Usefulness and Impact of Big Data in Libraries: An Opportunity to Implement Embedded Librarianship

Subhajit Panda[*]

Abstract

The concept of Big Data has been extensively considered as a technological modernisation in Library & Information centres. According to IDC, data volume is set to increase exponentially and envisages a data volume of over 160 zettabytes by the year 2025. Size is the first, and at times, the only dimension that leaps out at the mention of Big Data. Big Data is defined as information overload due to the volume, velocity, variety, variability & veracity of the data which must be processed to get value and better visualisation. Big Data contains the answer to several valuable questions related to patterns, trends & associations of user behaviour. It plays a major role in helping libraries to clearly understand the changing user needs, accordingly reshape & restructure their services & procedures. The primary focus of this study was to explore the concept of Big Data in a library environment, steps to introduce Big Data in libraries and use of Big Data in providing library services using the concept of data life cycle developed by DataONE. The main influential components to perform this study was the capabilities of Big Data analytics, need & usefulness of Big Data practices, its significant utilisation in libraries and discuss some globally taken practical initiatives. The study highlights the important role of Big Data analytics capabilities to uncover new challenges of the information utilisation, consequently helps a librarian to fulfil his role as an Embedded Librarian, both in theoretical & practical terms.

[*] *Assistant Librarian*, Chandigarh University, Mohali, Punjab (India)

INTRODUCTION

In the Reference Model for an Open Archival Information System (OAIS), data is defined as "[a] reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen." (CCSDS, 2012, p 20) To convert collected data into more contextual information, suitable for utilization, data processing is mandatory & usually performed by a data analyst or team of data scientists (Pearlman, 2018). And in this context, Big Data refers to a huge volume of data with massive output & exponential growth, and neither processed/analyzed by any common analyzing technique, nor stored in any traditional storage (Kiran, 2019). Though efficient & effective data processing tool & data storage required to deal with such high volume of data, it combines answer of a variety of valuable questions (Olendorf & Wang, 2017), reveal patterns, trends, and associations, especially relating to human behaviour & interactions and uncover insights (Sengupta, 2016) and improve the services in organisation level (Kiran, 2019).
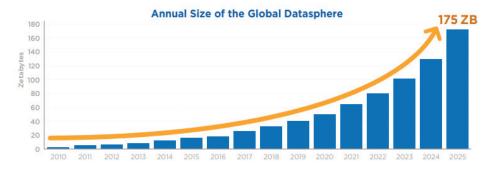
INFORMATION INFLATION & BIG DATA

Figure 1: Increase of Global Datasphere

Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018)

The digitisation of our world is advancing relentlessly and a far-reaching and serious side effect of this transformation is the fact that humankind is producing ever more data. Thanks to technology such as faster broadband connections, mobile technology, ever more powerful computers, larger memories and the Internet of Things, the data volume is set to increase exponentially in years to come. For instance, the market research company IDC envisages a data volume of over 160 zettabytes by the year 2025; it currently amounts to approximately 25 zettabytes (Reinsel et al., 2018). The global Big Data market size to grow from USD 138.9 billion in 2020 to USD 229.4 billion

by 2025, at a Compound Annual Growth Rate (CAGR) of 10.6% during the forecast period. The major growth factors of the Big Data market include the increasing awareness of Internet of Things (IoT) devices among organizations, increasing availability of data across the organization to gain deeper insights to remain competitive, and increasing government investments in various regions for enhancing digital technologies (MarketsandMarkets, 2020).

Characteristics of Big Data

Roughly first defined "Big Data (it) refers to any data that exceeds the user's ability to analyze it in one of three dimensions (the three Vs): Volume, Velocity and Variety" (Olendorf & Wang, 2017). Later, "Veracity/validity" becomes an essential characteristic to judge the authenticity of Big Data with "Value" as a lower entry (Jain, 2016) (Kiran, 2019). "Variability" & "Visualization" are often included as an additional quality of Big Data (Impact, 2016).

