2020

Characteristic Assessment of Advancement in Duties of Librarians and Function of Libraries in Data Science Era

Naimat Ullah Shah Mr.
DLIS, Islamia University, Bahawalpur, naimat784@gmail.com

Salman Bin Naeem Dr.
DLIS, Islamia University Bahawalpur, salmanbaluch@gmail.com

Characteristic Assessment of Advancement in Duties of Librarians and Function of Libraries in Data Science Era

By

Naimat Ullah Shah

Ph.D. Scholar,
Department of Library and Information Sciences,
Islamia Univesity Bahawalpur
Corresponding author's email: naimat784@gmail.com
Contact Number: +923327775222

Dr. Salman Bin Naeem

Assistant Professor,
Department of Library and Information Sciences,
Islamia Univesity Bahawalpur

Dr. Rubina Bhatti

Professor,
Department of Library and Information Sciences,
Islamia Univesity Bahawalpur

Abstract

Data sciences usually involve data management, its utilization, distribution as well as its re-utilization. All these components need to be focused while targeting data science. Thus data puts a significant burden on research institutes because it is the authority that decides the responsible for the whole course of the procedure. It is of prime importance for data science librarians serving in data-centric age to know regarding LIS principles, theories, and other related skills that are mandatory for management and support of data science. This paper sums up the reviews of researchers regarding the data science era.

Moreover, this paper includes diagnostic assessment of data science environment concerning recent advancements in data science and progress in duties of librarians, presentation of detailed data, the function of data science libraries as well as librarians concerning data users. It is supposed to be an exciting era to work in a library as its role is expanding with specific new challenges. It is the need of the current period to educate librarians, library science researchers, and students regarding understanding, utility, and management of data to meet the requirements of data science librarians.

Key words: Data Science, Data management, Data librarians, Data Managers, Data Scientists, Data Centers.

Introduction

. Researchers are primarily engaged in data creation and analysis, but the decisions they take in deciding what formats to use to collect and store their data, what metadata they will use to describe it, who owns it, who has access, what software they will use to analyze it, what outputs there will be from the research, and countless other activities will have an impact further along the track lived as a usual attachment of digital data-management and electronic resource tasks. Still, the main thing is the level of awareness and skills on how to know about domain considerate needed. (Henty, 2008). This ‘data-centric era' has required that universities establish strategies, policies, setups, and data-centric services for data-management through assisting the scholars in developing, gathering, using, assessing, managing, and sharing the digital data-sets (Pinfield, Cox, & Smith, 2014).

Certain nascent research corners are unveiled every decade. Data analysis and computer sciences are going to make us witness some quickly adopted revolutionary advancements. With the appearance of nascent areas of research, specific novel terminologies are also come into being like data mining, deep learning, etc. In a broader sense, these are a particular generalization of existing fields. With the significance of such generalizations, not only the appearance of new terms happens, but it also serves as an augmentation impulse for the quick progress of the field. Data science, being one of the newly established terms, is looking for a more targeted and purified definition (Dzemyda, 2018). The notion of data is the basis on which the entire building of data science relies. Data scientists have concluded that data are not the only goal-oriented but also neutral in itself. Information studies have defied this perception by regarding data be bias from inside. Scientific scholars have also concluded the same. The word data is described as the statistical observations and findings or collection and recording of proofs or a succession of the detached conclusions and observations (Zins, 2007).

What is data?

In a broader sense, the data may be defined as the assortment of digits or sequence of text or a thread of alphanumeric codes, which will not necessarily possess a firm meaning. Data itself rarely holds a unique value (Frederick, 2016). Data is an element or bit of information which has the observed volume to be gathered and kept in a way to be recognized, used again, and again.

"Data is famously divided into two types' structured-data and unstructured-data. (Ramkumar, 2018).

Data is regarded as of prime importance in research activities because of its towering worth in information. It is the data that concerns information to possess a level of significance that may be considered as critical for deciding on a particular matter. Data is supposed to have appeared in a variety of ways, including surveys, statistical declarations, research outcomes, graphical presentations, figures or illustrations, interviews, journal statements, etc (Mikalef, Pappas, Krogstie, & Giannakos, 2018). Data is considered as a vital factor for science and technology. It plays a crucial job in the generation of hypothesis, it is assessment and making of any desired amendments in a belief, putting forward of theories and suggesting models. It is data that has made it possible for technology to be practiced and show applications in day to day life (Oyelude, 2017).

