	Total
Yes	50	71.42	50	1	50		
No	4	5.71	4	1	4	17.5	**70 (100%)**
Don't Know	9	12.86	9	1	9		
Not Aware	7	10.00	7	1	7		

Table 6: Users' Opinion on the Condition of Automation of the Central Library

Table 6 explores the users' opinions on the automation of the central library of RLBCAU. There are 50(71.42%) users who said that the central library is automated, followed by 4(5.71%) library users who said that the central library is not automated. 9(12.86%) users told that they do not know that the central library is automated, and 7(10%) library users are not aware of library automation.

Awareness Status	Number of the Respondents	Percentage of Respondents (%)	X	F	FX	Arithmetic Mean ($\bar{X}$)	Total
Yes	46	65.71	46	1	46	35	
No	24	34.28	24	1	24		**70 (100%)**

Table 7: Users' Awareness about Library Automation

Table 7 explains the users' awareness about the automation of the central library of RLBCAU. There are 46(65.71%) users who admit that they are aware about the automation of the central library and only 24(34.28%) users said No that they are not aware about the automation of the central library according to received responses of the distributed questionnaire.

Level of Automation	Number of the Respondents	Percentage of Respondents (%)	X	F	FX	Arithmetic Mean ($\bar{X}$)	Total
Fully Automated	41	58.57	41	1	41	17.5	70 (100%)
Semi-Automated	18	25.71	18	1	18		
Partially Automated	4	5.71	4	1	4		
Not Aware	7	10.00	7	1	7		

Table 8: Level of Automation of Central Library

From table 8 shows the level of automation in central library of RLBCAU. There are 41(58.57%) maximum number of users admit that the central library is fully automated followed by 18(25.71%) users said the central library is semi-automated. 4(5.71%) users said that this central library is partially automated. Lastly, 7 (10%) users responded to the questionnaire and stated that they were unaware of the automation of the central library at Rani Lakshmi Bai Central Agricultural University, Jhansi.

Present Status	Number of the Respondents	Percentage of Respondents (%)	X	F	FX	Arithmetic Mean ($\bar{X}$)	Total
Good	27	38.57	27	1	27	14	70 (100%)
Very Good	34	48.57	34	1	34		
Average	7	10.00	7	1	7		
Neutral	1	14.28	1	1	1		
Well	1	14.28	1	1	1		

Table 9: Present Status of Automation of the Central Library of RLBCAU

Table 9 reveals the status of automation at the central library of RLBCAU. A maximum of 34 (48.57%) users said that the current automation condition in the central library of RLBCAU is very good. Followed by 27(38.57%) users who admit the automation condition of the central library is good, thereafter 7(10%) speak average condition of automation of central library, 1(14.28%) neutral, and 1(14.28%) told that the status of automation in central library is in good condition as per the received responses.

Frequency of Utilisation of OPAC	Number of the Respondents	Percentage of Respondents (%)	X	F	FX	Arithmetic Mean ($\bar{X}$)	Total
Daily	14	20%	14	1	14	10	70 (100%)
Once a day	5	7.14%	5	1	5		
Twice in a day	6	8.57%	6	1	6		
Once a week	13	18.57%	13	1	13		
1-2 times a week	9	12.86%	9	1	9		
Once a month	18	25.71%	18	1	18		
Once a year	5	7.14%	5	1	5		

Table 10: Frequently Utilised the Library's OPAC System by Users

Table 10 examined the frequent use of the OPAC system at the central library of RLBCAU. There are 14(20%) users who daily use the OPAC facility of the central library. 5 (5.71%) users access the OPAC facility of the central library once a day, followed by 6 (8.57%) users who use the OPAC system twice a day. There are 13(18.57%) users who access the OPAC facility once a week. 9(12.86%) users use the OPAC facility of the central library at 1-2 times a week. The majority of 18(25.71%) users use the OPAC facility once a month only for accessing library collections, and 5(5.71%) respondents reported that they are using the OPAC facility in the central library only once a year.

Users Awareness	Number of the Respondents	Percentage of Respondents (%)	X	F	FX	Arithmetic Mean ($\bar{X}$)	Total
Yes	60	85.71	60	1	60	23.33	70 (100%)
No	7	10.00	7	1	7		
Not Aware	3	4.28	3	1	3		

Table 11: Users' Awareness About e-resources

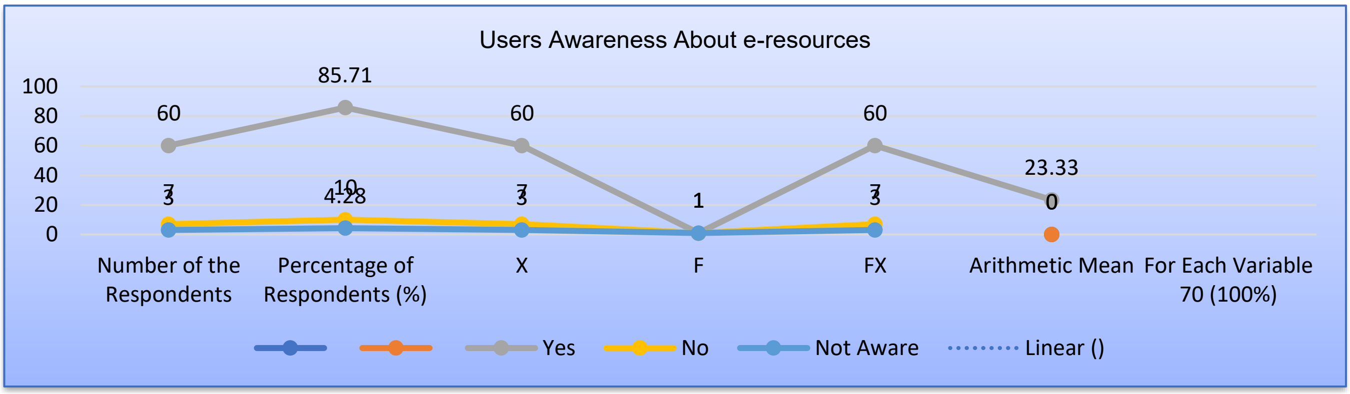


Figure 1 Users' Awareness About e-resources

Table 11 shows that a maximum of 60(85.71%) users were aware of the e-resources service provided by the central library, followed by 7(10%) users who responded " No and 3(4.28%) users who responded that they were not aware of the e-resources provided by the central library.

Purposes of accessing e-resources	Number of the Respondents	Percentage of Respondents (%)	X	F	FX	Arithmetic Mean ($\bar{X}$)	Total
Academic	39	55.71	39	1	39	14	70 (100%)
Research work	16	22.86	16	1	16		
Writing the papers	3	4.28	3	1	3		
Qualifying competitions	8	11.43	8	1	8		
Keep up-to-date on the subject	4	5.71	4	1	4		

Table 12: Purposes of Accessing e-resources

Table 12, discussed by the bar graph, provides a clear insight into the various purposes for which users access e-resources. Most respondents, 39(55.71%), indicated that their primary purpose is academic, highlighting the dominant role of digital resources in supporting coursework, learning materials, and classroom-based study. This is followed by 16(22.86%) of users who use e-resources for research work, reflecting the importance of online information in scholarly investigations and project-based studies. A smaller proportion, 8(11.43%), accesses these resources to prepare for qualifying competitions, suggesting that e-resources also serve as a tool for career advancement and competitive exam preparation. Only 4(5.71%) of respondents use e-resources to stay up-to-date with their subject, indicating limited usage for current awareness and professional development. Lastly, just 3(4.28%) reported using e-resources for writing papers.

