FROM SHELVES TO SERVERS: BIG DATA IN LIBRARY AND INFORMATION SCIENCE

Kamala Mohanty
Research Scholar
Rashtrasant Tukadoji Maharaj University, Nagpur
Email: kamala.nts@nitandhra.ac.in

Shalini R. Lihitkar
Professor
Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur
Email: shalinilihitkar2021@gmail.com

Crossref DOI - https://doi.org/10.5281/zenodo.23415209

Abstract :

Libraries have a long history of collecting, organising, preserving, and disseminating information resources for a wide range of user communities. Since data in all its forms has always been essential to library and information science (LIS), the rise of big data is not a completely new idea but rather the technological and quantitative extension of an existing field. Technological developments as well as the changing demands of academia and society have led to the rise in importance of fields like scholarly communication, data sharing, and data curation in this changing environment. Current innovations like digital repositories, networked research infrastructures, and open science are redefining these fields. This study analyses how Big Data is changing library and information science (LIS) practices, roles, and technologies. The study also looks into how Big Data may affect areas like data curation, academic communication, information access, and the changing role of the librarian in a networked, digital research environment.

Keywords : Big Data, Library and Information Science, Digitisation, Technological Advancement.

Introduction :

Libraries are now thriving centres of knowledge and technology, having long been the quiet retreats of scholars and bookworms.The advent of digital libraries has opened up a new world of innovation, accessibility, and convenience.This shift has been fueled by the pressin g need to adjust to the fast-paced, on-demand information era. This change requires digital catalogues; for this change to occur, digital catalogues are crucial. The days of looking through card catalogues are over due to online technologies that make it simple for users to browse vast collections. Many academic papers, research resources, and multimedia can be accessed with a few keystrokes. Patrons are no longer limited by physical presence and can now access resources from anywhere in the world.

The transformation of libraries from traditional repositories to digital wonders is a prime example of adaptability. As we celebrate the peaceful coexistence of technology and

timeless wisdom, libraries continue to uphold their commitment to fostering education, curiosity, and connection. The transformation in library administration exemplifies the mutually beneficial relationship between innovation and tradition, which ensures the long-term viability of these beloved establishments.

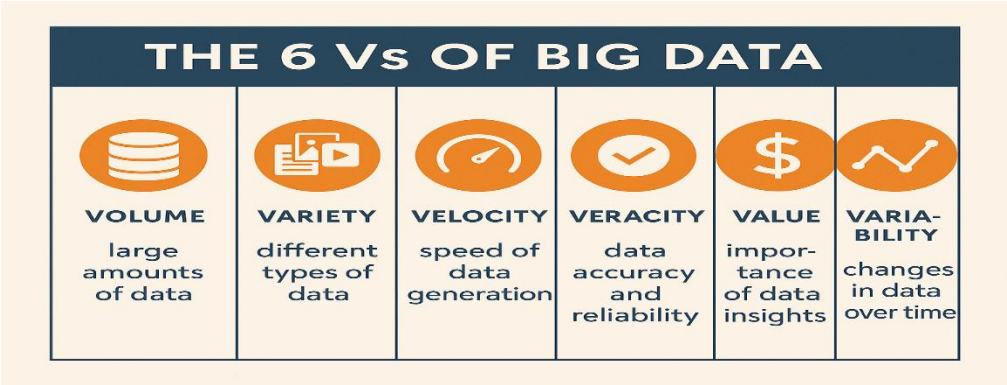
Big Data is prevalent in libraries as they are known to store a wide range of materials in various formats, whether they be academic, recreational, social, cultural, health, or religious. Since OCLC and other data-sharing organizations automated cataloguing and circulation systems, libraries have started using big data. World Cat is the earliest example of big data. The largest library content and service network in the world, World Cat facilitates knowledge discovery across a number of domains.

