

**APPLICATION OF BIG DATA IN ACADEMIC AND RESEARCH LIBRARIES:
PROSPECTS AND CHALLENGES IN THE DIGITAL AGE****By****Sotonye Orji****and****Hope Ebereamaka Chiekezie (CLN)****Abstract**

This paper discusses the prospects and challenges of Big Data application in academic and research libraries. It examines the concept of Big Data as an emerging phenomenon in the digital era. It maintains that there is increasing volumes of data creation and distribution across the world by different sources. The paper highlights various components of Big Data that influenced its application in academic and research libraries. It discusses the process of Big Data creation, control, tools and techniques of providing Big Data services for research and development in academic and research libraries. Also, the paper examines challenges that disrupt data service provision such as lack of awareness among librarians, lack of advanced technology facilities, skills gap, issues related to data security, storage and preservation. It concludes by recommending that academic and research libraries should create proper awareness of the current trends in Big Data services, provide advanced technology facilities, bridge the skills gap and provide adequate storage devices that can accommodate large volumes of data for efficient services.

Keywords: Academic Libraries, Big Data, Digital Age, Researchers, Service Delivery

Introduction

In the digital age the trending concept in academic and research libraries is obviously the application of Big Data in library services. Academic and research libraries provide access to data and information for the purpose of research and instructions. Over the years, academic and research libraries were using traditional method of data collection and management. This method covers the process of collection and application of hard copy data records in the form of inventory held in traditional storage systems. The hard copy data records are used to provide services such as reference and research information.

In the present digital age, there is paradigm shift in the application of data records in academic and research libraries. This shift is largely impacted by the advances in new information systems and massive production of data for research and development purposes. It led to the outset of Big Data as a concept that shapes the pattern of data production, capturing, storing, analyzing, searching, sharing and dissemination. Ularu et al (2012) opined that Big Data is another concept advanced by new technology thereby becoming a new phenomenon to library organizations and brings new competitive advantage to them.

Panda (2018) described the concept of Big Data as technological modernization of library and information centres. Big Data refers to the large amount of data that has been produced to solve a research problem. This type of data according to Sonawane (2018) is enormous and complex to the extent that traditional system of data processing is incapable to perform the capturing, analysis, storage, curation, sharing and dissemination. Sengupta (2020) asserted that Big Data is a broad term used to denote large-scale of both structured and unstructured data which is massive in nature and difficult to process using traditional techniques. It is commonly seen as massive data sets that may be analyzed computationally to reveal patterns, trends and associations especially relating to human behavior and interactions. Kamupunga and Chunting (2019) observe that Big Data continue to develop due to increase in library digitization. It helps libraries to support researchers, educationists, scholars and policy makers. The focus of this paper therefore is to examine the prospects and challenges of application of Big Data in academic and research libraries.

Components of Big Data

Big Data have certain components or characteristics which are crucial in the process of managing it. These components show the dimension of Big Data as an emerging concept in the digital age. According to Francesco et al (2021), Middleburg (2019) and Sonawane (2018) the components of Big Data are generally known as four Vs indicated below:

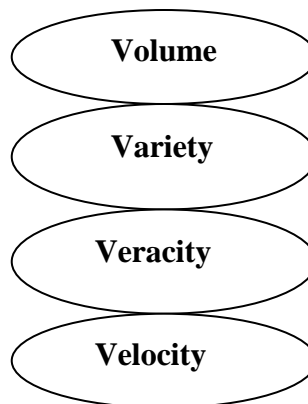


Fig1: Components of Big Data

1. **Volume of Big Data:** this denotes the actual volume of the data sets that need to be considered for examination, review, assessment, appraisal, processing and evaluation. There is rapid increase in the amount of data production due to information explosion. The data is too large for libraries to handle thereby impacting the pattern of processing and dissemination. According to Ku (2021) a data whether active or passive is growing exponentially and technology is enhancing the speed and efficiency at which large data sets can be processed. Furthermore, new, innovative and cost-effective technologies are continually emerging and providing solutions for libraries to improving large data handling.