Table 1: Characteristics of Big Data

S.N.	Attribute	Meaning
1	**Volume**	With the application of IoT (Internet of Things), the enormous growth of information leading to the information explosion. With ever-increasing user-base & increased demand, evaluate & forecast such a huge amount of data requires extra attention.
2	**Velocity**	Velocity is the speed in which data is process and becomes accessible. Big Data enables high speed information transfer – as close to real-time.
3	**Variety**	Diversity of information retrieved from heterogeneous sources with diverse forms and formats. Also, user need is also getting diversified with libraries serving heterogeneous population with diverse subject's and interdisciplinary research.
4	**Veracity**	Authentic, credible and accurate information to reduce inconsistencies and uncertainty.
5	**Value**	Enabling & empowering users to be able to convert information into something useful and valuable.
6	**Variability**	It's different from variety. With time the meaning, value, validity, visualization of same Big Data file changes, which have a huge impact on data homogenization.
7	**Visualization**	To understand enormous amount of information, proper visualization is critical. Using charts and graphs to visualize large amounts of complex data is much more effective in conveying meaning than spreadsheets and reports chock-full of numbers and formulas.

Literature Review

The study of Hamad et al. (2020) attempted to reviewed Big Data analytics, its research direction & challenges to implementing it in academic libraries in Jordan. Ball (2019) provided an overview of the possibilities and chances of using large volumes of data in libraries, presented theories and explained realistic examples. Kamupunga & Chunting (2019) addressed the problems of compiling data from various sources and institutions in one location in real-time, which can save time. In addition, Blummer & Kenton (2019) reviewed the Big Data literature and libraries to recognize relevant trends and highlight possibilities and issues with enormous dataset management. While, Kaladhar et al. (2018) discussed various forms of datasets and their characteristics, consulted IT (Information Technology) experts on the relevancy of Big Data technology to the library operations. Sonawarne (2018) clarified the Big Data concept, its functionality, required skills & infrastructure for its implementation in libraries, also discussed the application of Big Data in libraries. Olendorf & Wang (2017) explored the possibilities of Big Data use in libraries and critically examined the benefits, costs and risks associated with using Big Data. The study also provided use cases, guidance to getting started and briefly outline tools and resources for working with Big Data. Similarly, Sengupta (2016) introduced the concept of Big Data and its usefulness in the library. Wang et al. (2016) identified the different characteristics of datasets in the library, reviewed the Big Data research on the library and then summarized the applications in this area. This study of Jantti (2016) introduced how the University of Wollongong Library (UWL) and the University's Data and Analytics Hub (formerly known as the Performance Indicators Unit) partnered to build new data and reporting models (comprised tools namely Value Club & Marketing Club) to join Library usage data with students' demographic and academic performance data.

Objectives of the Study

The main aim of this research is to spread awareness about Big Data among library professionals, investigate its analytic strategies and use-case in libraries. The primary objectives of this study are:

   (i)   To investigate the concept of Big Data in a library environment;

  (ii)   To structure the basic steps to introduce Big Data in libraries;

 (iii)   To examine the opportunities relating to Big Data services in libraries;

 (iv)   To investigate Big Data analytics tools used in the libraries; and

  (v)   To present the international initiatives using Big Data in various library project.

Statement of the Problem

Library and Information Centers as a part of either a higher education institution (Academic Library) or directly connected to society (Public Library) operate in a highly competitive environment with different varieties of data (viz, Structured, Semi-Structured & Unstructured) generated at a large scale with the exponential growth rate. Nowadays, libraries compete with each other and with other information providers and suppliers such as bookstores, authors and literary discussion groups (Hamad et al., 2016). As mentioned previously, Big Data contains the answer to several valuable questions related to patterns, trends & associations of user behaviour, it plays a major role in helping libraries to clearly understand the changing needs of their users, and to reshape and restructure their services and procedures accordingly (Gibbons, 2013). All forms of Big Data can require high-performance computing and specialized software to analyze. Therefore, the library professionals should carry a clear understanding of the need & importance of Big Data in libraries and basic knowledge to operate such specialized tools for information utilization.