Definition of Big Data :

The concept of "Big Data" first appeared in debates about massive datasets in both academia and industry in the 1980s. However, with the increase in the amount and diversity of data made possible by significant developments in data science, the idea gained a lot of traction (Diebold, 2000, as quoted by Yan, 2013). Information gathered from various sources and available in a variety of formats, such as structured, semi-structured, and unstructured data, is referred to as big data.

To better understand the complexity and characteristics of Big Data, it is often described through the **6 V's**:

1. **Volume** : The vast amount of data produced every second by digital platforms, sensors, transactions, social media, and other sources is referred to as volume.
2. **Velocity** : This indicates how quickly fresh data is created, gathered, and processed often in real time.
3. **Variety** : Reflects the wide range of data types (e.g., text, images, videos, logs) originating from multiple channels.
4. **Veracity** : Concerns the accuracy, reliability, and quality of the data, acknowledging possible biases, noise, and inconsistencies.
5. **Value** : Emphasizes the importance of deriving meaningful insights and actionable intelligence from raw data.
6. **Variability** : Refers to the fluctuations in data flow and meaning, which can complicate analysis and decision-making.



(Fig-1)

While the "Vs" offer a practical taxonomy for understanding big data, there is ongoing debate within academic and professional communities about their comprehensiveness and applicability. Nonetheless, these characteristics collectively highlight why big data presents both an opportunity and a challenge for fields such as Library and Information Science.

Literature Review :

1. According to Alice A. Bamigbola (2023), big data is a new subject that has drawn interest from all around the world and has drastically altered the way we work, live, and think. The use of big data in libraries has revitalized the profession and given librarians new duties. Because traditional librarianship abilities are no longer sufficient to operate in data-intensive library contexts, librarians must acquire new skills to assist make enormous datasets more visible, valuable, and accessible to users.
2. David et al. (2021) study said Libraries can boost the efficacy and efficiency of their operations while providing their patrons with more engaging and individualized experiences by utilizing big data analytics. Libraries that embrace big data will be well-positioned to prosper in the digital era as data volume and complexity continue to increase.
3. (Zhan & Widen, Citation2019) examined the idea and general comprehension of big data in librarianship. Thirty-five papers that defined big data were collected and analyzed by the writers. Since there were so few definitions examined, the study combined content analysis with a statistical description of the definitions. After reviewing the collected definitions, the writers compiled five key components. Their explanation of librarians' big data expertise and cultural knowledge of libraries in connection to big data was facilitated by these five elements.
4. Subhajit Panda (2021) the study emphasizes how crucial Big Data analytics capabilities are in revealing new information usage difficulties, which in turn aids a librarian in carrying out his responsibilities as an embedded librarian in both theoretical and practical ways.

Objective of the Paper :

The study's primary goal was to investigate big data's function, significance, and prospects in library and information science (LIS).

Below are the few points of objective :

1. To determine the application of Big Data in Morden libraries.
2. To evaluate the influence of Big Data on library services, decision-making processes, and user engagement.
3. To investigate the obstacles and difficulties encountered by LIS professionals in implementing Big Data technologies.

Methodology :

In order to fully investigate the function, applications, and implications of big data in library and information science (LIS), this research employs a mixed-approaches approach that blends qualitative and quantitative methods. Current trends, usage patterns, and the shift in library services from conventional methods to data-driven solutions are the main topics of the study.

Why a transformation is needed in the scholarly landscape (LIS) field :

The evolving nature of information in the digital age has expanded far beyond traditional formats. While conventional LIS research has primarily centred on print materials and structured digital sources such as bibliographic databases, the rise of Big Data introduces complex, unstructured, and real-time data streams. These include social media interactions, IoT sensor outputs, web logs, and multimedia content. The following points highlight the reasons why a transformation is essential in the LIS field.

Integration of Data Science Skills :

- Data mining, machine learning, and analytics skills are essential for LIS workers.
- Data science must be incorporated into LIS research objectives and curricula in schola rly programs.