Status of Users Satisfaction	Number of the Respondents	Percentage of Respondents (%)	X	F	FX	Arithmetic Mean ($\bar{X}$)	Total
Fully Satisfied	21	30.00	21	1	21	11.67	70 (100%)
Satisfied	29	41.43	29	1	29		
Partially Satisfied	13	18.57	13	1	13		
Very Satisfied	2	2.86	2	1	2		
Dissatisfied	1	1.43	1	1	1		
Neutral	4	5.71	4	1	4		

Table 13: Satisfaction Status of Users

There is a maximum of 29(41.43%) users satisfied with the services of the central library of RLBCAU. Thereafter, 21(30.00%) users responded that they are fully satisfied with the central library facilities and services that are provided to the patrons in an automated library environment. Meanwhile, 13(18.57%) of users are partially satisfied with automated library services. Followed by 1(1.43%) user was dissatisfied with the automated library environment, and 4(5.71%) respondents responded as neutral with automated library services.

Limitations of the Study

Despite the significance of this study in assessing user awareness and satisfaction with automated library services at Rani Lakshmi Bai Central Agricultural University (RLBCAU), Jhansi, certain limitations must be acknowledged:

i. The study may include only a specific group of users, i. e., Undergraduate, Postgraduate, and Research Scholars from RLBCAU, to collect data regarding the automated central library of Rani Lakshmi Bai Central Agricultural University.

ii. The purpose of this survey is to find out the type of facilities and services provided to the users of RLBCAU through the central library.

iii. Based on survey and interview replies, user awareness and satisfaction levels could be influenced by social desirability effects, personal biases, or misunderstandings.

iv. The study may not examine all potential automation technologies utilised in contemporary libraries, instead concentrating on automation tools (such as KOHA, RFID, and OPAC).

v. The research is conducted within a specific time frame, which may not capture long-term changes in user awareness and satisfaction as library services evolve.

Findings of the Study

i. There are 60% male and 40% female respondents who responded to the questionnaire for this study.

ii. The maximum 95.71% users are under the age of 25 years, and 4.28% users are above the age of 30 years at RLBCAU, Jhansi.

iii. There are 52.85% users who are undergraduates. 40% users are post-graduate, and 7014% users are research scholars at RLBCAU, Jhansi.

iv. **Library Visits**: 37.14% visit regularly, 11.42% don't, 30% sometimes, 14.28% never, and 20% only when needed.

v. **Visit Frequency**: 30% daily, 17.14% twice a week, 25.71% thrice weekly, 4.28% once weekly, and 22.86% as needed.

vi. **Library Automation Awareness**: 71.42% confirmed automation, 5.71% denied it, 12.86% didn't know, and 10% were unaware.

vii. **Automation Awareness**: 65.71% knew about it, while 34.28% did not.

viii. **Automation Status**: 58.57% said fully automated, 25.71% semi-automated, 5.71% partially, and 10% were unaware.

ix. **Automation Condition**: 48.57% rated it very good, 38.57% good, 10% average, and 14.28% neutral/well.

x. **OPAC Usage**: 20% daily, 5.71% once daily, 8.57% twice daily, 18.57% weekly, 12.86% 1-2 times weekly, 25.71% monthly, and 5.71% yearly.

xi. **E-Resources Awareness**: 85.71% knew about them, 10% did not, and 4.28% were unaware.

xii. **E-Resource Purpose**: 55.71% academic, 22.86% research, 11.43% competitions, 5.71% subject updates, and 4.28% paper writing.

xiii. **Satisfaction Level**: 41.43% satisfied, 30% fully satisfied, 18.57% partially, 1.43% dissatisfied, and 5.71% neutral.

Conclusion

Rani Lakshmi Bai Central Agricultural University is situated in the district of Jhansi of the Indian state of Uttar Pradesh. This central university is purely dedicated to the sector of Agriculture. In conclusion, the study on user awareness and satisfaction with automated library services at Rani Lakshmi Bai Central Agricultural University (RLBCAU), Jhansi, highlights the critical role of technology in enhancing library experiences. The findings underscore the importance of user-friendly automated systems in improving accessibility, efficiency, and overall satisfaction among library users. While the adoption of such services has shown positive outcomes, the study also identifies areas for improvement, such as increasing awareness programs and addressing technical challenges to ensure optimal utilisation. In this study, it was found that 37.14% users visited the library on a regular basis. There were 65.71% users aware of library automation in the central library. There were 4857% users who responded that the automation condition in the central library is very good. 18.57% library users were fully satisfied with the library automation facility in the central library.

References

Abayomi, O. K., Adenekan, F. N., Abayomi, A. O., Ajayi, T. A., & Aderonke, A. O. (2021). Awareness and perception of artificial intelligence in the management of university libraries in Nigeria. *Journal of Interlibrary Loan, Document Delivery & Electronic Reserve*. 29(1-2): 13-28.

Aparna, V. R. (2021). Agricultural knowledge management initiatives by ICAR: An Overview. *Journal of Agricultural Extension Management*. 22(2): 175.

Araya, T. W., & Mengsteab, A. (2020). Designing web-based library management system. *International Journal of Engineering Research & Technology*. 9(10): 272-277.

Basiru, A., & Adebayo, E. L. (2017). Implementation and utilization of an open-source library management system: Experience with Koha 3.06 at Redeemers University Library, Nigeria. *Journal of Information Engineering and Applications*. 7(9): 10-16.

Bhoi, N. K. (2017). Use of information communication technology (ICT) and library operation: An overview. In: International Conference on Future Libraries: From Promises to Practices, Bangalore, India, 2017, 445-456.

Hossain, H. (2024). AI technology applications in library science: Transforming information management and user experience. *College Libraries*, 39(4): 62-69. https://collegelibraries.in/index.php/CL/article/view/181

Kuhar, M., & Merčun, T. (2022). Exploring user experience in digital libraries through questionnaire and eye-tracking data. *Library & Information Science Research*. 44(3): 101175.

Matonkar, P. V., & Kumar, M. (2021). Study of library automation and computing skills of library professionals in Goa. *Library Philosophy and Practice*. 6140.

Moghe, G., Nagarkar, S., & Pradhan, A. (2024). Assessment of KOHA open-LMS at MKSSS group libraries. *College Libraries*. 39(4): 53-61. https://collegelibraries.in/index.php/CL/article/view/180

Ojebisi, I., & Choji, G. D. (2016). Design and implementation of a computerized library system for the National Institute of Policy and Strategic Studies, Kuru Nigeria. FULafia *Journal of Science and Technology*. 2(1): 84-88.

Olubiyo, P. O., & Awoyemi, R. A. (2021). Automation of academic libraries in Nigeria: Issues and practices. *Library Philosophy and Practice (e-journal).* 5613.

Ranawella, T. C., & Rajapaksha, M. P. (2017). Use and user satisfaction with library services and resources at the main library of the General Sir John Kotelawala Defence University, Sri Lanka. *Journal of the University Librarians Association of Sri Lanka*. 20(2).