PRIMARY STEPS TO INTRODUCE BIG DATA IN LIBRARIES

In library and information centres Big Data analytics is influenced & affected by two basic challenges. First, due to the massive volume of information, its management & exponential growth overwhelmed the traditional management, processing & storage system. And, secondly, the analytical methods & tools used to analyze Big Data are extensive & complex to understand. (Al-Barashdi & Al-Karousi, 2019) To resolve these two reversely proportional issues library should go through the following steps to implement the use of Big Data in libraries.

Identify the Framework

The framework of identification the area for implementation Big Data in the library environment can be understood typically through the following 7 Vs framework. These V-characteristics were first proposed by Doug Laney (first 3 Vs) (2001), now a Gartner consultant, in the context of emerging business conditions.

Need Analysis

It is important to first understand the potential benefits and current needs. This will not only prevent unnecessary allocation of resources, but it will also result in a better product. Libraries can have to primary interests in Big Data.

Firstly, libraries can use Big Data to optimize their collections, betterutilize space, asses their instruction and to provide information to their users. Secondly, Big Data can help libraries in the field of research data services.

Researchers often find themselves without sufficient skills and resources to adequately manage the data during their projects; here libraries have always played the role of housing societies academic and research output.

Required Infrastructure & Staffing

To implement Big Data services in library efficient infrastructure required both in the form of instruments & manpower. That's why, before implementation, it is necessary to estimate the cost of the implementation and identify the resources library already have that fit well with it.

Instrumental resources include the High-Performance Computing (HPC) centre & Cyber Infrastructure (CI) centre associated with servers, storage & database software, operating systems, statistical & analysis software owned or licensed to use. While the cloud is the most satisfactory infrastructural resolution to support the storage and process needs of Big Data analytics applications. Staffing resources are critical & there may be untapped skills within the working library staffs. It should be identified & ensured the capability of currently employed Information Technology staffs while starting implementation.

Solutions Identification

With identified essential needs and data about available resources in hand; at first libraries need to find solutions to fill the gap between the availability & minimum requirement to implement Big Data analytics. The components areas where the maximum concentration should be given to find solutions are,

Data store - For one-off solutions, the type of data store may be less critical. CSV or excel files may be sufficient for some projects, but they may not scale well with true Big Data projects, even one-offs, so it may be necessary to use SQL databases or document based databases such as Mongo DB or Hadoop. The cost of making a mistake isn't so great however, mainly in extra time spent fixing the mistake or time spent working with an inefficiency.

(a) **Administration** – For service level data stores libraries will need to provide an administrative layer. Some solutions come with built in administrative front ends.

(b) **Policies of use & access** – One important non-technical issue to resolve are policies of use and access. At the start, a document that states the scope of the project, and policies governing use should be drafted. To ensure focus & set limits, the scope specifies the data that are expected to be included in the data store and what they will be used for.

(c) **Analysis & visualization** – The types of analyses and visualizations that are needed may drive the solutions chosen. While most solutions

provide a wide array of analyses and solutions some are limited in what they can do (i.e. upto their skills & potentials).

(d) **Query access** — Query access is providing refers to providing a low level interface to the data. Usually in the form of an HTTP based API to allow the extraction of data to be used in online visualizations. This feature is not necessary, but many data warehouses provide. Implementing this requires some thought, including access control.

Implementation

The final implementation is completed after extensively going through the above-mentioned steps when a library's organising committee is satisfied with its current situation (i.e., infrastructural availability & staff efficiency) to fulfil the criteria for implementing Big Data analytics. However, if service level projects, will exceed the capabilities of any library to support them with the necessary infrastructure & expertise, collaborations with other institutional departments such as high-performance computing, advanced cyber-infrastructure, or similar units may be advantageous.

