Trends in Information Management (TRIM)
ISSN: 0973-4163
13(2), pp. 29-54

Systematic Literature Review of Data Curation Research and Stewardship of Libraries: An Opportunity to Untangle the Dilemma of Big Data Disputes

Pooja Rana
Pramod Kumar Singh

Abstract

This paper examines the crucial role of data curation in scientific development, particularly within the evolving landscape of research libraries. Recognizing the exponential growth of research data, the authors conducted a systematic literature review to analyze trends, themes, and gaps in data curation research. The findings highlight a surge in data curation research publications, particularly between 2018 and 2022, indicating a growing awareness of its importance. The study reveals the dominance of the United States and the United Kingdom in this field, emphasizing their commitment to data science. The paper underscores the significance of libraries as stewards of data, responsible for its management, preservation, and accessibility. It identifies challenges related to infrastructure, resources, and the evolving complexities of data management. By synthesizing these findings, the paper advocates for libraries to be at the forefront of data-driven scholarship, urging them to develop robust data services and advocate for the necessary resources required for the successful incorporation of these practices within libraries.

Keywords: Data curation, Libraries, Data librarians, Scientific discovery, Scientific Transparency, Data visibility, Research Productivity, Systematic Literature review

Introduction

The emergence of ICT support systems and techniques has prevalently increased digital data in academic and scientific research, in the league libraries have an important role in supporting the data management efforts. Libraries as an element of their traditional role can provide expertise in organizing and preserving data and facilitating its reuse through proper documentation and metadata creation. Curating data through layers of identification, documentation, and preservation enriches the integrity and quality of data, which is crucial for trustworthy scholarship. The data curation component helps maintain the long-term accessibility and usefulness of research data, thereby contributing to advancing knowledge and science.

If we dig into the concept of data curation, it specifically focuses on two major keywords i.e. accessibility and usability over time. Data curation involves organizing and integrating data collected from various sources, ensuring its quality and integrity, and making it available for discovery and reuse. Data curation is crucial in tackling the challenges associated with big data. The vast amounts of data generated by various devices, services, and individuals present several issues such as data quality, accessibility, and the ability to extract actionable insights (Siddiqa et al., 2016; Bello-Orgaz et al., 2016). In the context of modern libraries, data curation is highly relevant as libraries adapt to the digital age and take on new roles in the knowledge lifecycle (Tammaro et al., 2019). Many funding agencies require data management plans and long-term availability of research data as a way to avoid multiple efforts to generate what has already been revealed. Access to curated data facilitates a broader community for validation of research findings, and reusability of data across disciplines, thereby fostering new research collaborations (Johnston et al., 2017).

The inception of big data alongside creating problems, has also opened ventures for libraries to change their perception of data. By offering data curation services, library institutions are crucial in supporting the data management lifecycle and enhancing the overall impact and efficiency of the research conducted within their organizations (Yakel, 2007; Federer et al., 2016). To understand the full scope of data curation practices in libraries, this paper attempts to provide a systematic review of the emergence of the data curation concept in libraries and presents a statistical analysis of the possible implications of data curation services in libraries, identify best practices, highlight gaps, and suggest areas for further research and development.

Data Curation and Libraries

Data curation services in academic libraries are increasingly becoming a vital part of the libraries' offerings to support research and scholarship. With the explosion of data generated by academic and scientific research, libraries are evolving to include data management and curation as a part of their services. This includes teaching data management best practices, working with researchers to improve data management, creating subject guides on data management, assisting with compliance with funding agency and publisher data requirements, and sometimes getting involved with technical aspects like metadata and repository services (Surkis and Read, 2015).

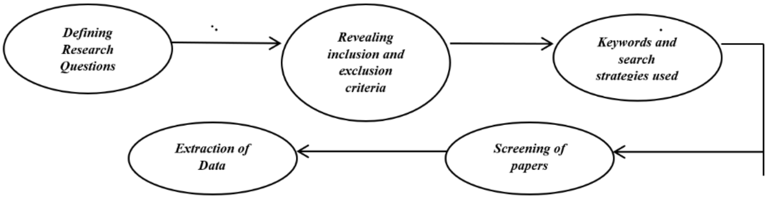
Libraries are also expanding their roles to offer more specialized services such as data curation-as-publishing, which involves making datasets public and ensuring they are suitable for reuse while maintaining the integrity and ownership of the data creators (Koltay, 2019). Libraries tailor their data services to meet the varied needs of different user groups and often require coordination between various library units and IT services to fulfill these roles effectively (Whitmire et al., 2015).

The implementation and expansion of these services depend on the library's capacity, mission, and strategic planning, with many libraries beginning to manage research data associated with faculty publications and theses. It is also important for libraries to stay updated with new approaches and methodologies related to managing research data brought forth by the evolving field of data science (Koltay, 2019). Data curation services are a growing and integral part of what modern academic libraries offer to support the research process and digital scholarship (Wang et al., 2013).

Methodology

The study incorporates a systematic mapping as a means to analyze the increasing culture of data curation practices in libraries. Figure 1 highlights the steps of the procedure followed by a brief description of each step.

Fig. 1: Steps followed to perform SLR on impact and mapping of data curation in library science.


Defining Research Questions
Revealing inclusion and exclusion criteria
Keywords and search strategies used
Screening of papers
Extraction of Data

Review Questions

R1: How has the increasing focus on data curation reshaped the realm of scientific research and discovery?

R2: How do data curation practices impact research and scholarship and its infrastructural requirements?

R3: What are the pressing challenges that libraries face in the deals of data curation?

Selecting Inclusion and Exclusion Criteria

The study selects certain inclusion criteria according to the research questions. For 1st Research question inclusion criteria are a) peer-reviewed papers focusing on data curation in libraries, (b) documents on data curation in scientific endeavor, (c) data curation in general (d) publications from the last ten years to ensure current practices are reviewed, (e) published between 2013-2024 and (f) written in or translated into English; and exclusion criteria i.e. (g) non-peer-reviewed literature, such as opinion pieces or editorials and (h) studies not specifically focused on data curation practices in library settings and scientific zones.

For 2nd Research question, particularly those research studies included that mention a) data curation impact on research and scholarly communication, role, and skills, education and training, policies all together b) that are in English language c) that are published during 2013-2024 and excluded those studies d) that mention data curation in general.