Shivakumaraswmay, K. N., & Narendra, B. K. (2016). Library automation: A case study of the Central Library of BGS Institute of Technology, BG Nagara. *Journal of*

Suman, A. K., Patel, M., & Kishore, A. (2024). Web presence of state universities of Bihar, India: A webometric analysis. *College Libraries*. 39(3): 23-32. http://collegelibraries.in/index.php/CL/article/view/165

About Authors

Santosh Kumar Kannaujia is currently pursuing his PhD in the Department of Library and Information Science at Mahatma Gandhi Central University, Bihar. He has a strong academic background, having completed both his Bachelor's and Master's degrees in Library and Information Science from Banaras Hindu University. In addition, he earned a B.A. (Hons.) in Sociology from the same institution. He qualified the UGC-NTA Junior Research Fellowship (JRF) in July 2019. Demonstrating his research aptitude, he has published two papers in peer-reviewed refereed journals and has presented four papers at national and international conferences and seminars. He has also actively participated in several online and offline workshops and training programmes, reflecting his continued engagement with academic and professional development.

Dr. Madhu Patel is serving as an Assistant Professor in the Department of Library and Information Science at Mahatma Gandhi Central University, Bihar. She completed her PhD from Babasaheb Bhimrao Ambedkar University (BBAU), a Central University in Lucknow, in 2018. Currently, she is guiding four PhD scholars, and under her supervision, one M.Phil. dissertation and nineteen Master's level dissertations have been successfully completed. In recognition of her contributions, she has received a Prashasti Patra from the Ministry of Information and Broadcasting, Government of India, as well as a Certificate of Appreciation from the Department of Library and Information Science, BBAU, Lucknow. She has published more than sixteen research articles in UGC CARE-listed and refereed journals, contributed six book chapters, and presented seven papers in national and international conference proceedings. Additionally, she has presented over twenty papers at national and international seminars and conferences. Dr. Patel has also served as an External Examiner and delivered invited talks at the university level. Within the university, she holds several important administrative and academic roles, including Associate Dean of Students' Welfare and Assistant Librarian (In-charge). She is also an active member of various committees such as the MHRD programme 'Ek Bharat Shreshtha Bharat', G20 University Connect Programme, Equal Opportunity Cell, Departmental Research Committee (DRC), Board of Studies, and the School Board. She is a life member of professional bodies including the Indian Library Association, CS&IRS, BBAU Alumni Association, and the Central Government Library Association. Furthermore, she has contributed in various capacities such as Chairperson, Convener, Guest of Honour, Keynote Speaker, and Coordinator at several national and international webinars, seminars, committees, and academic competitions.

Mr. Anubhaw Kumar Suman is presently serving as an Assistant Librarian at the University of Lucknow, Uttar Pradesh. He is UGC NET/JRF qualified and a recipient of the National Fellowship for Doctoral Research. He completed his MLIS, BLIS, and B.A. (Hons.) degrees from Patna University, Patna, and also holds a PGDCA from Mahatma Gandhi Antarrashtriya Hindi Vishwa Vidyalaya, Wardha. He has gained professional experience working at prestigious national-level institutions such as the Indian Institute of Technology (IIT) Delhi and the National Inland Navigation Institute (NINI), Patna. He has published several research papers in reputed national and international journals and is a life member of various professional bodies and associations in the country. His academic profiles can be accessed through his Vidwan profile (https://vidwan.inflibnet.ac.in/profile/572148) and ORCID (https://orcid.org/0000-0002-4199-799X).

Soma Tanti is a PhD scholar in the Department of Library and Information Science at Mahatma Gandhi Central University. She possesses a strong academic background, having completed both her Bachelor's and Master's degrees in Library and Information Science from the University of Delhi. In addition, she earned a B.A. (Honours) in Geography from Miranda House and an M.A. in Geography from the Delhi School of Economics. She qualified the UGC NET in Library and Information Science in 2017 and was awarded the NFSC in 2019. Her research contributions include four published research articles and one book chapter. She has also actively participated in the academic community by presenting four papers at national conferences.

Blind Peer Reviewed
(Previously UGC-CARE)
ISSN: 0972-2750

Manuscript Received on –
21st January, 2026
Primary Reviewed on –
10th February, 2026
Peer Reviewed on –
22nd February, 2026
Accepted for publication
9th March, 2026

Application of Artificial Intelligence for Library Document Classification: A Systematic Review

Moirangthem Sajan Singh, Rajesh R. Aldarthi, Manoj Kumar Sinha, & Luckey Pathan

Mr. Moirangthem Sajan Singh
Research Scholar (PhD)
Dept. of Lib. & Inf. Sc.
Assam University
sajanmoirangthem@gmail.com

Dr. Rajesh Rangappa Aldarthi
Assistant Professor
Dept. of Lib. & Inf. Sc.
Assam University
rajeshaldarthi@gmail.com

Dr. Manoj Kumar Sinha
Professor
Dept. of Lib. & Inf. Sc.
Assam University
manoj.kumar.sinha@aus.ac.in

Mr. Lucky Pathan
Library Trainee
OP Jindal Global University
luckeypathann@gmail.com

ABSTRACT

Artificial Intelligence is fundamentally transforming and reshaping the way we live and work. AI can also be adopted in libraries for technical services such as classification and cataloguing, and for library management, such as staffing and decision-making. This study is guided by the PRISMA framework to analyse studies on the use of machine learning or artificial intelligence in library classification. A quantitative approach was adopted to assess the applicability of machine learning or artificial intelligence (AI) for automating the classification of DDC, UDC, LCC, and CC. The research articles were retrieved from the SCOPUS, Web of Science, and Google Scholar databases. The search results found 535 documents in Scopus, 99 in WOS, and 473 in Google Scholar. The researcher identified 27 documents as closely related and retrieved their full texts for review. This analysis revealed that the DDC scheme is commonly used to automate classification. SVM and BERT are the most popular models among the studies. Only 20 of the 27 studies provided formal evaluation metrics such as accuracy, precision, recall, and F1-score. Highest accuracy found in Dewey Decimal Classification (DDC) with Support Vector Machine (SVM) model, with F1 scores of 0.80 to 0.85. The challenges in achieving fully automated classification included insufficient datasets, the complexity of classification schemes, and the technical limitations of machine learning models. This research has made a valuable contribution to the existing body of knowledge. It is the first systematic review to examine machine learning in library classification quantitatively. It has also provided real-world problems for the application of machine learning or AI technologies in Classification.

Keywords: *AI, Artificial Intelligence, Machine Learning, library classification, Automatic classification, Systematic Review*

Introduction

The advent of Information and Communication Technology (ICT) has transformed how libraries manage resources and deliver services. In developing countries, libraries actively automate the household operations to meet the growing and modern information needs of users (Komolafe-Opadeji & Ojo, (2019) [1]. In India, regulatory bodies such as AICTE emphasise the adoption of library automation. Their guidelines highlight the need to automate library functions, including acquisition, technical processing, circulation, and serials control, to enhance service in academic libraries. Library classification is one of the backbones of academic, public, and research libraries. Classification enables the arrangement of books, journals, monographs, handbooks, gazetteers, and other holdings in a systematic order. Classification schemes such as the Dewey Decimal Classification (DDC), Library of Congress Classification (LCC), Colon Classification (CC), and Universal Decimal Classification (UDC) are the most popular classification schemes adopted globally (Lund, 2023) [2]. Thanks to classificationists who developed these schemes, classifiers and catalogers can classify vast resources through structured hierarchies and controlled vocabularies (Hjørland, 2017) [3].

Traditionally, trained classifiers manually classify library documents according to the organisation's classification scheme. This is labour-intensive, difficult to scale, and requires highly trained classifiers with in-depth knowledge of classification schemes. With the exponential growth of information, the manual classification of books and journals is time-consuming, and more effort is required to provide a relevant and accurate classification number (Palve & Arora, 2025) [4]. For instance, the National and University Library of Slovenia digitised over 850,000 items. But many of these documents are not classified. Several challenges hampered the classification process. Among these, the complexity of the classification scheme (DDC, UDC, LCC, and CC) made it difficult for classifiers to assign accurate class numbers when libraries have large collections.