Using Big Data in Library Services

Libraries are in a unique position to use Big Data for their managerial purposes and also provide library services around Big Data analytics. Many libraries are already offering research data services such as data management planning, data collection, data curation and data archiving (Sugimoto et al., 2012). The extent of the provided services is limited according to the users' need & the current situation of the library/organisation. Here some of the possible services are discussed that can be provided by the libraries using the data life cycle developed by DataONE (DataONE, n.d.). In this life cycle, there are some stages where libraries can play an important role, while in other stages libraries don't have much importance. So, libraries will need to choose carefully what services to provide and how to resource them.

Planning

In the case of planning a organizational level research project, libraries can act as a research assistance and support researchers in improving their data analytics skills by helping them conceptualize how data is searched (e.g. by using different search terms) to get maximum efficient result. Librarians' database design and development skills can prove useful for organization and data mining processes in Big Data (DDS, 2021). Assisting with Data Management Plans (DMPs) is one of the most common services offered by academic libraries. Working with a researcher on a Big Data project only differs from typical DMPs in several ways.

Collection

Though, collection of Big Data may seem to be outside the range of typical library services, there are two ways libraries still can help researcher in collecting Big Data. First, some libraries may already have in house skills common Big Data sources, such as social media feeds, or, libraries can acquire commercially available Big Data sources and treated as part of their library collection. Second, acquiring and working with Big Data can be computer intensive. Libraries can collaborate with other units such as high performance computing to provide computing assistance and also the technical infrastructure to work with Big Data. Libraries primary role here might be to ensure that the data collected is properly organized and documented to ensure that it can be easily archived later.

Assure, Analyze and Integrate

After collection of required data Quality assurance (QA) is mandatory to validate its acceptance for the research. After assure the quality of the collected data it is analyzed with the help of relevant tools and techniques to organize, classify and summarize the data being collected for better comprehension and interpretation leading to understand and explore answers or solutions to the research problem which originally triggered the research. Quality assurance (QA), analysis and integration of the data is typically the responsibility of the researcher. However, if librarians are supplying data as part of their collection, there may be some responsibility to perform quality assurance on this data and provide adequate documentation for researchers to conduct their own QA methods.

Describe

For proper understanding any data file should be describe with adequate documentation and metadata information. Documentation is human readable information that provides details about various aspects of the data. Metadata is typically machine readable information, often machine generated and in addition to providing some information that aids in using the data, also functions to identify and describe the data. Providing data with adequate metadata and documentation fall squarely within the strengths of libraries and this is a task researchers struggle with providing proper documentation and metadata as well. Libraries should provide services to help researchers document their Big Data and supply it with metadata.

Preserve

Along with data description, archiving and preservation is perhaps the other primary service libraries can provide. Many libraries have institutional repositories and some also have dedicated data repositories. However, in most

cases they are limited by their technology to smaller and simpler datasets. This presents a problem to researchers who are using Big Data and whose data sets may not easily fit into the institutional repository solutions. Digital preservation in cloud may be one solution for better handle Big Data sets in an organization level.

Big Data, by itself, is also introducing new challenges for digital preservationists – especially related to indexing, efficient search and future accessibility & adequacy (Martin, 2011).

Discover

Aside from some of the more common Big Data sources, e.g. social media platforms like Twitter, discovering useful data can be difficult. Librarians involved in research data support can serve a useful role in working with researchers to find relevant and useful datasets. In addition to locating the datasets, librarians as informaticists should also be familiar with techniques for efficiently acquiring the data, storing and processing the data.

Big Data Tools

To implement Big Data in library services, the first requirement is efficient & effective tools to handle & analyze such enormous amount of data into usable information. Following are ten (10) such most popular Big Data tools identified & listed in Computer Business Review (Nunns, 2015):

- **Splice Machine:** Splice Machine is a highly scalable (from gigabytes to petabytes) real-time SQL-on-Hadoop database compliant with ANSI SQL standard. Its scalability range support .NET, Java & Python and ensure the developers who have found MySQL & Oracle as less scalable one to fulfill their desired limits.