For 3rd Research question, particularly those research studies selected a) that mention data curation challenges in libraries b) Papers in the English language c) papers published between 2013-2024 and excluded those studies c) that present data curation in general d) that explain the same results and outcomes, e) that contain pages fewer than 4.

Search Strategies and Keywords/Phrases Used

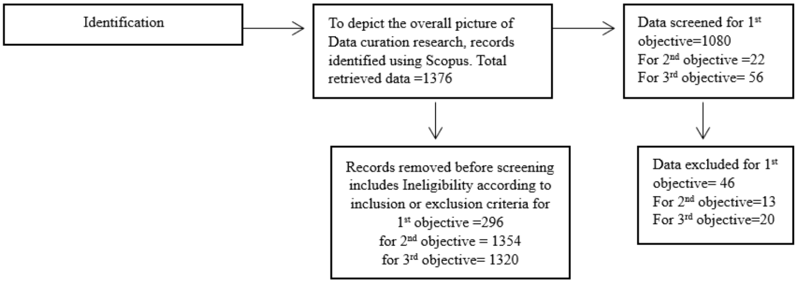
The strategy used for literature search includes academic databases such as Web of Science, Scopus, and Library and Information Science Abstracts. Use keywords, and phrases to search the literature relevant to the topic:

Used keywords	Phrases used	Combinations
Data Curation Academic Libraries Metadata Research data management Digital repositories Data Science	“Impact of data curation on scientific discovery” “Data curation and scientific collaboration” “Impact of data curation and research transparency” “Application of data curation in libraries”	“Data curation and research productivity” “Data curation and Scientific breakthroughs” “Data curation and libraries” “Research data management in libraries”

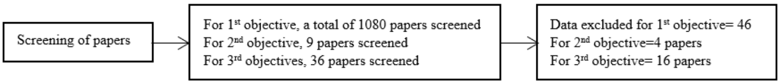
Screening and Extraction of Data

References from relevant articles are inspected to find additional literature. For data extraction, the data are collected on the authors, publication year, methodology, key findings, and recommended practices from the included literature. The data are synthesized thematically to identify the changing landscape of scientific research on data curation, impact, and challenges. The Graphs are prepared using Tableau (https://www.tableau.com/products/tableau), the visualizing tool. The quality and relevance of each source are assessed using a standardized tool i.e. PRISMA (a systematic literature review tool). The flow diagram for the extraction of data is shown in Fig 2.

Fig 2: Identification of papers/studies via databases


Identification
To depict the overall picture of Data curation research, records identified using Scopus. Total retrieved data =1376
Data screened for 1st objective=1080 For 2nd objective =22 For 3rd objective= 56
Records removed before screening includes Ineligibility according to inclusion or exclusion criteria for 1st objective =296 for 2nd objective = 1354 for 3rd objective= 1320
Data excluded for 1st objective= 46 For 2nd objective=13 For 3rd objective=20

Screening of papers

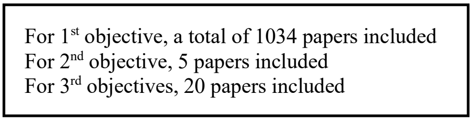


Screening of papers

For 1st objective, a total of 1080 papers screened
For 2nd objective, 9 papers screened
For 3rd objectives, 36 papers screened

Data excluded for 1st objective= 46
For 2nd objective=4 papers
For 3rd objective= 16 papers

Studies Included



For 1st objective, a total of 1034 papers included
For 2nd objective, 5 papers included
For 3rd objectives, 20 papers included

Analysis and Findings

Big Picture of Data Curation in the Scientific Endeavour

Global landscape of data curation: A country-wise Analysis

Fig 3 depicts the country-wise analysis examining the geographical distribution of publications related to "data curation" to identify leading countries and regions contributing to the field. Each dot or pattern on the map (represented in *Figure 4*) signifies a publication. Also, it presents a visual representation of the global reach of data curation research and represents countries that understand the potentially increasing significance of data curation research for collaboration and knowledge exchange. This analysis reveals the following key observations:

United States Dominance: The United States (*n=414*), *where "n" represents the number of publications,* exhibits the highest concentration of publications on data curation. This suggests a robust research landscape in the US, potentially driven by factors like substantial research funding, early adoption of open science principles, and the presence of influential institutions engaged in data curation practices.

European Union as a Leading Region: European Union countries collectively demonstrate a significant presence in data curation research. The UK (*n=119*), in particular, stands out with a high concentration of publications, reflecting its active research community and potential influence in shaping data curation practices.

China's Emergence: China's contribution (*n=81*) to data curation literature is notable, indicating a growing emphasis on research data management alongside its increasing investment in research and development.

Contributions from Australia and Canada: Australia (*n=35*) and Canada (*n=34*) also demonstrate a considerable number of publications, highlighting their engagement in data curation research.

Representation from Developing Countries: While developed countries appear to dominate the publication landscape, contributions from developing countries like India (*n=31*) are also evident, suggesting a growing global awareness and interest in data curation practices.

Fig. 3: Global distribution of research output on data curation research

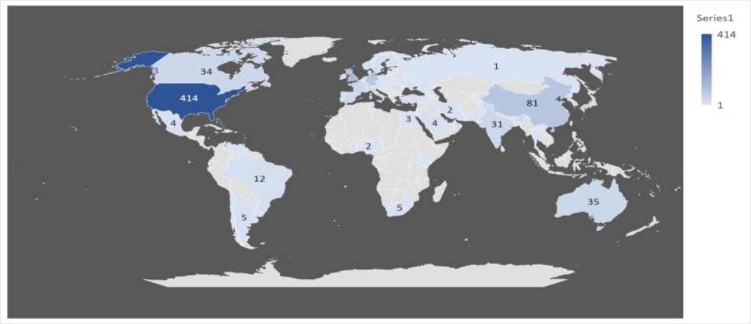
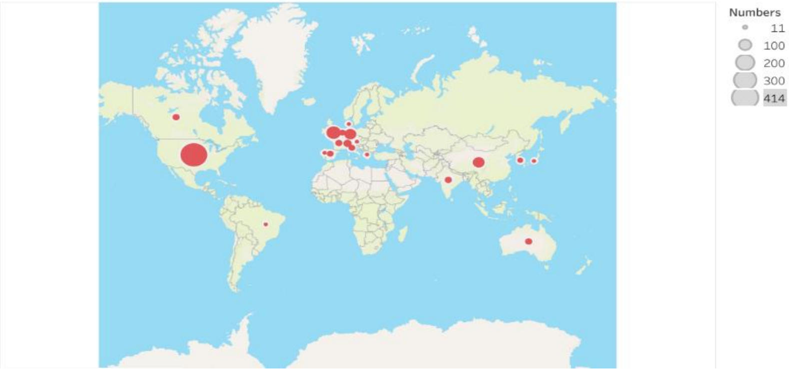