The initial studies employed rule-based systems and string-matching methods, but these approaches often failed to assign accurate notations to complex titles. However, machine learning, the data-driven algorithms, assist to learn patterns and making predictions with higher accuracy. It supports real-world problem-solving, such as spam filtering in email, product and movie recommendations, speech recognition, text mining, and image classification. Models such as k-nearest neighbours (KNN), naïve Bayes (NB), and support vector machines (SVM) produce promising results in assigning classification codes to text data (Ullah & Ullah, 2017; Kragelj & Kljajić Borštnar, 2021) [5][6]. Recently, deep learning and natural language processing (NLP) models, such as Bidirectional Encoder Representations from Transformers (BERT), have enhanced the understanding and semantic analysis of texts. The National Library of Sweden trained a Swedish BERT model (KB-BERT) using over 3.5 billion words from its national collections to automate the classification (Kragelj & Kljajić Borštnar, 2021) [6]. Kragelj and Borstnar developed a supervised model and trained it on scholarly texts to assign UDC codes to over 200,000 older Slovenian texts [6]. This opens the door to automating classification in libraries. This review systematically analysed published research to identify the Machine Learning models employed, the classification schemes targeted (e.g., DDC, UDC), the corpora used for training and testing, and the reported accuracy of the models in automating library classification.

Literature Review

A systematic literature review (SLR) on AI adoption in libraries using the PRISMA framework, Harisanty et al., examined AI implementation across public and academic libraries [7]. Their review found that chatbots, expert systems, and intelligent classification modules are gradually emerging and adoption in libraries remains in its early stages. The study also highlights the limited infrastructure, readiness, and professional capacity. Similarly, Marasinghe et al. conducted a systematic review of 88 studies from LENS.ORG and Google Scholar identified three dominant thematic clusters [8]. AI tools and applications, academic integration of AI, and AI-enabled service delivery. Martínez Concha et al. studied 56 articles from the WoS, highlighting the necessity of ethical awareness, user immersion, and blending between technology and human-centred service design in AI-enabled library services [9]. Shahzad et al. also conducted a systematic review on the application of robotic technologies in academic libraries [10]. Robotics technology would play a strategic role in automating repetitive, time-consuming tasks, including shelving and sorting books. Also, can assist patrons with disabilities and provide 24/7 customer support. Adesina and Zubairu conducted a literature review of studies in Google Scholar to examine the uses and challenges of artificial intelligence in modern libraries [11]. Areas where AI can improve in library functions include technical assistance, reference services, circulation, resource management, and information retrieval. Supervised machine learning algorithms, such as support vector machines (SVM), k-nearest neighbours (k-NN), and transformer models such as BERT, are increasingly deployed to assign subject headings and class numbers. Chou and Chu demonstrated the effectiveness of the BERT model for assigning Library of Congress Subject Headings in public-domain texts with high precision [12]. In a related context, Yunus et al. emphasised the role of AI in supporting innovative technology frameworks that encompass data mining, RFID tracking, and semantic classification to enhance physical and digital collections[13].

Objectives:

1. To identify AI, machine learning (ML), or natural language processing (NLP) models applied in library classification.
2. To assess the transition of the methodology of automatic and machine learning applications in library classification over time.
3. To review datasets employed for training and testing.
4. To review the accuracy of adopted machine learning models in library classification.
5. To find the limitations in applying AI to Library classification.

Methodology:

The researcher adopted PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to conduct this systematic review. Research articles are retrieved from three databases viz Web of Science (WoS), Scopus, and Google Scholar. Advanced search query was constructed with different keyword such as *"artificial intelligence," "machine learning," "automatic classification," "library classification," "bibliographic metadata,"* and classification schemes *"DDC," "UDC," "LCC,"* and *"Colon Classification"* along with Boolean operators (*AND*, and *OR*) in combination with search fields *TITLE-ABS-KEY* in Scopus and *TOPIC* from WoS. *Publish or Perish*, a software program that can retrieve and analyse academic citations, was used to retrieve relevant studies from Google Scholar using the same keywords.

Researchers searched in March 2025, without restrictions on publication year. Only peer-reviewed research articles published in the English language were considered for screening. Initially, 5,325 documents from Scopus, 99 documents from Web of Science, and 473 documents from Google Scholar were retrieved. The researchers screened the titles and abstracts of all retrieved documents to determine relevance. Based on this preliminary screening, 27 studies from Scopus, 19 from WoS, and 17 from Google Scholar were selected for full-text assessment. After removing duplicates and excluding unrelated studies, 27 studies were included in the systematic review (Figure 1).

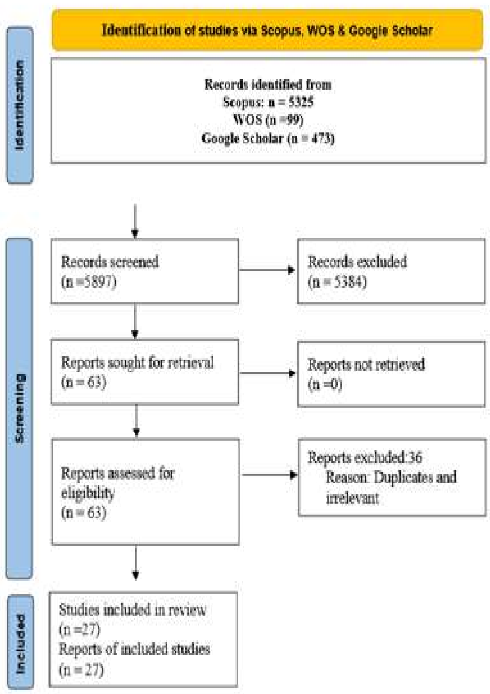


Figure 1 Flow chart of the methodology adopted (PRIMSA)

Inclusion and Exclusion Criteria

Inclusion: This review includes published research articles that examined the application of artificial intelligence (AI) in library classification using formal schemes such as the DDC, UDC, LCC, CC, and CLC. Only English-language studies were considered, with no restrictions on publication year.

Exclusion: The researchers excluded proceedings, reviews, and book chapters. If the studies focused exclusively on indexing, clustering, or keyword extraction, or lacked clear methodological details regarding machine learning or AI models and classification schemes, they were also excluded.

Result

Implemented classification schemes and Models in Automatic Library Classification.

Dewey Decimal Classification (DDC) is the predominant focus of automated classification research, appearing in 18 of the 27 studies examined (Figure 2). Following DDC, Universal Decimal Classification (UDC) and Library of Congress Classification (LCC) appeared 5 times each. Notably, only one study used the Colon Classification (CC) system, and another employed multiple schemes, including the DDC, LCC, UDC, and faceted classification. This distribution suggests a clear preference for traditional hierarchical classification systems, such as the DDC, within AI-based implementations.