Noval Ways of Data Implication

Presently the research studies are more data-centric, and scholars are facing novel type of barriers in accessing, disseminating and management of digital data, so libraries are starts to offering data services, which includes training and instructing, management of data, planning and data guidance, data-curation and stewardship, and data imagining (Federer, 2018). Science is not only the owner and creator but also the managing authority of the most significant share of universal data. The general opinion of the public regards data to be brought into existence by the scientist community for scientific purposes. However, if we talk about data concerning digital stuff, it may be utilized for multiple intellectual goals (Clement, Blau, Abbaspour, & Gandour-Rood, 2017).

Similarly, the Data may also be employed for the production of nascent ideologies, novel protocols as well as invention and deployment of scientific information in novel ways for the significantly improved customary computational, experiment-based, physical observations based as well as theoretical tracks for the discovery of new experimental stuff and their utilities for the public. The procedure can be initiated generation of an idea followed by the scheduling of record for data collection as well as its storage for future services. In academia, such initiation involves an advance proposal as the first step. Individuals from scientific backgrounds are usually not

well trained regarding the preservation techniques of data for more extended periods (Cao, 2018).

Literature review

Historical Development of Data Science

Data Science has been acknowledged explicitly in the present era as an area of study that deals with vast data stores. As far as the Data is concerned, an enormous variety of perceptions are suggested by Data scientists and experts of its allied areas, including statistics, library science, and computer science for long. Back in the late 5th decade of the 20th century, a nascent term Bit was introduced by Turkish scientists and was phrased by Claude Shannon in one of his papers (Shannon, 1948). After being presented in the article "A Mathematical Theory of Communications" by Claude Shannon, another researcher John W. Tukey mentioned it in his paper "The Future of Data Analysis" (John W Tukey, 1962). In the year 1977, J.W. Tukey put a vast emphasize on the utilization of data in various phases of research, including hypothesis, assessment, and the final analysis(John Wilder Tukey, 1977). The *International Federation of Classification Societies(IFCS)* biennial conference held at Kobe, Japan, became the pioneer who uses the Data Science term, which included in conference title (Davenport & Patil, 2012). It is said supposed that the initial introduction of data science as a term in the official documents and manuscripts was in 1974 in the preamble to Nauru book "Concise Survey of Computer Methods" (Belzer, 1976). In the intro of publication, the definition of data science is mentioned that science which involves dealing with the data just after their establishment as well relating of this data along with their findings and presentations to other subfields of data science. Word data science was coined by statisticians academically to point out the subject as "big data, data analysis, and broader trends," which highlights statistical foundations and the new mathematical techniques to make sure by the richness of data (Gilmore, 2016).

Data Science Era

There is the greatest challenge for science in the twenty-first century is that how do we report to the new data science era. In experimental and theoretical research, data science is recognized as an emerging model and also in computer and information science recreations of natural phenomena, which is required new ways, techniques, and tools of working (Peyne &

Chan, 2017). Data Science, being also known as e-science, is regarded as a combination of scientific procedures and technologies that assist in data alliance and collaboration for making it possible to assess as well as explore the data along with the supporting of communication and propagation between the scholars (Antell, Foote, Turner, & Shults, 2014). Scientific data is mainly composed of every corner of natural, artificial, and socials systems, all of which act as a source for obtaining experimental data. The data collected from these components or sources is passed via a chain of scientific protocols to get the desired goals. The scientific data is a source of immense opportunities for the data sciences, introducing it to a variety of new prospective (Federer, 2016). The data has to be transformed into a compatible and useful format before exposing it to be processed by a computer. That is regarded as the equation's input portion. It is preferred to be obtained in a useful context. This is considered to be the equation's input portion. The output is supposed to be associated with the input, just like suggested by the adage, "Garbage in, garbage out" (Balachandran & Kamalanathan, 2018).