Fig.4: Dot density map for representing the concentration of research productivity of different countries on Data curation



Impact and Visibility of Data Curation Research: An Analysis of Citation Count

Figure 5 depicts the Citation count of the major title in data curation. This analysis investigates the citation impact of publications related to "data curation," focusing on how title characteristics relate to citation frequency. The Citation count patterns can provide insights into which topics or framing within data curation research have garnered the most attention and influence within the scholarly community. Here is the list of the top 20 highly cited articles in data curation (*Table 1*).

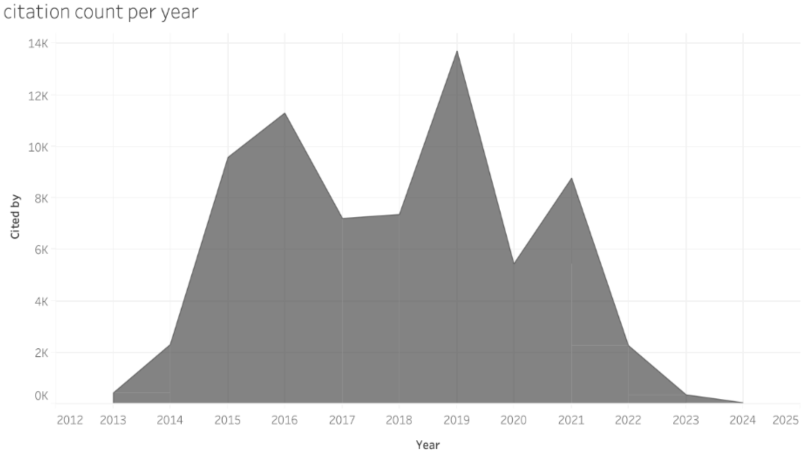
Table 1: List of Top 20 highly cited publications in Data Curation research

S.No.	Title	Cited by	Index Keywords
1	Comment: The FAIR Guiding Principles for scientific data management and stewardship	8698	Data Collection; Data Curation; Database Management Systems; Database management system; Information processing.
2	UniProt: A worldwide hub of protein knowledge	5032	Data Curation; Databases, Protein; Knowledge Bases; Molecular Sequence Annotation; Proteome; Sequence Analysis, Protein; proteome; access to information.
3	UniProt: the universal protein knowledge base in 2021	4213	Computational Biology; COVID-19; Data Curation; Databases, Protein; Humans; Internet; Knowledge Bases; Molecular Sequence Annotation; Pandemics.
4	CDD: NCBI's conserved domain database	2599	Amino Acid Motifs; Amino Acid Sequence; Conserved Sequence; Data Curation; Databases, Protein; Protein Structure, Tertiary
5	CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database	1641	Biological Ontologies; Computational Biology; Data Curation; Databases, Genetic; Drug Resistance, Microbial.
6	PANTHER version 11: Expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements	1571	Computational Biology; Data Curation; Databases, Genetic; Gene Ontology.
7	Clinical-grade computational pathology using weakly supervised deep learning on whole slide images	1330	Breast Neoplasms; Carcinoma, Basal Cell; Decision Support Systems, Clinical; Deep Learning

S.No.	Title	Cited by	Index Keywords
8	SMART: Recent updates, new developments, and status in 2015	1302	Data Curation; Databases, Protein; Protein Interaction Mapping; Protein Structure, Tertiary
9	The Immune Epitope Database (IEDB): 2018 update	1165	Antibodies; Antigens; Autoimmune Diseases; Data Curation; Databases, Protein; Epitopes; Forecasting; Gene Ontology;
10	Expanded Microbial genome coverage and improved protein family annotation in the COG database	1068	Archaeal Proteins; Bacterial Proteins; Data Curation; Databases, Protein; Genome, Microbial
11	The Bio-GRID interaction database: 2019 update	919	Animals; CRISPR-Cas Systems; Data Curation; Databases, Factual; Drug Discovery; Genes; Humans
12	The Reactome Pathway Knowledgebase 2022	857	Antiviral Agents; COVID-19; Data Curation; Genome, Human; Host-Pathogen Interactions; Humans
13	RCSB Protein Data Bank: Biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy	822	Biomedical Research; Biotechnology; Data Curation; Databases, Protein; Protein Conformation; Software; access to information; accuracy
14	The Bio-GRID interaction database: 2017 update	729	Animals; Computational Biology; Data Curation; Data Mining; Databases, Genetic; Humans; Protein Interaction Mapping
15	Recovery of genomes from metagenomes via a dereplication, aggregation, and scoring strategy	696	Algorithms; Animals; Computational Biology; Data Curation; Gastrointestinal Microbiome; Genome, Bacterial; Humans