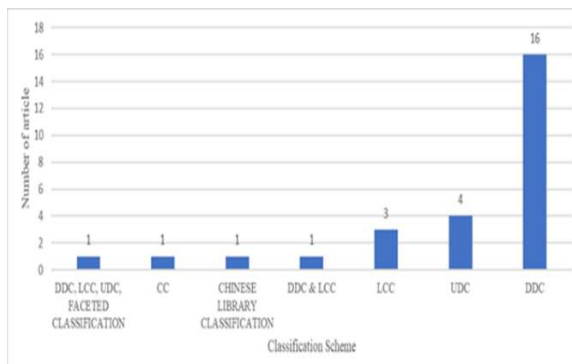


Figure 2 Classification schemes adopted among the studies

Figure 3 shows the frequencies of methods and models used across studies. Support Vector Machine (SVM) appeared in seven different works. This dominance suggests that SVM is a popular choice for text classification tasks in library classification, likely due to its efficiency in handling high-dimensional data typical of bibliographic records. BERT Models appeared in 5 documents, indicating a significant shift toward Natural Language Processing (NLP) techniques. It indicates that researchers are moving away from traditional statistical methods toward deep learning architectures that can better capture the semantic context of book titles and abstracts.

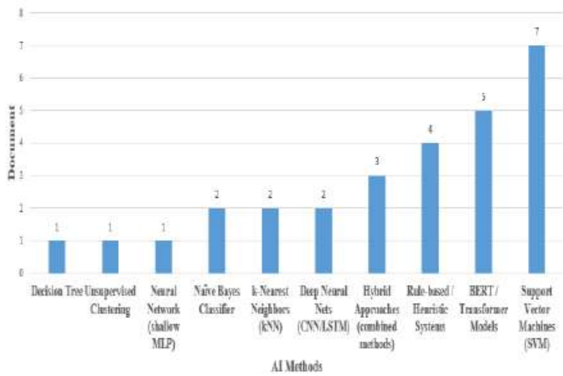


Figure 3: methods and models used across studies

Rule-based and Heuristic Systems appeared in four documents; these entries largely match the initial phase of automation research in the late 1990s and early 2000s. This suggests that while rule-based logic established the foundation for automated classification. Various other methods, such as Deep Neural Nets (e.g., CNNs and LSTMs), Naïve Bayes, and k-Nearest Neighbours (kNN), also appear in the studies. Finally, researchers used Decision Trees, Unsupervised Clustering, and Shallow Neural Networks as comparative baselines rather than primary solutions.

Methodological Transitions in Automatic Library Classification

Table 1 highlights studies worldwide that automate library classification using major classification schemes, such as DDC, UDC, LCC, and CC. During the 1980s and 1990s, Garland (1983) and Jenkins et al. (1998) adopted rule-based and similarity-driven approaches, particularly in DDC and LCC, to automate the classification process [14],[15]. In the Early 2000s, Frank & Paynter (2004) and Wang (2009) adopted supervised machine learning, such as Support Vector Machines and Naïve Bayes models, for DDC and LCC classification [17],[18]. During the 2010s, the introduction of hybrid and weakly supervised methods was observed in the study of Lin et al (2017) [19]. The transition towards deep learning and transformer-based models, such as fastText and BERT, to capture greater semantic complexity is shown in the works of Golub et al. (2020) and Borovic et al. (2022) [20],[31]. Still, no single machine learning model can solve the problems of library classification. Golub et al. (2024) and Mahmud (2024) employed multiple models, including SVM, fastText, and ensemble learning, to enhance robustness and stability in automation [36],[21].

Year	Reference	Title	Models / Techniques Used	Scheme
1983	Garland (1983) [14]	*An Experiment in Automatic Hierarchical Document Classification*	Agglomerative clustering using keyword similarity	LCC
1998	Jenkins et al. (1998) [15]	*Automatic Classification of Web Resources Using Java and Dewey Decimal Classification*	Rule-based classification, Dice similarity, hierarchical filtering	DDC
1999	Jenkins et al. (1999)[16]	*Automatic RDF Metadata Generation for Resource Discovery*	Rule-based keyword matching	DDC
2001	Shafer (2001) [22]	*Automatic Subject Assignment via the Scorpion System*	Information Retrieval ranking, DDC matching	DDC
2002	Kim & Lee (2002) [23]	*Designing a Knowledge Base for Automatic Book Classification*	Rule-based system, structured knowledge base, PMEST facets	CC

Year	Reference	Title	Models / Techniques Used	Scheme
2004	Frank & Paynter (2004)[17]	*Predicting Library of Congress Classifications from LCSH*	Hierarchical SVM, binary classifiers	LCC
2006	Slavic (2006) [24]	*UDC in Subject Gateways: Experiment or Opportunity*	Rule-based NLP, keyword matching, and stemming	UDC
2008	Pong et al. (2008) [25]	*A Comparative Study of Two Automatic Document Classification Methods*	k-NN, Naive Bayes	LCC
2008	Joorabchi & Mahdi (2008) [26]	*Bootstrapping an Automatic Text Classification System*	Naive Bayes (TWCNB), SVM (LIBLINEAR)	DDC
2009	Wang (2009) [18]	*An Extensive Study on Automated Dewey Decimal Classification*	Naive Bayes, SVM, interactive learning	DDC
2009	Mehler & Waltinger (2009) [27]	*Enhancing document modelling by means of open topic models: Crossing the frontier of classification schemes in digital libraries by example of the DDC.*	SVM	DDC
2011	Joorabchi & Mahdi (2011) [26]	*An unsupervised approach to automatic classification of scientific literature utilising bibliographic metadata*	Citation-network analysis (BB-ATC)	DDC
2011	Hou & Xue (2011) [28]	*Knowledge Base for Automatic Indexing Based on CLC*	Rule-based mapping, Dice similarity, thesaurus matching	*CLC
2013	Joorabchi & Mahdi (2013) [29]	*Classification of Scientific Publications Using Controlled Vocabularies*	Wikipedia Miner, semantic similarity, MARC/WorldCat matching	DDC
2015	Khoo et al .(2015) [30]	*Augmenting Dublin Core Metadata with DDC*	Rule-based classification, Lucene TF weighting	DDC
2017	Lin et al. (2017) [19]	*Mapping Metadata to DDC Classification Structures*	Vector similarity mapping (DISTIL)	DDC
2020	Golub et al. (2020) [20]	*Automatic Classification of Swedish Metadata Using DDC*	SVM, Naive Bayes, CNN, RNN, BM25, TF-IDF	DDC
2021	Kragelj & Kljajić Borštnar[7] (2021) [6]	*Automatic Classification into UDC*	SVM, MLP, Logistic Regression, Naive Bayes, k-NN, fastText	UDC
2022	Borovic et al. (2022) [31]	*Hybrid UDC Recommendation for Slovenian Digital Libraries*	BM25, BERT	UDC
2022	Chou & Chu (2022) [12]	*BERT for Assisted Subject Indexing*	Longformer, SVM, k-NN, Random Forest, Neural Networks	LCC

Year	Reference	Title	Models / Techniques Used	Scheme
2022	Gooding et al. (2022) [32]	*Subjectifying Library Users to the Macroscope Using Automatic Classification Matching*	OCLC Classify API, DDC/LCC mapping	DDC, LCC
2022	Golub et al. (2022) [33]	*Automatic Classification Using DDC on the Swedish Union Catalogue*	SVM, Multinomial Naive Bayes, TF-IDF	DDC
2022	Modden (2022) [34]	*AI and Machine-based DDC Short Number Classification*	Support Vector Machine (SVM)	DDC
2022	Piros (2022) [35]	*Automatic Interpretation of UDC Numbers*	Rule-based parsing, syntactic decomposition	UDC
2024	Golub et al. (2024) [36]	*Automated DDC Using Annif Software*	SVC, fastText, Omikuji Bonsai, TF-IDF, ensemble	DDC
2024	Boryaev (2024) [37]	*Intelligent Global Bibliographic Search System*	SyntaxNet, Word2Vec, semantic networks	UDC
2024	Mahmud (2024) [21]	*AI in Automating Library Cataloguing and Classification*	Supervised ML, Deep Learning, NLP, and knowledge graphs	DDC, LCC, UDC, **FCS