- **MarkLogic:** MarkLogic provides users with real-time geographical data, combined with content, location & relevance filtering tools through real-time updates and alerts. It has the ability to deal with heavy & unstructured data loads. It supports flexible API's such as Node.js Client API, NoSQL and it also offers Samplestack to help show developers how to implement a reference architecture using key MarkLogic concepts and sample code.

- **Google Charts:** Google Charts with its giant search engine fingers in many pies and app developer tools is another area where the company has a strong offering.

This free tool comes with various capabilities,

- It visualise data from a website such as hierarchical tree maps or just simple charts,

- It enables easy implementation by embedding JavaScript code on a website,

- It allows sorting, modifying & filtering data as well as the ability to connect to a database or pull data from a website,

It also offers support for popular languages.

- **SAP in Memory:** SAP's HANA platform offers a number of advantages over some of the competition, such as the ability to integrate and analyse large workloads of data to be analysed in real time. This is extremely beneficial for the developer who is looking for speed to market. It offers users the choice between Eclipse and Web based tools which allows for a more collaborative model of development.

- **Cambridge semantics:** Using the Anzo Software Suite, this open platform helps users to collect, integrate and analyse Big Data to build Unified Access solutions. The software has a data integration machine that streamlines data collection and assists with analytics. It also has the ability to combine data from multiple sources and customised dashboards to make analysis easy.

- **MongoDB:** This is an open-source documental database comes with full index support and the flexibility to index any attribute and scale horizontally without affecting functionality. The document-based queries and GridFS for storing files mean that the users shouldn't have issues with compromising the stack. MongoDB is also scalable and includes third party log tools such as Edda and Fluentd.

- **Pentaho:** Pentaho, open and embeddable platform comes with extensive analytics capabilities with data mining and predictive analysis. It has the ability to joins data integration and business analytics for visualising, analysing and blending Big Data. It has the connectivity to any type of data source or source of data with native support with Hadoop, NoSQL and analytic databases. In this data integration tool users do not require coding in SQL or writing MapReduce Java functions.

- **Talend:** Talend's Open Studio is a combination of open-source tools, include products for developing, testing and deploying data management and software integration. The company also manages the full data lifecycle, even across enterprise boundaries.

- **Tableau:** Tableau is a well-known data visualisation software that offers many tools for developers and supported by an active community. Moreover, it has an in-memory analytics database and supports advanced query language, e.g., API, XML, User Scripts, Python, & JavaScript and several more browser extensions.

- **Splunk:** Splunk harness machine data created from several different sources, such as websites, applications & sensors and enables developers to write code using any technology platform, language or

framework. Extension tools have been developed for Visual Studio for .NET developers to build applications and use the Splunk SDK for C#. A plug-in for Eclipse contains a template for building Splunk SDK for Java applications and the company also provides logging libraries to help log activity from .NET or Java application.

INTERNATIONAL INITIATIVES

Following are some examples of international projects initiated by many organizations or institutions to make Big Data a dynamic & exhaustive knowledge base in providing valuable information source.

Library of Congress: Creating and Providing a Twitter Archive

The partnership between the short message service Twitter and the Library of Congress (LoC) is the best and preliminary example of the use of Big Data, announced in 2010. The goal of this partnership was archiving and retaining every tweet. Unfortunately, later this project seemed much trickier than initial assumption.

Creation of a Metadatabase for Geophysical Data in Australia

This project aimed at creating a metadatabase using geophysical data from the oil concern shell in Australia, which also required proficient librarians of Australia. It is beyond the ability of traditional cataloguing to manage & record such complex Geophysical ("big") data . Generally, Big Data contains petabytes of data in a wide variety of file formats, media forms and licensing conditions. It also ranges from raw data to processed and interpretable data. So, this project required collaboration among librarians, geophysicist, geophysical data analysts, IT experts and database developers.

Big Data Applications for Books (Harvard University Library)

Metadata on over 12 million materials, such as books, videos, audio recordings, images, manuscripts, maps and other content began to publish by Harvard University Library, the most comprehensive university library system in the world in April 2012. It was never possible to give free access to all those materials, due to some copyright issues. David Weinberger, Co-Director of Harvard Library Lab, describes "Big Data for Books" as a valuable trove of data.