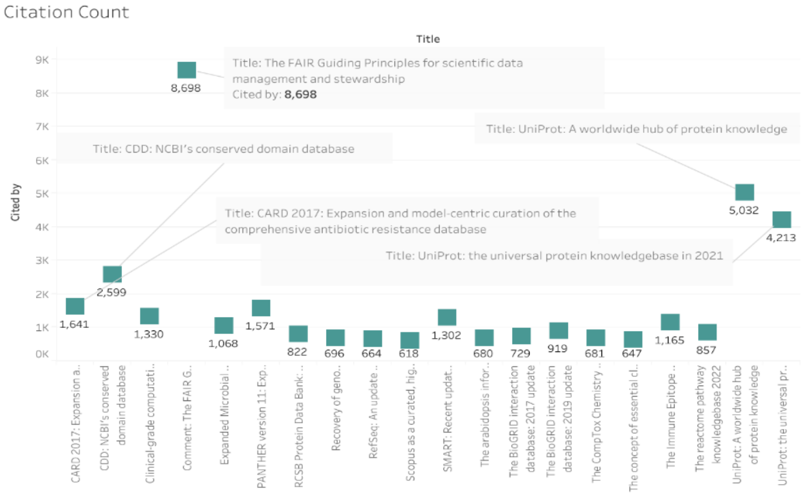
S.No.	Title	Cited by	Index Keywords
16	The CompTox Chemistry Dashboard: A community data resource for environmental chemistry	681	Bioassay data; Compound database; Computational toxicology; Data curation; EDSP21; Environmental chemistry
17	The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome	680	Alleles; Arabidopsis; Arabidopsis Proteins; Data Curation; Databases, Genetic; Genetic Association Studies; Genome, Plant; Arabidopsis; Arabidopsis thaliana; plant DNA
18	RefSeq: An update on prokaryotic genome annotation and curation	664	Archaea; Bacteria; Data Curation; Databases, Nucleic Acid; Databases, Protein; Eukaryota; Forecasting; Genome
19	The concept of essential climate variables in support of climate research, applications, and policy	647	Earth atmosphere; Meteorology; Climate research; Climate variables; Communities of Practice; Global climate observing systems
20	Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies	618	Citation linking; Content Selection and Advisory Board; CSAB; Data cleaning; Data clustering; Data curation

Fig 5: The Citation Count per Year of the Publication from 2013-2024



The top 20 publication that are cited more contains the keyword "data curation. Also, Publications with titles containing the keyword "Data management", "Data cleaning", "Data curation", "Knowledge-base", "Databases", "Genome", "Gene ontology" etc. had a significantly higher average citation count of average Citation Count (*for 20 top highly cited publications*) 39828 / 20 = 1991.4 as compared to those with keywords specifically related to their disciplines or without the descriptors such as data curation. Fig6 highlights the major titles in data curation that attain highest citation among other titles in data curation research.

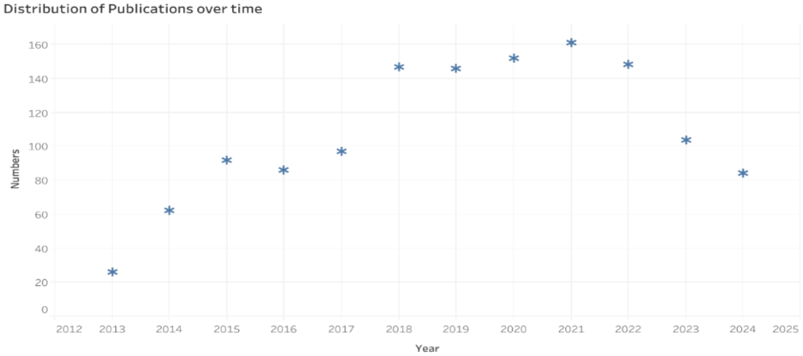
Fig 6: The Most Effective Titles in Data Curation that are highly cited



Temporal Trends in Data Curation Publications: A 2013-2024 Analysis

The year-wise analysis of the publications (*depicted graphically in Fig 7*) related to Data curation highlights the pattern and trends in research over time. The temporal analysis allows us to understand the field's evolution and highlights the period of significant growth. The Temporal dynamics of data curation research revealed the following:

Fig. 7: The Temporal Year-wise distribution of Data Curation Research Output



The increasing number of publications on data curation over the past decades suggests a growing recognition of its importance in research and scholarship. The observed peaks in publication from 2018-2022 often coincide with key developments in the field, such as the global push towards Open Science practices, with funders and policymakers increasingly mandating data sharing and open-access publishing.

The FAIR principles, published in 2016, gained widespread adoption, further driving the need for robust data curation practices. This likely led to a surge in research and publications focused on implementing and aligning with these principles. However, the years 2023-2024 show temporary fluctuations and further research is needed to determine if this dip represents a true decline or if it may be an indexing delay.

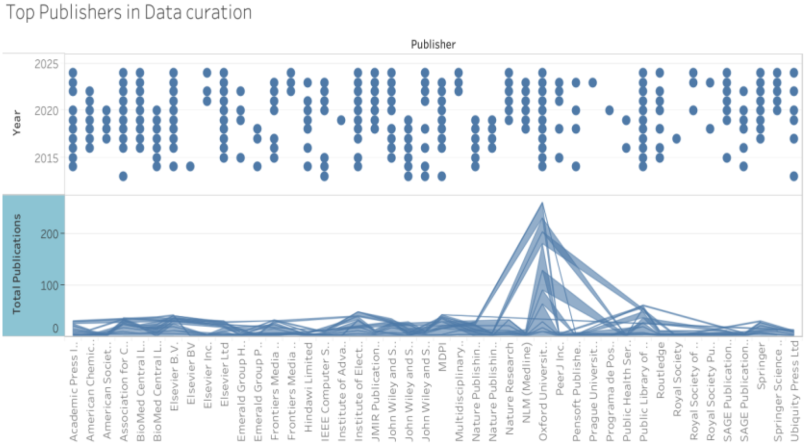
Top Publishers in Data Curation

The top publishers in the field of Data curation are presented graphically in *Figure 8*. *Figure 8* illustrates the prominence of top publishers along with their yearly productiveness. This analysis examines the leading publishers in the field of data curation, aiming to identify those who have made significant contributions to the dissemination of knowledge in this rapidly evolving domain.

This analysis identifies Oxford University Press (*n=261)* as a leading publisher in the field of data curation, based on their significant publication output and citation impact. Among other top publishers are the Public Library of Science (*n=62*), Institute of Electrical and Electronics Engineers Inc. (*n=49*), MDPI (*n=43*), Nature Research (*n=33*) and Springer (*n=31*).

These findings provide valuable insights for researchers, practitioners, and policymakers seeking to stay abreast of the latest developments in data curation and identify key sources of information in this domain.