Table 4 Machine learning and other methods used in library Classification

*Chinese Library Classification, ** Faceted Classification

Datasets Employed in AI-driven Library Classification

The effectiveness of ML-based classification depends on the type, scale, and quality of the datasets used for model training and prediction. Selected studies have used a wide range of data sources, including national bibliographic repositories, institutional metadata collections, full-text corpora, web-based documents, citation networks, and knowledge-rich resources, demonstrating the heterogeneity of the dataset. Table 2 highlights that most AI-driven classifications rely on national or union catalogues for model training. Borovic et al. (2022) trained a hybrid BERT-based model for UDC using 114,485 records from the Slovenian Open Access Infrastructure [31]. Modden (2022) applied the Annif platform with SVMs to millions of records from the German National Library for DDC assignment, while Golub et al. (2020, 2024) evaluated multiple machine learning and ensemble models using Swedish Union Catalogue metadata for Dewey classification [34][20],[36]. Joorabchi and Mahdi (2011, 2013) used MARC records, WorldCat data, and citation graphs to assign DDC numbers automatically. Kim and Lee (2002) developed a PMEST-based knowledge base for Colon Classification, while Frank and Paynter (2004) trained SVM classifiers on Library of Congress Subject Headings for LCC prediction [26][29][23].

Recently, Mahmud (2024) combined deep learning with knowledge graph embeddings across multiple schemes, Hou and Xue (2011) used lexical resources for Chinese Library Classification, and Boryaev (2024) employed multilingual semantic and structural analysis for UDC interpretation [28][21].

Data Category	Study	Dataset / Source
National Repositories	Borovic et al. (2022) [31]	114,485 records from Slovenian Open Access Infrastructure
	Modden (2022) [34]	Millions of records from the German National Library
	Golub, Hagelbäck & Ardö (2020); Golub et al. (2024) [20],[36]	Swedish Union Catalogue metadata
Full-Text Corpora	Chou & Chu (2022) [12]	3,671 digitised books from Project Gutenberg
	Kragelj & Borstnar (2021) [7]	70000 publications and 200000 historical docs from dLib.si
Web / Unstructured	Jenkins et al. (1998, 1999) [15],[16]	UK web documents crawled via the WWLib project
	Shafer (2001) [22]	Web pages (Scorpion system)
	Garland (1983) [14]	Simulated document sets
Metadata / Citations	Joorabchi & Mahdi (2013) [29]	Structured metadata, citation graphs, MARC, WorldCat
Knowledge-Rich	Kim & Lee (2002) [23]	PMEST-based knowledge base
	Frank & Paynter (2004)	Library of Congress Subject Headings (LCSH)
	Hou & Xue (2011) [28]	Cilin Dictionary & Chinese bibliographic datasets
	Mahmud (2024) [21]	Multimodel data (text, ontologies, visual descriptors)
	Boryaev (2024) [37]	Global bibliographic corpus with semantic networks

Table 5 Taxonomy of Data Sources in AI Library Classification

Comparative Accuracy of Machine Learning Models Across Classification Schemes

To identify the most effective machine learning models for automated classification, studies employing explicit machine learning techniques were considered. Rule-based systems, APIs, and purely heuristic approaches were excluded to enable direct performance comparison Table 3. Dewey Decimal Classification (DDC), Support Vector Machine (SVM) model achieved F1 scores in the range of 0.80 to 0.85, particularly when combined with TF-IDF features (Wang, 2009; Mehler & Wallinger, 2009; Golub et al., 2020, 2022) [27],[20],[33]. Hybrid SVM approaches outperform Naive Bayes and rule-based classifiers, especially at narrow classification for three-digit main class notation. Neural and deep learning models show mixed results, often constrained by class imbalance and limited training data.

However, accuracy in the Library of Congress Classification (LCC) is generally low, possibly due to its complexity. Chou and Chu (2022) reported that SVM achieved approximately 87% accuracy but only 41.8% exact class prediction. Pong et al. (2008) similarly validated k-NN over Naive Bayes, achieving an F1 score of 0.802 [12], [25]. Hierarchical SVMs (Frank & Paynter, 2004) achieve 55.3% exact accuracy, but performance increases substantially when evaluated using top-k or higher-level matches [17].

Classification Scheme	Models	Accuracy
LCC	SVM, k-NN	~0.80–0.87% (partial / higher-level matches)
DDC	SVM + TF-IDF	F1 ≈ 0.80–0.85
UDC	SVM + TF-IDF, BM25 + BERT	F1 ≈ 0.85; ~90%

Table 6 Overall Comparative Accuracy of Machine Learning Models Across Classification Schemes

UDC studies show the highest reported accuracies when hybrid and feature-enhanced models are employed. Kragelj and Kljajic Borstnar (2021) report an F1 score of 0.849 using SVM combined with TF-IDF, with approximately 90% correct UDC assignment under human-validated evaluation. Borovic et al. (2022) demonstrate that the BM25 + BERT hybrid model outperforms standalone IR and neural models across ranking-based metrics, supporting multi-label UDC classification [31].

Limitations of machine learning or AI in Library Classification

Automated classification in a library is limited by several factors, including technical limitations, insufficient or imbalanced datasets, incomplete classification schemes, and limited human audit.

Technical limitations

Transformer-based models, such as BERT, have a limitation on truncation: they have a fixed number of tokens per pass. Therefore, it reduced the ability to predict class counts accurately.

Classical ML models, such as Naive Bayes and SVMs, can classify simple titles but struggle with hierarchical structures when document titles are complex and overlap with other subjects; these models often confuse and miss interdisciplinary connections.

Corpus limitations

One of the most persistent problems is a lack of high-quality training data, which results in poorer performance. Unusual topics are frequently underrepresented in predictions of DDC- and UDC-classification. Some studies used a small corpus, which hindered the ability to recognise the semantic patterns. A limited dataset and conflicts between the training and test sets compromise further reproducibility of the methodology.

Restricted access to full classification schemes

Researchers often do not have full access to complete classification schemes; using simplified, shortened, or partial versions of these schemes reduces models' capacity to assign accurate class labels. Incomplete class labels in the datasets further affect model training and testing. Machine learning models appear to perform well in training and high-accuracy tests but struggle to predict complex titles and are often misclassified due to restrictions on using the full versions of the classification schemes.

Limited Human Auditing Procedures

Human Audit can examine whether the assigned class numbers generated by models are correct, meaningful, and consistent with the classification scheme's guidelines. Without this information, it makes it difficult to compare results and validate models. When human auditing is neglected, errors and misclassifications in automated classification may go unnoticed, reducing the model's applicability.

Findings

This study discloses that the application of AI or machine learning in library classification is real, not merely hypothetical. The findings show that research is being conducted to automate classification in libraries using schemes such as DDC, LCC, and UDC. The Dewey Decimal Classification (DDC) is the most widely adopted scheme, followed by the Universal Decimal Classification (UDC) and the Library of Congress Classification (LCC).