Jisc & HESA Library Data Labs Project

On 30 November 2015, Heidi Plus was officially launched jointly by JISC (Joint Information System Committee) (Jisc, n.d.) and HESA (British Higher Education Statistics Agency) (HESA, 2019) to provide higher education

institutions & non-profit organizations with numerous tools to analyse enormous amounts of data ("Big Data"). The idea behind this web-based platform is to hand decision-makers a tool to save time and money on the one hand and gain swift and easy access to the information they need.

Brooklyn Public Library (BPL): Big Data for the Visualisation of User Data:

Another Big Data project was conducted by Brooklyn Public Library (BPL), one of the major public libraries in New York with numerous sites and branches. Two primary goals of this project were: being able to make swifter and more data-based decisions. But due to dependency on external advisers and outdated reporting systems, it was not possible earlier. Consequently, it has hard to know that from where the required data comes. So, BPL had to shift to Tableau (Tableau Software, 2018), an interactive data visualisation solutions provider. Tableau's data visualisation could take effect only after completion of the Sisyphean task.

Joint Big Data Initiative Between Ten US Libraries:

Another such sound example is the joint Big Data project between ten public libraries from all over the USA, the Institute of Museum and Library Services (IMLS) and CIVICTechnologies, a software provider for data analyses. According to their own claim, it is supposedly the first true Big Data project in history. Ten medium-sized to large libraries are cooperating on this project, with a population of over 7.8 million among which almost 52% (= 4 million) users have a library card, to use the possibilities of Big Data analyses jointly. The goal of this project was to get to know the customers & non-customers and their needs precisely to gear the services towards them.

Big Data and Embedded Librarianship:

Embedded librarianship offers a fairly new arena for library research. This service has emerged as one of the most effective avenues of instructional collaboration between distance education faculty and librarians (Figa et al., 2009; Shepley, 2009). The term embedded librarian derives from a similar phrase "embedded journalist" (Shumaker, 2009). Embedded librarians are integrated into the user environment, where they can provide services to their clients or patrons at the point of need and maintain a user-centric approach to library resources and services (Kesselman & Watstein, 2009).

The present study identified where Big Data resources assist embedded librarians to incorporate embedded librarianship in their organizations. The comprehensive, scalable, & dynamic nature of Big Data Services can meet the user's variable requirements physically or virtually. In this context, there is

a significant difference between reference librarian & embedded librarian. A reference librarian can help a user to fulfil his/her information need when they approach the librarian for the same. Moreover, when a student contacts a reference librarian for assignment help at the reference desk—physical or virtual—much time is spent by both the student re-explaining the assignment and the reference librarian, who must depend on the reference interview to understand his/her requirements and identify appropriate resources (Chesnut et al., 2009). In comparison, an embedded librarian integrates himself with the course curriculum and combines all the available resources into a single gateway (as a Big Data). As the embedded librarians are so directly involved in the course, they are able to identify potential points of need due to tricky assignments or troublesome assignment language and offer assistance proactively and to many more students at a time than in a traditional reference interaction (Lillard, 2003; Figa et al., 2009).

CONCLUSION

In recent years, especially at the time of COVID-19 lockdown libraries & librarians are offering online resources & services to their users proactively & enthusiastically. Libraries also started using social media and other online platform to promote their services and collecting updated information. Online information need to analyze to check their authenticity & accuracy with emerging technology and provide them to particular targeted user-base for adding value to their services. Moreover, to "walk with the footprints" of new-age information-inflation, Big Data can help libraries in better decision making, regarding demand-driven collection development, library space management & usage tracking. Big Data also helps in maintaining quality among regular & repetitive library works, viz. cataloguing, indexing, archiving. management, preservation & representation works. In conclusion, it can be said that, Big Data can help libraries make more cost-effective, innovative and informed decisions or recommendations for users, in order to contribute toward a more user-oriented service (Hamad et al., 2020).