It is evident that still, the importance of data science and its analytical approaches in enabling data-driven theory, economic progress, and professional improvement is going to be accepted with much more potential than ever. This is composed of not only the necessary fields of computer, information technology, and statistics but also specific other disciplines like business, social sciences, and health sciences (Cao, 2018). The establishment of a systemic educational model is mandatory to train our future's data professionals (including engineers, professionals, scientists as well as executives) that may make them capable to have an intellectual attachment with data, gain expertise regarding management of data, have a hand on computation of data, gain abilities to mine data, can successfully propagate data, know regarding the delivery of data and know-how and when to take action on data (Waller & Fawcett, 2013). It is an approach to research studies of “big data, predictive analytics, and data science" which is the requirement of researchers with the field of information and data management (Carillo, 2017)

LIS in Data Science Era

The data-science was only be cast as a discipline if a specified foundation is established for it. Moreover, advancements in technological fields, specification of subject areas, development of research map and goal, and establishment of practical tools are also considered as vital factors for such purposes.

The progress of information science research and education is going on with such a pace in LIS schools that can be regarded as the leading reign of the development of these fields. Keeping in view the challenges faced by and opportunities provided by data science, we are apparent that information science will pay incredible gifts to the area of data science and collaboration. These two will undoubtedly benefit society (Wang, 2018). Data science, along with a considerable extent of data and its applications, has become known as a nascent field in the present decade. Some i-schools have been established with fewer schools of library and information sciences (Zuo, Zhao, & Eichmann, 2017). Data science involves a variety of "Data Issues" that resides in cybernetics, information theory and system sciences which are supposed to be the customary basis of information system and convert them into a logical and theoretical idea. So data science is going to replace the conventional theories into a nascent theory source of IS presenting novel speculative references for it (Ge, Bangui, & Buhnova, 2018). "Data chain and information chains are intrinsically associated with each other. Basis of both of them relies on extracting value as well as insight from data" (Cai & Zhu, 2015). Data science entirely incorporated with information science. We may conclude that data can be proved as a research object for information sciences. Moreover, these both may have tremendous attributes alike. However, massive data never always represents an enormous potential for LIS School. In case we manage to successfully prove the association and common aspects in data and information studies, it can be possible to propose to data studies may become a component of information science and may lead to a beneficial mode of education in LIS schools (Wang, 2018).

Role of Librarians in Data Science Era

The role and responsibilities of librarians have greatly changed. The involvement of data science in data management libraries has become an integral portion of the research protocol from the very start of the life cycle (Walek, 2017). The nascent job title of data librarians has come into existence in such a data science scenario. The initiation of data librarians may be linked to social sciences, bioinformatics, and data management research activities (Koltay, 2017). Currently, the progress of data library professionals, particularly those related to academics and research, is seen to be very fast. Data librarian is actually produced by an intellectual broadening of the term of an academic librarian, keeping in view the new

requirements, research policy needs, and nascent data management performances (Federer, 2016).

According to Lyon & Mattern, (2017) stated that "*There are six roles of data science managers they are: data archivist, data curator, data librarian, data analyst, data engineer, and data journalist*." The mentioned six characters are enclosed as data-science characters, some of the other roles whose derives by the commercial site, which define just analysts of data kind job oriented as data-scientists.

Data Science works as a critical element in information centers, so building librarians with data-science techniques enhance their skills in keeping up this data-centered studies (Burton, Lyon, Erdmann, & Tijerina, 2018). Data can become familiar as the earliest period article, which may experience tests, peered review, and circulation, use again and again, and is mainly under consideration. The data reuse blesses the society in the capacity of reuse of an idea in a journal article. The libraries and its staff have been found to pay duties of collection, preservation, and dissemination of intellectual results of the society (Carlson & Johnston, 2015). Past studies reveal the absence of an agreed guideline regarding the duties and requirements with respect to data science qualification, skills, and capabilities that are supposed as prerequisites for the position of "Data Librarian" (Yoon & Schultz, 2017). Librarians should also be skillful in handling and assessing technologies for the future data-science era. Usually providing services through technology, the literature suggests that library information science may require specific care in outgoing, and gaining undesirable approaches towards technology (Lyon & Mattern, 2017).

Functions of Libraries in Data science Era

With the progressive involvement of digital data dealing, the libraries have either made the boarders of professional duties of their existing staff more full by making them endow data services to their concerned groups or have managed to appoint a full time and committed additional professional team that has expertise in rendering data services (Lyon, 2012).