Fig. 8: Prominence of Top Publishers in Data Curation Research from 2013-2024



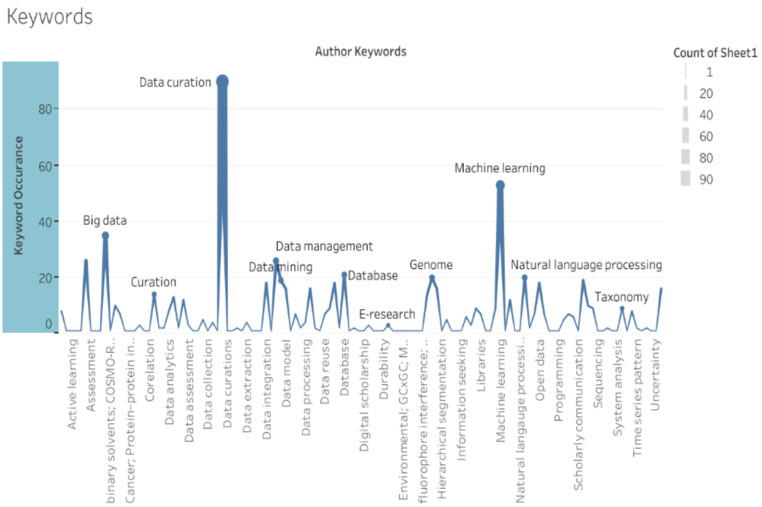
Keyword Analysis of the Publications

The keyword analysis of the scientific publication related to “Data Curation” (*depicted graphically in Figure 9*) provides understanding and insights into the emerging trends, themes, and subfields within data curation research and practice. A collection of 1034 research articles indexed in Scopus and Web of Science revealed a high frequency of keywords like “Data curation”, “Machine learning”, “Big data”, “Data management” and “Data mining”. These keywords underscore the central role of these concepts within data curation.

The emergence of “Open data” and "Scholarly communication" as major keywords suggests a growing emphasis on making data findable, accessible, interoperable, and reusable.

The presence of Keywords such as “Genome”, “RNA-Sequencing”, “Machine learning”, “Data Processing”, “Natural Language Processing”, “Information seeking” “Digital scholarship” and “Libraries” highlights the increasing culture of data curation research in the fields of Health Science, Computer science, and Library Science. Other emerging keywords like “Data collection”, “Data extraction”, “Data integration” and “Data reuse” highlights the major stages involved in the process of Data curation.

Fig. 9: represents the major emerging themes and keywords in data curation



Data Curation, Research Impact, and the Infrastructure They Demand

Data curation is essential for robust and impactful research, moving beyond simple data management to become a cornerstone of scholarly progress. By making research data readily available and reusable, curation fosters a collaborative environment where researchers can build upon existing knowledge, confirm findings, and avoid unnecessary duplication of effort. This accessibility, coupled with well-organized and documented datasets, streamlines the research process, allowing researchers to quickly find and interpret relevant information, ultimately saving time and resources. Furthermore, data curation promotes transparency and openness by making data publicly accessible and traceable, leading to greater scrutiny and validation of research, and ultimately, more reliable results. This commitment to open access aligns with the principles of Open Science, fostering a more inclusive and equitable research landscape. Importantly, data curation ensures the long-term preservation of valuable research data, safeguarding it against loss and making it accessible for future generations of scholars. This preservation is vital for building a robust and reliable body of knowledge. Finally, by making data from different fields more accessible and understandable, curation facilitates interdisciplinary research, leading to new collaborations and insights. In conclusion, embracing data curation principles paves the way for a more efficient, transparent, and collaborative research environment that accelerates the pace of discovery and ultimately benefits society as a whole.

Infrastructural Requirements

Modern libraries, facing an evolving landscape of digital information, require a robust and adaptable technological infrastructure to effectively curate their collections and services. This infrastructure goes beyond simply housing physical books and journals; it necessitates sophisticated data storage and management systems capable of handling vast digital archives, specialized databases, and multimedia resources. Libraries need dedicated software tools designed for digital curation tasks, such as metadata creation, preservation, and rights management as indicated by Tammaro et al. (2019). Furthermore, secure and high-speed networks are crucial, enabling seamless data sharing and access for researchers, students, and the broader community, often across institutional boundaries.

The ability to exchange information seamlessly with other libraries and digital repositories highlights the importance of interoperability and standardization (Johnston et al., 2017). By adopting standardized formats for metadata, digital objects, and communication protocols, libraries can ensure that their collections are discoverable, accessible, and reusable within a global network of information. This interoperability fosters collaboration, expands access to knowledge, and maximizes the impact of library resources.

As technology continues to advance, libraries must prioritize the technological proficiency of their staff. Librarians and information professionals need to stay abreast of the latest tools, trends, and best practices in digital curation to effectively manage, preserve, and provide access to their evolving collections (Laferty-Hess et al., 2020). This ongoing professional development ensures that libraries remain at the forefront of digital information management.

The increasing demand for efficient and user-friendly access to library resources drives innovation in data management technologies tailored specifically for the library context. This demand fosters the development of new platforms, tools, and workflows designed to streamline digital curation tasks, enhance discovery and access, and improve the overall user experience.

Finally, building and maintaining this robust technological infrastructure is a shared responsibility. Libraries, institutions, funding agencies, and technology providers must collaborate to invest in, develop, and support the systems, standards, and expertise needed for effective digital curation. This collaborative approach ensures that libraries can continue to fulfill their essential role as custodians and disseminators of knowledge in the digital age. Moreover, the major specifications related to infrastructural/technology requirements, Policies, Roles, and Skill requirements highlighted during the review process are indicated in Table 2.

Table 2: The Impact of Data Curation on Research and Scholarship and the Infrastructural Requirements for Successful Implementation of Data Curation Practices

Author	Infrastructure and Technology	Roles and skill developments	Impact on research and scholarship
Darch et al. (2020)	Installed base, Specific technologies (SDSS), and Cyberinfrastructure.	A unique set of skills, industry collaborations, data cleaning, validation, classification, and Familiarity with relevant tools and applications for data analysis and interpretations.	Long-term utilization, supporting peer-reviewed publications, increasing efficiency, promoting openness and collaborations, and Enhancing research integrity.
Laferty-Hess et al. (2020)	Refers to infrastructure and technology as critical components to support data curation services including systems that provide functionality for provenance.	Skill development in data validation and verification, metadata creation and standardization, knowledge of domain-specific data standards, and familiarity with tools and technologies used for data curation and repository management.	Data curation leads to more reliable and reproducible research, supporting core values of transparency and openness in science. Ensures long-term value of data, enabling future research and innovation by preserving data integrity and context.
Tammaro et al. (2019)	Addresses the need for infrastructure that supports data creation, preservation, and reuse, but the components are not detailed.	Expertise in data management planning and workshops, good communication and instructional skills, technical skills in archiving data, adaptation to changing environment, leadership, and vision.	Enhancing data access and reuse, Improving research efficiency.