The present study displays the steps to implement Big Data in libraries, possible Big Data library services, tools to analyze Big Data & international initiatives in a combine theoretical framework. Therefore, the possible extension of the contemporary study includes,

- the actual perception of Indian librarians about the introduction of Big Data services in their libraries (survey-based),
- opportunities & challenges in this introduction and,
- the systematic effect of Big Data on the financial status of the libraries. (Al-Barashdi & Al-Karousi, 2019)

References

• Al-Barashdi, H., & Al-Karousi, R. (2019). Big Data in academic libraries: literature review and future research directions. *Journal of Information Studies & Technology (JIS&T)*, *2018*(2). https://doi.org/10.5339/jist.2018.13

• Ball, R. (2019). Big Data and Their Impact on Libraries. *American Journal of Information Science and Technology*, *3*(1), 1. https://doi.org/10.11648/j.ajist.20190301.11

• Blummer, B., & Kenton, J. M. (2019). Big Data and Libraries: Identifying Themes in the Literature. *Internet Reference Services Quarterly*, *23*(1-2), 15–40. https://doi.org/10.1080/10875301.2018.1524337

• CCSDS. (2012). *CCSDS Recommended Practice for an OAIS Reference Model* (Recommended Practice, Issue 2, p. 20). CCSDS Secretariat. https://public.ccsds.org/Pubs/650x0m2.pdf

• Chesnut, M. T., Henderson, S. M., Schlipp, J., & Zai, R. (2009). Value-added library resources and services through Blackboard. Kentucky Libraries, 73(1), 6–12.

• DataONE. (n.d.). *Data Life Cycle | DataONE*. Retrieved January 10, 2021, from https://old.dataone.org/data-life-cycle#:~:text=The%20DataONE%20data%20life%20cycle%20has%20eight%20components%3A

• DDS. (2021). *The Importance of Libraries (and Librarians) to the Data Science Movement*. https://www.discoverdatascience.org/resources/data-science-and-librarians/#:~:text=Libraries%20can%20support%20data%20scientists

• Figa, E., Bone, T., & Macpherson, J. R. (2009). Faculty-Librarian Collaboration for Library Services in the Online Classroom: Student Evaluation Results and Recommended Practices for Implementation. *Journal of Library & Information Services in Distance Learning*, *3*(2), 67–102. https://doi.org/10.1080/15332900902979119

• Gibbons, S. (2013). Techniques to understand the changing needs of library users. *IFLA Journal*, *39*(2), 162–167. https://doi.org/10.1177/0340035212472846

• Hamad, F., Fakhuri, H., & Abdel Jabbar, S. (2020). Big Data Opportunities and Challenges for Analytics Strategies in Jordanian Academic Libraries. *New Review of Academic Librarianship*, 1–24. https://doi.org/10.1080/13614533.2020.1764071

• Hamad, F., Tbaishat, D., & Al-Fadel, M. (2016). The role of social networks in enhancing the library profession and promoting academic library services: A comparative study of the University of Jordan and Al-Balqaa' Applied University. *Journal of Librarianship and Information Science*, *49*(4), 397–408. https://doi.org/10.1177/0961000616656043

• HESA. (2019). *HESA - Experts in higher education data and analysis*. https://www.hesa.ac.uk/

• IFLA. (2018). *A concept data science framework for libraries*. https://www.ifla.org/files/assets/big-data/publications/a_concept_data_science_framework_for_libraries.pdf

• Impact. (2016, April 7). *The 7 V's of Big Data*. https://impact.com/marketing-intelligence/7-vs-big-data/

• Jain, A. (2016, September 17). *The 5 V's of Big Data*. Watson Health Perspectives. https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/

- Jantti, M. (2016). Libraries and Big Data. In *Quality and the Academic Library: Reviewing, Assessing and Enhancing Service Provision* (pp. 267–273). Chandos Publishing. https://doi.org/10.1016/b978-0-12-802105-7.00026-9

- Jisc. (n.d.). *Jisc: Powering world-class education and research*. https://www.jisc.ac.uk/