Methodological Innovation :

- Classical qualitative and bibliometric methods are insufficient for analysing Big Data.
- There is a need to adopt computational, algorithmic, and statistical research methods in LIS scholarship.

Ethical and Privacy Challenges :

- Scholarly work must address ethical issues like user privacy, data bias, and misinformation in Big Data environments.
- New frameworks are needed for responsible data stewardship and digital ethics.

Relevance to Contemporary Information Needs :

- Public and academic libraries are already leveraging Big Data for user analytics and service design.
- LIS research must guide best practices and innovations to remain practically relevant.

Redefining Core LIS Concepts :

- Traditional concepts like cataloguing, classification, and metadata must be reinterpreted in the context of Big Data environments.
- Scholarly inquiry should explore how these concepts apply to large-scale, dynamic datasets.

The image shows sequences from the Traditional Library to the Virtual Library :

(Fig.2)

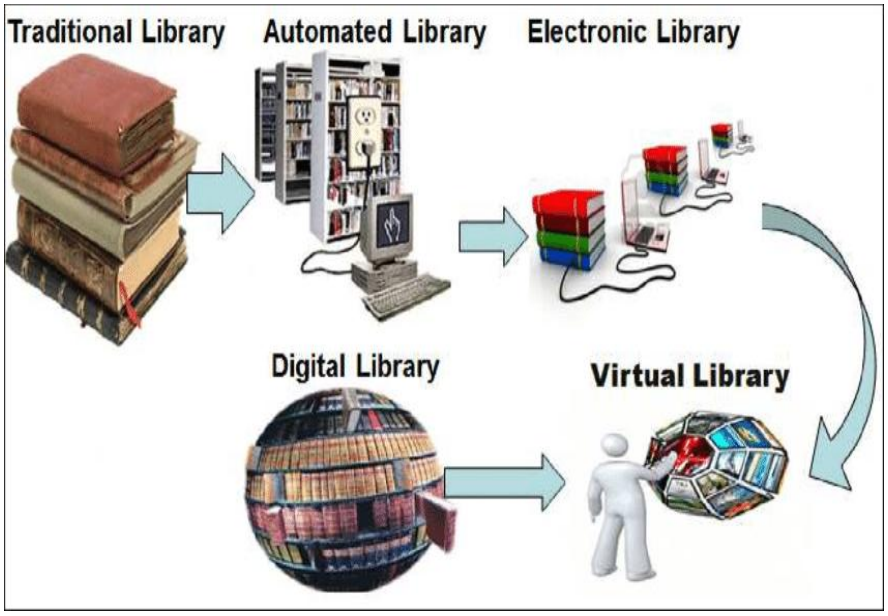


Image source : (https://www.researchgate.net/figure/Sequences-from-Traditional-Library-to-Virtual-Library_fig1_281838709)

Importance of Big Data in Library and Information Science :

There are several reasons why libraries should investigate big data more thoroughly. These include the ability to interpret scholarly data, the connection to open online resources, and the shift away from the notion of cataloguing correctness.

1. Big data has a direct effect on libraries since they may use big data methods to analyse their enormous data collections, gain a deeper understanding of their users, and offer new or improved services. For instance, to offer new or improved services and gain a

deeper understanding of their own users.

2. As academic researchers use big data extensively in their work, libraries will be indire ctly impacted by this trend.
3. For financial reasons as, for example, big data can result in cost savings, automation, a nd quicker and more accurate decision-making.
4. Big data analysis is crucial and full of opportunities. There are still plenty of chances to be involved in big data initiatives, as just 0.5% of the world's data has been examined to date.

Major Initiatives of Big Data :

Several examples of national and international projects have been initiated by various organisations and institutions to harness Big Data as a dynamic and comprehensive knowledge base for delivering valuable information resources.