Author	Infrastructure and Technology	Roles and skill developments	Impact on research and scholarship
Johnston et al. (2017)	Data repository, Data curation technologies and tools, Storage solutions.	Data curators, Library staff/subject liaison, Shared staffing model, Policy engagement, skills for data curation, soft skills that support collaboration and adaptability within the data curation process.	Indirectly addresses the impact on research and scholarship by discussing the roles and services provided by academic libraries in data curation.
Kong (2016)	Explored a variety of software applications, digital infrastructure configurations, and methods that the library might offer to support the curation, organization, and identification of geospatial data.	Professional skills from geospatial experts to optimize data-sharing platforms and manage data effectively depending on project requirements.	Libraries can assist researchers in managing their data throughout their full research cycle, from collection to sharing, management, and curation
Carlson (2013)	Repository systems, Metadata management systems, Preservation tools, Data management platforms, and Software for data analysis.	Understanding of research data, Metadata creation and management, Data preservation and access, Technological proficiency, Policy knowledge, Professional development, and soft skills.	Enhanced data discoverability, Data reuse, Transparency and verification, Long-term preservation, and interdisciplinary research.

Navigating The Data Deluge: Challenges for Librarians in Data Curation

Libraries face numerous challenges in establishing robust data curation practices within the evolving data-driven research landscape. A review of research articles (Table 3) highlights several key hurdles that are summarized following different points such as:

Data Reliability: Ensuring the dependability of sourced data, including verification of integrity, authenticity, and accuracy, is crucial. Libraries must act as gatekeepers of reliable and trustworthy data (Kumar and Singh 2023).

Low Data Reuse: Libraries need to actively promote a culture of data sharing and reuse by enhancing data discoverability, accessibility, and interoperability (Sheridan et al. 2021).

Lack of Proficiency and Policies: A significant challenge is the lack of professional proficiency and comprehensive policies for data curation. Competent authorities need to implement workshops and training programs, collaborating with specialized RDM partners (Masinde et al. 2021). Librarians with curation knowledge can effectively integrate practices and create value from unstructured data.

Clear Contribution Rationale: Researchers need compelling incentives to contribute data to curated repositories. Libraries can foster data sharing by developing comprehensive data policies that acknowledge and reward contributions, highlighting benefits such as increased research visibility and collaboration.

Communicating Data Value: Effectively communicating the value of curated data requires overcoming data storytelling obstacles. Libraries must translate complex datasets into accessible formats and craft compelling narratives for diverse audiences (McDowell 2023). Table 2 further details the major challenges identified in the review process.

Table 3: The Pressing Challenges in the adoption of Data curation Practices within Library Settings

Authors	**Challenges**	**Results**
Valdo Pasqui (2024)	Dynamic digital content and Open Science challenges. Selecting and deploying digital preservation platforms complexities. Cloud services reduce technical infrastructure costs for preservation platforms.	Digital curation and preservation are intertwined for long-term usage.
Prince (2023)	Challenges include lack of support, direction, and defined roles	Proposes ways to overcome obstacles in libraries defining the weakest and strongest skill sets required.
McDowell (2023)	Lack of time, high cost, lack of support for training	Identifies obstacles to library data storytelling
Kumar and Singh (2023)	Negligible curation efforts in academic institutions compared to research institutions. Identifying causes of negligible research data curation efforts in India	Three prominent challenges: dependability of sourced data, low reuse, and contribution rationale for researchers
Perkins et al. (2022)	Challenges, successes, lessons learned in libraries-research computing collaboration. Implementation of coordinator role, standard procedures, structured sharing venues	Implemented coordinator role, standard procedures, and structured idea-sharing venues for success.
Ashiq and Warraich (2022)	Lack of infrastructure, organizational support for data-driven services Missing data policies, limited training opportunities, lack of skills	Major challenges include missing data policies, limited training opportunities, lack of skills and expertise.

Authors	**Challenges**	**Results**
Chigwada (2021)	Data accuracy, confidentiality, security. Lack of skills and technology in libraries.	Libraries generate big data; challenges include data accuracy, security, skills gap, and technology availability.
Laskowski (2021)	Managing unstructured data, Ensuring ethical research data practices	Academic libraries are key to ethical research data future; need to structure services for unstructured data.
Lynch et al. (2021)	Challenges in data collection and use by African libraries	Highlighted challenges faced by libraries, focusing on African perspectives and the social context of knowledge.
Vitale et al. (2020)	Gap between expertise and local curation needs. Unfamiliar niche data formats and evolving research methods.	Developed data curation primers as a community resource, potentially inspiring new learning experiences.
Jeffery (2020)	Data volume, variety, veracity, value, analytics challenges. Preservation of software, equipment, sensors, and decision-making.	Increased motivation for data curation; use cases yielded significant scientific results.
Spiering and Lechtenberg (2020)	Lack of conceptual connections, diversity, and multiple perspectives in assignments	Curation assignments lack critical and collaborative elements; proposed revisions focus on collaborative, conceptual, and critical curation.
Koltay (2019)	Research Data Management, Low recognition of data curation importance	Academic libraries should engage in data curation. The overall level of recognition is low.

Authors	**Challenges**	**Results**
Kim et al. (2017)	Sustainable collection and selection of research information. Maximizing accessibility of collected information	Proposed autonomous data collection for up-to-date research information.
Creamer (2015)	Lack of consistency in handling student research data. Difficulty in building collections of data related to ETDs	Lack of consistency in handling student research data. Libraries exploring approaches for curating ETD-related data.
Kasianovitz et al. (2017)	Data management of acquired datasets impacting preservation. Unifying approaches across libraries to avoid working in isolation	Envisioning a library data service based on FAIR principles
Johnston et al. (2018)	Dissatisfaction with the current state of data curation. Gaps and opportunities for academic libraries to improve services	Researchers find data curation activities highly important. Majority dissatisfied with current state of data curation at institutions.
Kong (2016)	Extensive metadata is needed for unique research datasets. Uncertainty on which datasets to preserve and when to intervene.	Identified the library's role in research stages. Successfully addressed data problems in collaborated projects.
Kung et al. (2016)	Challenges include data sensitivity, replicability, metadata quality, and accessibility barriers.	A majority believe some data should not be permanently preserved.
Zhang and Chen (2015)	Lack of common standards and best practices. Varied data management and curation practices	Academic institutions and government agencies lead in data contribution.