- Kesselman, M. A., & Watstein, S. B. (2009). Creating Opportunities: Embedded Librarians. *Journal of Library Administration*, *49*(4), 383–400. https://doi.org/10.1080/01930820902832538

- Kaladhar, A., Naick, B. R. D., & Rao, K. S. (2018). Application of Big Data Technology to Library data: A review. *International Journal of Library and Information Studies*, *8*(2), 25–30. http://www.ijlis.org/img/2018_Vol_8_Issue_2/25-30.pdf

- Kamupunga, W., & Chunting, Y. (2019). Application of Big Data in Libraries. *International Journal of Computer Applications*, *178*(16), 34–38. https://doi.org/10.5120/ijca2019918955

- Kiran, R. (2019, December 20). *What are the Characteristics of Big Data? | 5V's, Types, Benefits*. https://www.edureka.co/blog/big-data-characteristics/

- Lenny, D. (2001). *Application Delivery Strategies* (p. 949). META Group Inc. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

- Lillard, L. L. (2003). Personalized instruction and assistance services for distance learners: Cultivating a research relationship. *Research Strategies*, *19*(3-4), 204–212. https://doi.org/10.1016/j.resstr.2004.11.003

- MarketsandMarkets. (2020, March). *Big Data Market by Component, Deployment Mode, Organization Size, Business Function (Operations, Finance, and Marketing and Sales), Industry Vertical (BFSI, Manufacturing, and Healthcare and Life Sciences), and Region - Global Forecast to 2025*. https://www.marketsandmarkets.com/Market-Reports/big-data-market-1068.html

- Martin. (2011). *Big Data, BI, digital preservation and the cloud*. https://www.martinsights.com/?p=745

- Nunns, J. (Ed.). (2015, May 6). *10 of the most popular Big Data tools for developers*. Computer Business Review. https://www.cbronline.com/big-data/10-of-the-most-popular-big-data-tools-for-developers-4570483/

- Olendorf, R., & Wang, Y. (2017). Big Data in Libraries. In S. C. Suh & T. Anthony (Eds.), *Big Data and Visual Analytics* (pp. 191–202). Springer International Publishing. https://doi.org/10.1007/978-3-319-63917-8_11

- Pearlman, S. (2018). *What is Data Processing? Definition and Stages - Talend Cloud Integration*. https://www.talend.com/resources/what-is-data-processing/

- Reinsel, D., Gantz, J., & Rydning, J. (2018). *The Digitization of the World From Edge to Core*. https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

- Sengupta, S. (2016). Big Data: The Next Big Opportunity For Librarians. *International Conference on Contemporary Innovations in Library Information Science, Social Science & Technology for Virtual World*, 89–91. http://eprints.rclis.org/34349/1/final-proceeding-iicclist16-2s-view-50-51.pdf

- Shepley, S. E. (2009). Building a Virtual Campus: Librarians as Collaborators in Online Course Development and Learning. *Journal of Library Administration*, *49*(1-2), 89–95. https://doi.org/10.1080/01930820802312821
- Shumaker, D. (2009). Who let the librarians out? Embedded librarianship and the library manager. Reference & User Services Quarterly, 48(3), 239–257.
- Sonawane, A. S. (2018). Big Data: Application in Libraries. *International Journal of Scientific Research in Multidisciplinary Studies*, *4*(1), 22–23. https://www.isroset. org/pdf_paper_view.php?paper_id=529&6-LI-01.pdf
- Sugimoto, C. R., Ding, Y., & Thelwall, M. (2012). Library and information science in the Big Data era: Funding, projects, and future [a panel proposal]. *Proceedings of the American Society for Information Science and Technology*, *49*(1), 1–3. https://doi. org/10.1002/meet.14504901187
- Tableau Software. (2018). *Tableau: Business Intelligence and Analytics Software*. https://www.tableau.com/
- Wang, C., Xu, S., Chen, L., & Chen, X. (2016, June). Exposing library data with Big Data technology: A review. *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*. https://doi.org/10.1109/icis.2016.7550937