National Initiatives :

Shodhganga (INFLIBNET) : Shodhganga, a flagship initiative of INFLIBNET in India that embodies the principles of open access and knowledge sharing in the academic community to provide a centralized digital repository of Indian Electronic Theses and Dissertations (ETDs). Shodhganga is a prime example of how Big Data and digital repositories are transforming Library and Information Science (LIS) in India.

Traditional Knowledge Digital Library (TKDL) : The Council of Scientific and Industrial Research (CSIR) and the Ministry of AYUSH, Government of India, collaborated to create the Traditional Knowledge Digital Library (TKDL). This project's primary goal is to preserve Indian traditional knowledge, particularly in the areas of naturopathy, yoga, siddha, Unani and ayurveda.

National Digital Library of India (NDLI) : NDLI aims to establish a one-stop shop where students of all levels, from elementary school to graduate and research, can obtain a variety of educational materials for free. The nodal institution in charge of the creation and administration of NDLI is the Indian Institute of Technology, Kharagpur (IIT KGP).

International Initiatives :

The creation and provision of a Twitter archive by the Library of Congress : In April 2010, the Library of Congress (LoC) and Twitter entered into an agreement in which Twitter granted the library a digital archive of all public tweets from the platform’s launch in 2006 through April 2010.

The creation of a geophysical data metadata base in Australia : It is a national initiative aimed at enhancing the management and sharing of geophysical data in alignment with the FAIR principles—Findable, Accessible, Interoperable, and Reusable.

Big Data Applications for Books (Harvard University Library) : Harvard University Library's notable application is the publishing of metadata for more than 12 million items, such as books and manuscripts, is a noteworthy use of the Harvard University Library. This allows for more efficient searching, cataloguing, and digitization of resources. Metadata is accessible via APIs and downloadable datasets. Data formats include MARC (Machine-Readable Cataloguing), MODS (Metadata Object Description Schema), and Linked Data models (e.g., RDF).

Tools and Technologies in Use :

To effectively engage with big data, libraries are beginning to adopt a range of emerging technologies:

- Data Analytics Platforms (e.g., Power BI, Tableau) for visualization and trend analysis.
- Machine Learning and AI for predictive modelling, recommendation systems, and natural language processing.
- Integrated Library Systems (ILS) with data analytics modules for enhanced decision-making.
- Cloud Storage and Distributed Systems (e.g., Hadoop, NoSQL databases) for handling large datasets.
- APIs and Interoperability Tools to enable seamless integration across systems and data sources.

The Benefits of Big Data for Library Users :

The function of libraries is shifting in a world which is becoming more and more digital. Once only holding books and information, traditional libraries are evolving into vibrant centres of study, knowledge, and community involvement. Big Data is one of the main forces behind this change. By utilizing big data, libraries can significantly enhance their community service, user experiences, and services. Here, we discuss the various benefits that Big Data provides to library patrons.

- **Suggestions for personalized content :** The ability of big data to offer tailored recommendations is a major benefit for libraries. Libraries can suggest books, articles, and other materials that suit users' interests by looking at their reading habits and choices. This customized strategy improves customer satisfaction and engagement while also making it easier to find fresh and pertinent material. Libraries can provide a more intuitive and user-friendly experience, much like the recommendation systems employed by websites like Netflix and Amazon.
- **Smarter Search Functions :** Big Data contributes significantly to improving library systems' search capabilities. Users can obtain the information they require more rapidly and effectively thanks to sophisticated algorithms that make search results