Several models are used to automate library classification, ranging from classical models such as Support Vector Machines (SVM), Naïve Bayes, and k-nearest neighbours to deep learning models such as BERT and Longformer. SVM has emerged as the most frequently used model across studies. Hybrid approaches combining neural networks and BM25 have also been used to improve classification performance in complex scenarios. Datasets used for training and testing in the models vary, for instance, national catalogues, full-text corpora, and citation graphs. However, some studies did not report the corpus's size or structure. Evaluation metrics for the model's performance, such as accuracy, precision, recall, and F1-score, were inconsistently applied and not clearly documented in all documents. Of 27 documents, only 20 reported formal evaluations. SVM, KNN, BERT, and BM25 models achieve promising F1 scores of 0.80-0.9. Furthermore, some studies are limited by imbalanced datasets, incomplete classification schemes, and restricted access to full data, which hinder the model's predictive capacity. Restricted access to complete classification schemes for training and testing the model, and the lack of human auditing processes, reduce the reliability of the automated classification process.

Conclusion

This systematic review highlights the current state of machine learning and AI applications for classifying library documents. This study identifies the potential of machine learning models and the limitations reported in studies on dataset characteristics, uneven evaluation metrics, and a lack of reproducibility. Future research can use the models identified in this study, and attention should be paid to dataset quality and quantity for training and testing to enhance classification accuracy and reliability. As machines cannot understand the principles of the classification scheme, human intervention remains necessary to ensure the assigned classes are correct, meaningful, and consistent with conventional classification schemes. The complex rules for class construction in DDC, UDC, LCC, and CC, such as multiple levels of classes and subclasses, the use of facets, auxiliaries, and symbols, subject headings for interdisciplinary topics, tables, add instructions, and exceptions, hinder predictions and lead to misclassification. So, models cannot learn until researchers provide well-structured, sufficient training data. As libraries evolve within digital and data-driven ecosystems, AI-powered classification can effectively assist classifiers and catalogers and serve as a foundational means to improve knowledge organisation.

Reference

1.Komolafe-Opadeji, H., & Ojo, R. A. (2019). A survey study on the adoption and implementation of automation projects in federal university libraries in the South-West of Nigeria. *The International Information & Library Review*, 51(2), 194–201. https://doi.org/10.1080/10572317.2019.1600784

2. Lund, B. D. (2023). Classification schemes: Universal, Special, National. In Elsevier eBooks (pp. 501–513). https://doi.org/10.1016/b978-0-323-95689-5.00010-9

3.Hjørland, Birger. "'Classification' (Reviews of Concepts in Knowledge Organisation)." Knowledge Organization 44 (January 2017): 97–128.

4. Palve, Pratik, and Gaurav Arora. AI-Powered Library Management System. 10, no. 4 (2025).

5. Ullah, A., Khusro, S., & Ullah, I. (2017). Bibliographic Classification in the Digital Age: Current Trends & Future Directions. *Information Technology and Libraries*, 36 (3),48–77. https://doi.org/10.6017/ital.v36i3.8930

6.Kragelj, M., & Kljajić Borštnar, M. (2021). Automatic classification of older electronic texts into the Universal Decimal Classification–UDC. *Journal of Documentation*, 77 (3), 755–776. https://doi.org/10.1108/JD-06-2020-0092

7. Harisanty, D., Anna, N. E. V., Putri, T. E., Firdaus, A. A., & Noor Azizi, N. A. (2025). Is adopting artificial intelligence in libraries urgency or a buzzword? A systematic literature review. *Journal of Information Science*, 51 (2), 511–522. https://doi.org/10.1177/01655515221141034

8. Marasinghe, M. M. I. K., Gunasekera, D., & Senevirathne, W. A. R. (2024). Application of Artificial Intelligence for Library Services: A Systematic Literature Review. Journal of the University Librarians Association of Sri Lanka, 27 (2). https://doi.org/10.4038/jula.v27i2.8089

9. Concha, K. M., Zenteno, F. P., & Alfaro, J. T. (2024). Use of artificial intelligence in libraries: a systematic review, 2019-2023. *South African Journal of Library and Information Science*. https://doi.org/10.7553/90-2-2387

10. Shahzad, K., Khan, S. A., & Iqbal, A. (2024). Factors influencing the adoption of robotic technologies in academic libraries: A systematic literature review (SLR). *Journal of Librarianship and Information Science*. https://doi.org/10.1177/09610006241231012

11. Adesina, A. S., & Zubairu, A. N. (2023). Contemporary Library and Artificial Intelligence Technology. Alexandria: *The Journal of National and International Library and Information Issues*, 33 (1–3), 18–30. https://doi.org/10.1177/09557490241231483

12. Chou, C., & Chu, T. (2022). An Analysis of BERT (NLP) for Assisted Subject Indexing for Project Gutenberg. *Cataloguing & Classification Quarterly*, 60 (8), 807–835. https://doi.org/10.1080/01639374.2022.2138666

13. Yunus, N., Nasir Ismail, M., & Osman, G. (2025). Smart library themes and elements: A systematic literature review. *Journal of Librarianship and Information Science*, 57 (1), 175–190. https://doi.org/10.1177/09610006231207098

14. Garland, Kathleen. 1983. "An Experiment in Automatic Hierarchical Document Classification." *Information Processing & Management* 19 (3): 113–20. https://doi.org/10.1016/0306-4573(83)90064-X.

15. Jenkins, C., Jackson, M., Burden, P., & Wallis, J. (1998). Automatic classification of Web resources using Java and Dewey Decimal Classification. *Computer Networks and ISDN Systems* , 30 (1–7), 646–648. https://doi.org/10.1016/S0169-7552(98)00035-X

16. Jenkins, C., Jackson, M., Burden, P., & Wallis, J. (1999). Automatic RDF metadata generation for resource discovery. Computer Networks,31 (11–16), 1305–1320.

https://doi.org/10.1016/S13891286(99)00012-2

17. Frank, Eibe, and Gordon W. Paynter. 2004. "Predicting Library of Congress Classifications from Library of Congress Subject Headings." *Journal of the American Society for Information Science and Technology* 55 (3): 214–27. https://doi.org/10.1002/asi.10360.

18. Wang, Jun. 2009. "An Extensive Study on Automated Dewey Decimal Classification." *Journal of the American Society for Information Science and Technology* 60 (11): 2269–86. https://doi.org/10.1002/asi.21147.

19. Lin, X., Khoo, M., Ahn, J.-W., Tudhope, D., Binding, C., Massam, D., & Jones, H. (2017). Mapping metadata to DDC classification structures for searching and browsing. *International Journal on Digital Libraries* , 18 (1), 25–39. https://doi.org/10.1007/s00799-016-0197-z

20. Golub, K., Hagelbäck, J., & Ardö, A. (2020). Automatic Classification of Swedish Metadata Using Dewey Decimal Classification: A Comparison of Approaches. *Journal of Data and Information Science* , 5 (1), 18–38. https://doi.org/10.2478/jdis-2020-0003

21. Mahmud, Md. R. (2024). AI in automating library cataloging and classification. Library Hi Tech News . https://doi.org/10.1108/LHTN-07-2024-0114

22. Shafer, K. E. (2001). Automatic Subject Assignment via the Scorpion System. *Journal of Library Administration* , 34 (1–2), 187–189. https://doi.org/10.1300/J111v34n01_28

23. Kim, J., & Lee, K. (2002). Designing a knowledge base for automatic book classification. *The Electronic Library*, 20 (6), 488–495. https://doi.org/10.1108/02640470210454010

24. Slavic, A. (2006). UDC in subject gateways: Experiment or opportunity? Knowledge Organization.

25. Pong, J. Y.-H., Kwok, R. C.-W., Lau, R. Y.-K., Hao, J.-X., & Wong, P. C.-C. (2008). A comparative study of two automatic document classification methods in a library setting. *Journal of Information Science*, 34 (2), 213–230. https://doi.org/10.1177/0165551507082592

26. Joorabchi, A., & Mahdi, A. E. (2011). An unsupervised approach to automatic classification of scientific literature utilizing bibliographic metadata. *Journal of Information Science*, 37(5), 499–514. https://doi.org/10.1177/0165551511417785

27.Mehler, A., & Waltinger, U. (2009). Enhancing document modeling by means of open topic models: Crossing the frontier of classification schemes in digital libraries by example of the DDC. *LibraryHi Tech*, 27 (4), 520–539. https://doi.org/10.1108/07378830911007646

28. Hou, H.-Q., & Xue, C.-X. Construction of the knowledge base for automatic classification of Chinese information in Chinese Library Classification. *Journal of the China Society of Indexer Classification.* –36.