2. **Variety of Big Data:** this component shows that Big Data comes from variety of sources in the digital age. Academic and research libraries engage in identifying different data sources that can support services and answer user queries. Different data sets are generated from various locations such as databases, websites, blogs, social media and files. The data is used in different ways to serve the need of variety of academic disciplines.
3. **Veracity of Big Data:** this is the essential attribute of Big Data that demonstrate its quality in terms of accuracy, authenticity, consistency and credibility before being analyzed by library. Here, library considers the value of data and the benefit it has for research and academic activities. Ceylan and Elif (2018) opine that data is an important asset; it must be reliable in order to achieve value when conducting analysis.
4. **Velocity of Big Data:** this stand for the speed in which data can be processed in libraries. Panda (2018) stated that Big Data enables high speed transfer from multiple locations more efficiently. According to Middleburg (2019) Twitter messages and Facebook posts are examples of data that can be generated with high velocity. Also, there is emergence of RSS feeds which is published and updated in real time online. It provides access to the newest data and information that are useful to library users who want to learn about the latest content.

Big Data Creation and Control in the Digital Age

There is increasing volumes of data creation and distribution across the world by different sources. This is often called mass production of data. The data is made available by content creators in large quantity in order to satisfy various needs and preferences. According

to Banton and Potters (2020) mass production of data is facilitated by ICT which is used to achieve high volume. The ICT revolution helps in mass production of data, efficiency in data flow and control. In the meantime, the data is large, complex and mostly generated from satellites, websites and cameras. Therefore, users have to select from the wide range of data that is constantly available.

In the contemporary academic and research environment, there is urgent need of data by scholars, researchers and students who are continuously using data to solve research problems. A high volume of data provides users with the opportunity to conduct thorough investigation and discovery. Diesel (2019) opined that Big Data production is the creation of “different facts, including numbers, measurements and translations that have been converted into a form that computers can process”. These data will empower researchers to identify problems, develop accurate theories, backup arguments and make informed decisions.

According to Lee and Strong (2004), there are three roles within Big Data production process. First, there are data creators such as individuals, organizations, institutions or other sources who create and generate data. Second, there are data custodians such as librarians, knowledge managers, data scientists and information managers who collect, store, process and disseminate data resources. The third category is the data consumers who retrieve and utilize data across various disciplines. They include researchers, scholars and students. This shows that academic and research libraries being custodians of data and information are in the middle between the data creators and data consumers. They serve as intermediaries in the data distribution channel. They provide a link to the data and saves the time of the consumers.

Academic and research libraries provide the basic technical services by ensuring that data is harmonized through processing, filtering and effective storage. It is worth noting that Big Data can only be harnessed effectively if libraries control the access process. Therefore, data control provides not only benefits to libraries but also give advantages to users for bibliographic access and use. The following figure shows Big Data creation and distribution channels in the present era of technology:

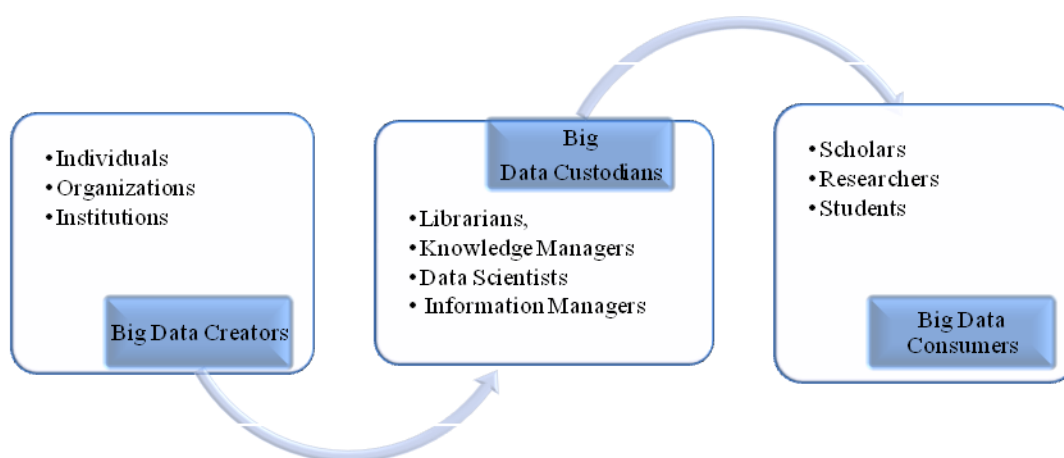


Fig 2: Big Data Creation and Distribution Channels

Prospects of Big Data in Academic and Research Libraries

Academic and research libraries are primarily concerned with the new ways of improving services in the digital age. The advent of technology compelled libraries to adopt the new concept of Big Data in service delivery. Sengupta (2020) observed that Big Data is now creating universal presence in the world of information and is likely to impact on all kinds of libraries. In addition, librarians need to know the basics of Big Data and how it affects the

nature of their work and how it can be used. There are different Big Data tools such as software and hardware that can be used for data gathering and analysis in order to support researchers.

Accordingly, libraries are expected to analyze both internal and external data. Salman, Abdullah and Sahid (2020) state that libraries are handling a lot of data which is adding value to the services they provide. These data include “customer data, service data (loan and return), research data, citation data and other data”. Kumar (2016) asserted that Big Data can be applied in subject analysis, development of library collections, tracking the use of collections, library administration and assessment of user satisfaction. The procedure allows for the new forms of processing user data and information. According to King (2018) libraries can use Big Data to create better services such as linking library databases to Geographic Information System Mapping (GIS Mapping). This technique helps libraries to connect the personal data of their users to a map in order to find out more about them and their location. The method further assists libraries to increase number of users and determine patronage from different geographic locations.

Hussain and Shahid (2022) observe that Big Data can improve the quality of library services whenever the data is made available via digital system. It can easily be harnessed using techniques in order to improve scholarly activities, creativity and innovation. According to Kumar (2016), Big Data affects libraries both directly and indirectly. The direct effect is in the use of Big Data tools to analyze data sets of libraries. On the other hand, the indirect effect is through the library users who are increasingly using Big Data in their research. Academic and research libraries play a fundamental role in facilitating access to data using latest

technology tools and techniques. As a result of this, libraries need to plan effectively and choose the most relevant tools and techniques. This will help achieving the best possible results in data handling and distribution.

Currently, the use of innovative tools and techniques in Big Data application in academic and research libraries is growing. This development increases the efficiency of libraries in the modern era. Library and Information Service Network (2022) suggested that libraries need to move away from a passive, reactive role to a more active one in order to respond to the demands of the researchers. The modern tools and techniques used in data management helps libraries to adequately respond to the research needs in different ways. Additionally, the following modern tools and techniques are imperative for academic and research libraries to provide Big Data service delivery to researchers, scholars and students in the digital age:

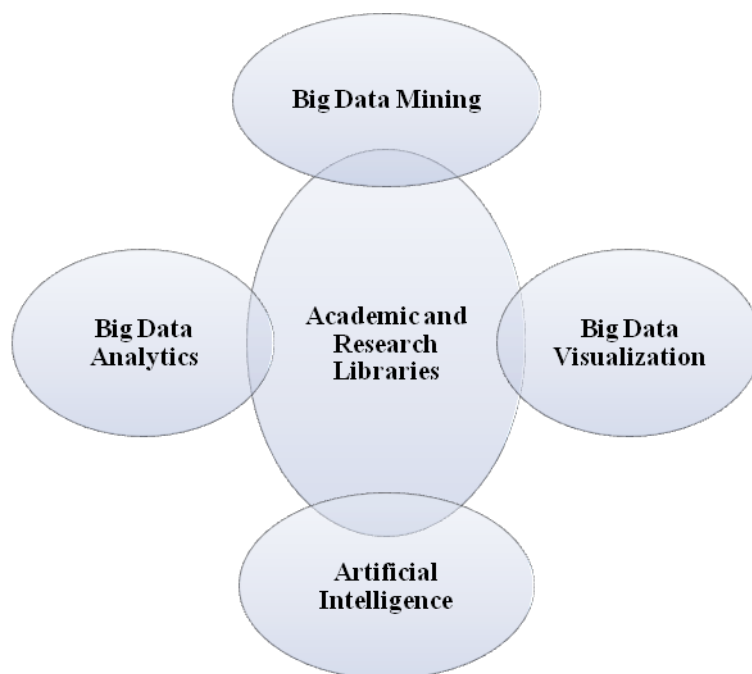


Fig 3: Tools and Techniques of Handling Big Data

1. Big Data Mining: librarians in academic and research libraries perform data mining activity when the data sets are too large. They extract relevant data using techniques and use it to support research, teaching and learning. Each data has a hiding value which librarians extracted to provide research evidence. Therefore, a meaningful knowledge can be obtained through this process. Dwivedi and Bajpai (2004) noted that librarians can use Big Data mining process to discover knowledge such as trends, patterns, characteristics or anomalies within huge data sets.