Results and Discussion

Data curation has been evolving for the past few decades. The analysis of the publication has shown considerable results in accepting that data curation has gained widespread momentum and is being utilized in many domains. The geographical analysis revealed the dominance of the United States in the realm of data curation. However, it's crucial to acknowledge that various factors contribute to this, including research funding priorities, institutional support for data management, and national policies on open science. The geographical distribution underscores the need for international collaboration and knowledge exchange in data curation. Sharing best practices, developing common standards, and fostering partnerships across borders will be essential for advancing data curation practices globally. Observing the evolving geographical landscape of data curation research will be interesting, particularly as more countries recognize the importance of research data management and invest in developing robust data infrastructures. Moreover, the temporal trends highlighted the increasing significance and major growth during 2018-2022. The advancement such as new tools and technologies such as data repositories with improved functionality, data management platforms, and metadata management systems, and the need to deal with the situation of data deluge likely spurred research into their applications and effectiveness, contributing to the increase in publications. Furthermore, the relation of year, citation count, and keyword analysis reveals the major terminologies that increased visibility for scientific communication. This helps researchers develop a deeper understanding of how title characteristics relate to citation frequency and helps them craft more effective titles that increase the visibility and impact of their work. Further research could explore the role of other factors, such as publication outlets and author networks, in shaping citation patterns in this field.

In addition to the above results, the study examined the scope of data curation in Libraries through the impact of data curation on research and scholarship, its infrastructural requirement, and the potential challenges in the implementation of data curation practices. The review of the publication particularly focused on these parameters reveals the major issues that libraries are or may be facing in the implementation of data curation practices in the present and future.

Data curation is crucial for impact research and scholarship and has implications beyond imagination. It reduces the efforts, avoids duplication, prepares backup for long-term preservation, and encourages reusability of data. The libraries need to create a collaborative environment and promote transparency and openness of data. In addition to these, libraries need to work on creating platforms for data sharing. The libraries should work on creating the infrastructural support for the handling of data curation activities. This included new tools for data curation, related software, and databases, investors, funders, the hands-on practices on new technologies related to data curation. Furthermore, libraries need to pay attention to the running and anticipated difficulties in incorporating these services in the libraries. These challenges are related to the complexities of digital preservation platforms, difficulty in collection of data from the sources, lack of education and training, gap in skills, missing data policies, lack of infrastructural support, etc. The future incorporation for the easy and quick implementation of these practices requires the attention of the competent authorities to eradicate these challenges by framing policies, emphasizing the value of data sharing, necessary workshops, and training programs, and filling the gap between the existing knowledge of the librarians and the necessary skills required for handling data curation activities.

Conclusion and Future Prospects

The data plays a crucial role in deciding who's going to lead the competitive market. Ultimately the one who knows how to handle data will have an extra edge in shaping success. The availability of big data has led us to a situation of data deluge and managing such sheer volume of data has overpassed the existing capabilities of the professionals in the field. Libraries have always been providers of information and their skills and capabilities somehow match the skills and competencies required for data curation activities therefore, they present libraries with unprecedented opportunities to frontier as data curators. The research landscape of data curation has shown significant growth and has marked its presence in the field of library science along with major disciplines such as Health Science, Computer science, etc. Data curation has implications beyond imagination. Libraries need to create collaborative platforms for data curation, indulge investors and funders, work on enhancing their skills in handling data and executing data curation-related activities, literate their users on sharing data, framing policies for data preservation and data sharing. This way the libraries can create value from the datasets and make the data reusable by the scholarly community. Data curation research is not merely a technical endeavor; it's a scholarly pursuit that demands critical thinking, domain expertise, and a commitment to the long-term stewardship of knowledge.

References

Ashiq, M. Warraich, N.F. (2024). Librarian's perception on data librarianship core concepts: a survey of motivational factors, challenges, skills and appropriate trainings platforms. *Library Hi Tech*, 42(3), 849-866. https://doi.org/10.1108/LHT-12-2021-0487.

Bello-Orgaz, G, Jung & J. J, Camacho, D. (2016). Social big data: recent achievements and new challenges. *Information Fusion,* 28, 45-59.

https://doi.org/10.1016/j.inffus.2015.08.005.

Carlson, J. (2013). Opportunities and barriers for librarians in exploring data: observations from the data curation profile workshops. University of Massachusetts Medical School, *Lamar Soutter Library,* 2(2). https://doi.org/10.7191/jeslib.2013.1042.

Chigwada, J. P. (2021). Opportunities and challenges of using big data applications in institutions of higher learning libraries and research institutions. In *Big Data Applications for Improving Library Services,* Dhamdhere, N.D. (Ed.), 107-122. Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-7998-3049-8.ch008.

Creamer, A. (2015). Current issues approaches to curating student research data. *Bulletin of the Association for Information Science and Technology*. 41(6), 22-25. https://doi.org/10.1002/bult.2015.1720410610.

Darch, P T., Sands, A. E., Borgman, C. L. & Golshan, M S. (2020). Library cultures of data curation: adventures in astronomy. *Wiley-Blackwell,* 71(12*)*, 1470-1483. https://doi.org/10.1002/asi.24345.

Federer, C., Yoo, M. & Tan, A. C. (2016). Big data mining and adverse event pattern analysis in clinical drug trials. *Mary Ann Liebert, Inc*, 14(10), 557-566. https://doi.org/10.1089/adt.2016.742.

Jeffery, K. (2020). Data Curation and Preservation. Zhao, Z., Hellstrom, M. (Ed.) *Towards interoperable research infrastructures for environmental and earth sciences*. Lecture Notes in Computer Science, 12003. Springer, Cham. https://doi.org/10.1007/978-3-030-52829-4_7.

Johnston, L. (2017). Curating research data, volume one: practical strategies for your digital repository. *Association of College and Research Libraries*. https://www.alastore.ala.org/content/curating-research-data-volume-one-practical-strategies-your-digital-repository.

Johnston, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R. & Stewart, C. (2018). How important is data curation: gaps and opportunities for academic libraries. *Journal of Librarianship and Scholarly Communication*, 6(1), 1-24. https://doi.org/10.7710/2162-3309.219.