29. Joorabchi, A., & E. Mahdi, A. (2013). Classification of scientific publications according to library-controlled vocabularies: A new concept matching-based approach. *Library Hi Tech*, 31 (4), 725–747. https://doi.org/10.1108/LHT-03-2013-0030

About Authors

Mr. Moirangthem Sajan Singh is a Research Scholar at Assam University, Silchar. He has completed his Master of Library and Information Science (MLIS) and has qualified for the UGC-NET examination. He is also a NFSC JRF Fellow. His areas of interest include, machine learning, sentiment Analysis, Scientometrics, Knowledge Organization, Library Automation.

Dr. Rajesh Rangappa Aldarthi is an Assistant Professor in the Department of Library and information science, Assam University, Silchar (A Central University). He Holds a Ph.D degree in Library and information science from Assam University, focusing on the cost-benefit analysis of library automation in central universities of North-East Indian Universities. He has more than 10+ years of experience in teaching and research in Library and Information Science at Assam University, Silchar. With Prior experience at Infosys, Wipro, and Tezpur University, He has authored several Books and Research Papers. His Interests include AI in Libraries, digital literacy and ICT-based knowledge management. He also holds an Indian patent titled "Method of Analysing Historical Texts Using Natural Language Processing to Identify Cultural Trends".

Prof. (Dr.) Manoj Kumar Sinha is a distinguished academician, researcher, and library professional in India. He currently serves as Professor and Head of the Department of Library & Information Science and as Dean of the Swami Vivekananda School of Library Sciences at *Assam University, Silchar*. Dr. Sinha holds Ph.D. degrees in Zoology and Library & Information Science from T.M. Bhagalpur University, Bhagalpur. He also earned a B.Ed. and a Post-Graduate Diploma in Higher Education. Prof. Sinha's research focuses on information and communication technology (ICT) in libraries, Internet literacy, e-resource access, digital library initiatives, and information behaviour studies. He has published numerous research articles in national and international journals and actively contributes to scholarly discourse in library and information science. He is also involved in professional activities, editorial boards, and academic committees, reflecting his leadership in LIS education and research.

Lucky Pathan is a library professional currently serving as a Library Trainee at The Global Library, O. P. Jindal Global University, Sonipat, Haryana. He previously gained hands-on professional experience as a Library Trainee at IIT Hyderabad. He holds a postgraduate degree in Library and Information Science. He has a keen interest in circulation management, acquisition processes, and information retrieval system.

Present Guest Teachers

Anamika Das, Associate Professor, NSOU
Arpita Roychoudhuri PhD
Avijit Kumar, Asst. Librarian, RBU
Durga Shankar Rath, Professor, VU
Kalyan Sundar Samata Librarian, Prabhu Jagatbandhu College
Radharani Rit Librarian, Bijoy Krishna Girls' College
Ranjan Samanta Librarian, Naba Ballygunge Mahavidyalaya
Santasree Basu PhD
Satarupa Saha Librarian, Surendranath Evening College
Sonali Dapsi Librarian, Raja Peary Mohan College
Sukanta Kumar Patra Librarian, Vidyasagar College for Women
Sushanta Bannerjee Librarian, WBSU

Ex Guest Teachers

Amalendu Ray
Amit Kumar Bandyopadhyay
Late Amitabha Basu
Amitabha Chatterjee
Amulya Ratan Chakraborty
Late Anil Chandra Pal
Anita Basak
Arjun Dasgupta
Late Arun Ghosh
Arun Kumar Chakraborty
Arup Roychowdhury
Ashis Kumar Pal
Asima Mandal
Atanu Das
Bapan Kumar Maity
Barun Kumar Mukhopadhyay
Bhabani Kumar Ghosh
Bibhuti Bhushan Mukhopadhyay
Late Bijoy Nath Mukhopadhyay
Binod Bihari Das
Biswajit Das Thakur
Late Biswajit Dey
Chaitali Dutta
Debashish Banerjee
Gopika Prasad Ghosh
Goutam Maity
Gurudas Ghosh
Harendra Nath Basu
Jagatpati Sarkar
Jayashree Sen
Joydeep Chanda
Late Juran Krishna Sarkhel
Keya Basu
Krishna Goswami
Laxminarayan Maity
Madhab Chandra Chatterjee
Madhusree Dutta
Manju Bala Bhakta
Mitali Ghosh Bhattacharya
Late Nandadulal Basak
Nitai Roychowdhury
Late Prabas Narayan Rath
Prabir Chandra Nag
Late Pramod Chandra Bandyopadhyay
Pratap Chandra Ray
Late Pulin Barua
Satyabrata Ghosal
Late Satyananda Mandal
Late Shibabrata Ghosh
Shivshankar Jana
Soumitra Sarkar
Late Sourendra Mohan Ganguly
Srabana Ghosh
Subal Chandra Biswas
Sujit Kumar Chatterjee
Sumanta Bandyopadhyay
Sunil Kumar Chatterjee
Late Swapan Bandyopadhyay
Tanushree Mandal
Late Tarun Kumar Mandal
Late Tushar Kanti Sanyal

PhD Sholars Awarded

Biswajit Das Thakur
Keya Basu
Madhab Chandra Chatterjee
Gurudas Ghosh
Bapan Kumar Maity
Sonali Dapsi
Tushar Acharya
Shyamal Ghosh
Sukanta Kumar Patra
Animesh Haldar
Sayan Palit
Mahitosh Das
Arup Bijali
Swarup Kumar Raj
Subhashis Karak
Kaushik Mandal
Falguni Pal
Abhijit Chakraborty
Satarupa Saha
Arpita Roychoudhuri
Madhushree Dutta
Sk Sofik

Present Research Sholars

Debashish Das
Laltu Aich
Rubi Acherya Manna
Mayukh Sarkar
Goutam Dinda
Asish Sharma
Pushpa Mandal
Swagata Khamrai
Sourav Debnath
Oishee Roy
Rakhi Dhibar
Mouli Purkait

Online Content of previous volumes available at:
https://lisrbu.wixsite.com/dlis/rbu-journal-of-lis

UGC CARE ENLISTED JOURNAL W.E.F 14.09.2019

Previously it was under UGC List of Journals

Upto February 10, 2025

Visit: https://rbu.ac.in/home/page/102

See Previous Volume's Content at:

https://lisrbu.wixsite.com/dlis/rbu-journal-of-lis

Published By

REGISTRAR

On behalf of the Dept. of Lib. & Inf. Sc.

Rabindra Bharati University

56A, B.T. Road, Kolkata 700050