2. Big Data Analytics: this has to do with analysis and interpretation of data for use by scholars and researchers. Academic and research libraries provide data analytics services in order to support research effort and draw conclusion. In this process, librarians apply analytical approach to develop models. In most cases, advanced analytics application (Analytics App) are use to explore, describe and confirm the accuracy of data for reliable solutions. These applications such as online and offline data analysis packages provide logical and contextual view of the data for access, understanding and reporting.

3. Big Data Visualization: here librarians perform some work of data visualizations in the form of charts and graphs design. This activity makes data easier and faster to understand by users. Reveal (2022) stated that data visualization is used in libraries to identify trends and understand complex data quickly. It is also used to identify patterns, relationships, frequency and communicate data using graphs, charts and maps.

4. Artificial Intelligence: this opened a new horizons for academic and research libraries. The libraries adopt and use artificial intelligence as reproduction of human intelligence processes using machine learning to support data management and operations. Artificial Intelligence

devices in the form of robot are now influencing how data is connected. Also, it provide new data discovery platform that help users find and use data on the library websites, blogs and social media. Another prospect is the application of chatbot or chatterbot in facilitating online chat conversation either written or spoken. It allows library users to interact with robot devices for data discovery and use.

Challenges of Big Data Handling in Academic and Research Libraries

There are several challenges facing academic and research libraries when handling Big Data in service delivery. Chigwada (2021) observed that the concept of Big Data is new but libraries are facing the challenges of awareness, skills, data accuracy, confidentiality, security and lack of technology. These challenges are hereby examined below:

1. Lack of Awareness of Big Data: In some instances, librarians lack proper awareness regarding how Big Data is structured. There is insufficient understanding of the basic selection tools for the acquisition and management of Big Data as a resource as well as how the data is integrated into library services.

2. Lack of Advanced Technology Facilities: Most libraries lack suitable software applications and technology integration and upgrade that can support data analytics. There are challenges related to the use of innovative solutions such as automation, Artificial Intelligence and robotic in handling Big Data as well as machine learning for data extraction and use.

3. Skills Gap: The skills gap is generally considered as a setback for libraries in handling Big Data services. At times, libraries are faced with the challenges of inadequate massive data professionals and analysts.

4. **Finance:** There is often financial constrain to invest money in training and recruitment of librarians with essential knowledge of massive data analysis.

5. **Issues on Data Security:** Data that have been falsified or manipulated usually emanate from scientific misconduct thereby posing a serious challenge for libraries. In this case, libraries are facing difficulties of distinguishing fake data from authentic. Furthermore, privacy and regulatory issues hinder access to valuable data. Often, a large sensitive data is secured by content creators thereby denying libraries free access.

6. **Issues on Data Storage and Preservation:** The size and complexity of the data often determines the nature of the storage hardware needed by libraries. Computer storage capacity and external hard drives have different storage capacities. In addition, data storage is a major concern of libraries particularly with regards to computer viruses, system failure and lack of backup facilities.

Conclusion

Big Data is a considerable amount of data that is too large and require effective control for accessibility. The emergence of Big Data as a latest concept has clearly shown a new library service prospect. Academic and research libraries have the opportunity to serve their clients more efficiently. In doing this, several library service tools and techniques emerged such as data mining, data analytics, data visualization and the application of Artificial Intelligence. These activities are often confronted by certain challenges that disrupt data service provision. The key challenges observed by this paper include lack of awareness among librarians, lack of advanced technology facilities, skills gap, issues related to data security,

storage and preservation. Academic and research libraries in the digital era are encouraged to harness the advantages offered by Big Data in information service delivery.