Kasianovitz, K. & Williamsen, J. (2019). From acquisition to access to archiving: creating library data services that provide end-to-end support for library-acquired data. Paper presented *at IFLA WLIC 2019* - Athens, Greece - Libraries: dialogue for change, Session 153a, Preservation and Conservation with Big Data SIG.

Kim, Y. K., Yang, J. A., Cho, J. M. & Kim, S. (2017). Data curation with autonomous data collection: a study on research guides at Korea University Library. https://library.ifla.org/id/eprint/1727/1/S06-kim-en.pdf.

Koltay, T. (2019). Data curation in academic libraries as part of the digital revolution. *Association of Polish Librarians,* 57(1), 28-36. https://doi.org/10.36702/zin.12.

Kong, N. (2016). The geospatial data curation, management, and discovery in academic libraries. http://library.ifla.org/1467/.

Kung, J. Y. C. & Campbell, S. (2016). What not to keep: not all data has future research value. *Journal of the Canadian Health Libraries Association Journal De Association Des bibliothèques De La Santé Du Canada*, 37(2). https://doi.org/10.5596/c16-013.

Lafferty-Hess, S., Rudder, J., Downey, M., Ivey, S., Darragh & Kati, R. (2020). Conceptualizing data curation activities within two academic libraries. *Pacific University Library,* 8(1). https://doi.org/10.7710/2162-3309.2347.

Laskowski, C. (2021). Structuring better services for unstructured data: academic libraries are key to an ethical research data future with big data. *The Journal of Academic Librarianship,* 47(4), 102335. https://doi.org/10.1016/j.acalib.2021.102335.

Lynch, R., Young, J. C., Jowaisas, C., Rothschild, C., Garrido, M., Sam, J. & Boakye-Achampong, S. (2021). Data challenges for public libraries: African perspectives and the social context of knowledge. *Information Development*, 37(2), 292-306. https://doi.org/10.1177/0266666920907118

Masinde, J., Chen, J., & Muthee, D. (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.

McDowell, K. (2023). Library data storytelling: obstacles and paths forward. *Public Library Quarterly*, 43(2), 202–222. https://doi.org/10.1080/01616846.2023.2241514.

Pasqui, V. (2024). Digital curation and long-term digital preservation in libraries. *JLIS.It.,* 15(1),109-125. https://doi.org/10.36253/jlis.it-567.

Perkins, D., Knuth, S., Lindquist, T., Johnson, Eichmann-Kalwara, N. & Dunn, T. (2022). Challenges and lessons learned of formalizing the partnership between libraries and research computing groups to support research: the center for research data and digital scholarship. In *Practice and Experience in Advanced Research Computing 2022: Revolutionary: Computing, Connections*, You (PEARC '22). Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3491418.3535165.

Prince, N. (2023). Continuing education and data training initiatives are needed to positively impact academic librarians providing data services. *Evidence-Based Library and Information Practice,* 18(3), 81–83. https://doi.org/10.18438/eblip30382.

Sheridan, H., Dellureficio, A. J., Ratajeski, M. A., Mannheimer, S. & Wheeler, T. R., (2021) Data curation through catalogs: a repository-independent model for data discovery. *Journal of eScience Librarianship,* 10(3). https://doi.org/10.7191/jeslib.2021.1203.

Siddiqa, A., Abaker, I., Hashem, T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A. & Nasaruddin, F. (2016). A survey of big data management: Taxonomy and state-of-the-art. *Journal of Network and Computer Applications,* 71(2), 151-166. https://doi.org/10.1016/j.jnca.2016.04.008.

Singh, D. M. K. & Kumar T. K. (2023). Research data curation in academic Institutions challenges & expectations. *DESIDOC Journal of Library & Information Technology,* 43(1), 39-44. https://doi.org/10.14429/djlit.43.01.18624.

Spiering, J. & Lechtenberg, K. (2020). Rethinking curation in school libraries a school library education: critical, conceptual, collaborative. *School Libraries Worldwide,* 26(1), 83-98. https://doi.org/10.14265.26.1.008.

Surkis, A. & Read, K. (2015). Research data management. *Journal of the Medical Library Association,* 103, 154-156. https://doi.org/10.3163/1536-5050.103.3.011.

Tammaro, A. M., Matusiak, K. K., Sposito, F. A. & Casarosa, V. (2019). Data curator's roles and responsibilities: an international perspective. *Libri,* 69(2), 89-104. https://doi.org/10.1515/libri-2018-0090.

Vitale, C., Hadley, H., Moore, J., Johnston, L., Kozlowski, W., Carlson, J., Blake, M. & Herndon, J. (2020). Extending the research data toolkit: data curation primers. *International Journal of Digital Curation,* 115(1), 1-14. https://doi.org/10.2218/ijdc.v15i1.713

Wang, D., Song, C. & Barabasi, A-L. (2013). Quantifying long-term scientific impact. *Science,* 342, 127-132. https://doi.org/10.1126/science.1237825.

Whitmire, A. L., Boock, M. & Sutton, S. C. (2015). Variability in academic research data management practices: Implications for data services development from a faculty survey. *Program*, 49(4), 382-407. https://doi.org/10.1108/PROG-02-2015-0017.

Yakel, E. (2007), Digital curation. *OCLC Systems & Services: International Digital Library Perspectives*, 23(4), 335-340. https://doi.org/10.1108/10650750710831466

Zhang, Y. & Chen, H.-l. (2015). Data management and curation practices: the case of using DSpace and implications. Proceeding of the Association for Information Science and Technology*,* 52 (pp. 1-4). https://doi.org/10.1002/pra2.2015.1450520100109.

Corresponding Author
Dr Pramod Kumar Singh can be contacted at:
pksingh22@gmail.com

Author Biographies

Pooja Rana is currently working as a Librarian at Government Degree College (GDC), Doda. She is also a PhD research scholar at DLIS, University of Jammu, with a strong passion for leveraging technology to enhance library services. Her research interests include data curation and the application of artificial intelligence (AI) in libraries, with a focus on improving information management, service efficiency, and accessibility.

Dr Pramod Kumar Singh is working as an Associate Professor in the Department of Library and Information Science (DLIS), University of Jammu. He has authored several research papers published in reputed national and international journals and has also edited two books. His research interests include Bibliometrics, Scientometrics, Information Technology Applications in Libraries, Open Access Initiatives, and Library Management, with a particular focus on recent trends and developments in the field of Library and Information Science.