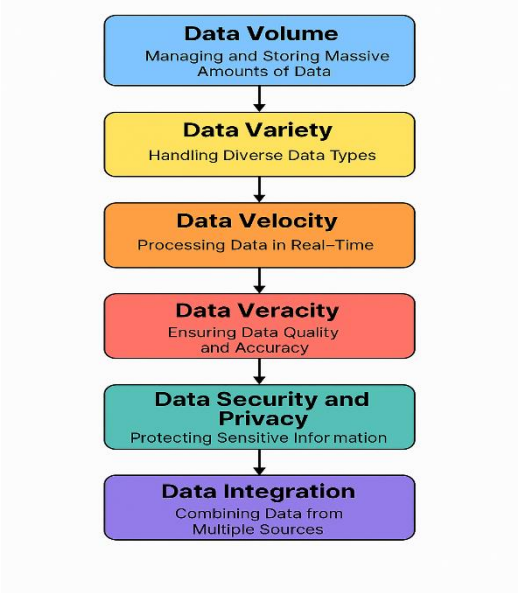
more accurate and pertinent. Furthermore, semantic search—which deciphers the context and intent of user queries facilitated through the integration of natural language processing.

- **Enhanced efficiency and accessibility of information :** Libraries may better manage and arrange their vast digital holdings by using big data, which provides users with instant access to a vast array of e-books, journals, and multimedia content from anywhere. Especially users who are unable to physically visit the library, this remote availability is quite helpful as it ensures fair access to necessary resources. Libraries can now provide 24/7 resource access thanks to big data-driven technologies, which increase productivity and make services more user-focused.
- **Robust User Assistance :** Big Data facilitates the creation of AI-capable chatbots and virtual assistants that provide users with quick support and guidance. The user's interaction is greatly enhanced by these smart tools, which can direct patrons through library systems, offer responses to frequently asked questions, and suggest relevant materials.

Challenges and Concerns :

The flow chart represents the challenges and concerns about the big data implications.

(Fig.3)



How Big Data Shaped Library Operations During the COVID-19 Pandemic ?

For academic libraries, big data is essential during the COVID-19 pandemic. Big data technology is necessary for librarians to effectively serve their user groups. Additionally, every library should provide during the COVID-19 pandemic. All library patrons can receive services at their doorstep by utilizing big data technologies and approaches. With the use of big data, libraries may now provide consumers personalized, remote, real-time, and virtual services. The amount, velocity, diversity, and veracity of data generated by various library

technologies are also rapidly increasing, creating new opportunities to better assess and improve user interactions in the library environment.

Conclusion :

Libraries and librarians have been actively and enthusiastically providing their patrons with the web information and services in recent years, particularly during the COVID-19 lockdown. This shift has led to increased use of social media and other digital platforms for promoting library services and disseminating updated information. However, with the surge in online content, it has become essential to analyse information for authenticity and accuracy using emerging technologies. To summarize all big data helps libraries make better choices in areas like tracking usage, utilizing library space, and developing collections based on demand. Additionally, Big Data improves the efficiency and consistency of standard library activities like cataloguing, indexing, archiving, management, preservation, and information representation.

References :

- Bamigbola, A. A. (2023). " BIG DATA" Revitalizing Librarianship: Impact on Librarians. Library Philosophy & Practice.
- Panda, S. (2021). Usefulness and impact of Big Data in libraries: An opportunity to implement embedded librarianship.
- Kamupunga, W., & Chunting, Y. (2019). Application of big data in libraries. International Journal of Computer Applications, 178(16), 34-38.
- Zhan, M., & Widén, G. (2019). Understanding big data in librarianship. Journal of Librarianship and Information Science, 51(2), 561-576.
- Golub, K., & Hansson, J. (2017). (Big) data in library and information science: a brief overview of some important problem areas. Journal of universal computer science (Online), 23(11), 1098-1108.
- Itmazi, J. (2013). The Importance of Digital Libraries in the Palestinian Universities. In Second International Conference on Information and Communication Technologies for Education and Training, Tunis.
- Tella, A. (2021). Librarians' perception of opportunities and challenges associated with big data in public libraries. Internet Reference Services Quarterly, 24(3-4), 89-113
- Xie, I., & Matusiak, K. (2016). Discover digital libraries: Theory and practice. Elsevier.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity.
- Privacy-Preserving Data Mining: Models and Algorithms. Edited by Charu C. Agarwal and Phillip S .U, Kluwer Academic Publishers, 2007.