Recommendations

The following recommendations were made in order to enhance application of Big Data in academic and research libraries for better services:

1. Academic and research libraries need to create awareness about the trends and new development on Big Data management thereby helping librarians to be conversant with the present and future services as well as how they can use data to support researchers, scholars and students.
2. Academic and research libraries should be equipped with advanced technology facilities. Without technology hardware and software libraries cannot partake in massive data generation, storage, analysis and dissemination.
3. Librarians need to be equipped with updated knowledge that will help them adapt to the changing digital environment such as knowledge of Artificial Intelligence. They should possess competencies to provide Big Data analytics services to researchers, scholars and students. Librarians should acquire digital literacy skills in order to distinguish fake data from the real one in a digital environment.
4. Academic and research libraries should improve the security of their information management systems by ensuring that data is protected and relevant backup are provided to preserve the integrity of data for research.

5. Data storage and preservation facilities should be made available to academic and research libraries. Storage devices that can accommodate large amount of data should be acquired to provide safe and proper storage of large volumes of data.

References

- Banton, C & Potters, C (2020) Mass production. Retrieved from <https://www.investopedia.com/terms/m/mass-production.asp>
- Ceylan, O & Elif, O (2018) A review of credit scoring research in the age of Big Data. *Journal of Financial Regulation and Compliance*. 26 (3) 382 – 405
- Chigwada, J. (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*. Pp 107 – 122
- Dwivedi, R.K & Bajpai, R.P (2004). *Use of Data mining in the field of Library and Information Science: An overview*. Inflibnet. India.
- Diesel, P.T (2019) The importance of data collection: 10 reasons why data is so important. Retrieved from www.diesel-plus.com/the-importance-of-data-collection-10-reasons-why-data-is-important/
- Francesco, C. Et al (2021) Big Data for creating and capturing value in the digitized environment: Unpacking the effects of volume, variety and veracity on firm performance. *Journal of Product Innovation Management*. 38 (1) 49 – 67
- Hussain, A & Shahid, R (2022) Impact of Big Data on library services. Prospect and Challenges. *Library High Tech News*. Emerald Publishing. Retrieved from <https://www.emerald.com/insight/content/doi/10.1108/LHTN-02-2022-0022/full/html>
- Kamupunga, W & Chunting Y (2019) Application of Big Data in libraries. *International Journal of Computer Applications*. 178, (16) 34 – 38
- King, D.L (2018) Big data and libraries. Retrieved from <https://www.davidleeking.com/big-data-and-libraries>
- Ku, L (2021). The impact of big data in business. Retrieved from <https://www.plugandplaytechcenter.com/resources/impact-big-data-business/>

- Kumar, A.J (2016) Big Data Technology: Big opportunity for librarians. In *Librarianship in ICT Age*. Pp 1 – 9
- Lee, Y.W & Strong, D.M (2004) Knowing- why about data processes and data quality. *Journal of Management Information Systems*. 20 (3) 13-39
- Library and Information Science Network (2022) Library effectiveness. Retrieved from www.lisedunetwork.com/library-effectiveness
- Middleburg, J.W (2019) The 4 characteristics of big data. Retrieved from <https://www.bigdataframework.org/th-four-vs-of-big-data>
- Panda, S (2018) Usefulness and impact of big data in libraries: An opportunity to implement embedded librarianship. *Technological Innovations and Environmental Changes in Modern Libraries*. Pp. 45 - 60
- Reveal, B (2022) Data Visualization. Retrieved from www.revealbi.io/glossary/data-visualization.
- Salman, M.S, Abdullah, M.J & Sahid , S.N (2020) Big Data analytic concepts in libraries; A systematic literature review. *International Journal of Academic Research in Progressive Education & Development*. 9 (2)
- Sengupta, S (2020) Big Data: *The next big opportunity for librarians*. Chandmal Tarachand Bora College. India.
- Sonawane, A.S (2018) Big Data: Application in libraries. *International Journal of Scientific Research in Multidisciplinary Studies*. 4 (1)
- Ularu, G.E (2012) Perspectives on big data and big data analytics. *Data Base System Journal*. III, (4)