Libraries are the sites that are expected to possess the skill sets, durability and the essential skeleton for the accomplishment of tasks for a variety of data. In case of negligence from libraries in engagement in such tasks, the public will alternatively originate specific new institutional setup to deal with digital data. This situation will mainly direct us towards the textual output and related data's dissociation that will, unfortunately, result in deprivation of society to be benefited from its own knowledgeable assets (Heidorn, 2011). Libraries usually possess an attraction for data. Almost the men's entire actions make libraries need data for the accomplishment of their technical tasks. In the absence of data, the needy person won't be able to meet the desired information. Circulation of books and payments of staff may also be disturbed (Frederick, 2016). Current Emerging technologies in stats and software engineering tethered with a richness of information have provided a novel expert bio network known as data science. Data-science techniques, tools can change trade, health science, and management, and it will also change libraries like other areas. *"There is the need for Investments to enlarge the recent flow of support in the area of data science"* world is more and more imbued with information, information professional was a critical job in growth and future of the data-science (Burton et al., 2018). Data science's fundamentals specifically theory basis and procedures for analyzing and gaining knowledge towards certain corners, including numerical computation, algebra, geometry, graph, probability, and philosophy of information science (Van Der Aalst, 2016).

Objective of the study

- To diagnose the environment concerning recent advancements in data science in libraries with the help of related literature.
- To assess the duties of librarians and function of libraries in data science era with the help of related literature.

- To dig out the current advancement in libraries in data science era with the help of related literature.

Methodology

This paper sums up the reviews of researchers ideas and thoeries regarding the data science era. A theoretical lens was drawn from insightful articles, books and some other sources applicable to this specific issue, core of research, or hypothesis, and gives a basic assessment of the previously works done in connection to the research objectives being explored.

Concluding Discussion

The data scientists are not likely to come from conventional library backgrounds, predominantly career researchers who are gone via a period of work as a data scientist as a mandatory part of long term research career track. Who will possibly provide such advice? It is probably the data librarian serving in university, who may also accomplish a task of administration of local data sets for small projects, the need that may be held in case the considerable gaps in national level stipulation are stopped in the future.

It is the need of the current era to educate librarians, library science researchers, and students regarding understanding, utility, and management of data. Science is not only the owner and creator but also the managing authority of the most significant share of universal data.

There is a need for the current scenario that information professionals inquire about what data science roles can fill out through existing assets by librarians. Libraries are in an excellent position to contribute to providing the necessary infrastructure needed for support of data science, particularly admittance to data, utilization, reuse, storage, and preservation of data. It is supposed to be engaging in the data science era to work in a library as its role is expanding with specific new challenges.

References

Antell, K., Foote, J. B., Turner, J., & Shults, B. (2014). Dealing with data: Science librarians' participation in data management at the Association of Research Libraries institutions. *College & Research Libraries, 75*(4), 557-574.

Balachandran, K., & Kamalanathan, S. (2018). Data Management in Clinical Research. *Thesis Writing for Master's and Ph. D. Program*, 83-92.

Belzer, J. (1976). A concise survey of computer methods. Peter Naur. New York: Petrocelli Books, 397 p.(1975). *Journal of the American Society for Information Science, 27*(2), 125-126.

Burton, M., Lyon, L., Erdmann, C., & Tijerina, B. (2018). Shifting to Data Savvy: The Future of Data Science In Libraries.

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal, 14*.

Cao, L. (2018). *Data Science Thinking: The Next Scientific, Technological, and Economic Revolution*: Springer.

Carillo, K. D. A. (2017). Let's stop trying to be "sexy"–preparing managers for the (big) data-driven business era. *Business Process Management Journal, 23*(3), 598-622.

Carlson, J., & Johnston, L. R. (2015). *Data information literacy: Librarians, data, and the education of a new generation of researchers* (Vol. 2): Purdue University Press.

Clement, R., Blau, A., Abbaspour, P., & Gandour-Rood, E. (2017). Team-based data management instruction at small liberal arts colleges. *IFLA Journal, 43*(1), 105-118.

Davenport, T. H., & Patil, D. (2012). Data scientist. *Harvard business review, 90*(5), 70-76.

Dzemyda, G. (2018). *Data Science and Advanced Digital Technologies.* Paper presented at the International Baltic Conference on Databases and Information Systems.

Federer, L. (2016). Research data management in the age of big data: Roles and opportunities for librarians. *Information Services & Use, 36*(1-2), 35-43.

Federer, L. (2018). Defining data librarianship: a survey of competencies, skills, and training. *Journal of the Medical Library Association: JMLA, 106*(3), 294.

Frederick, D. E. (2016). Data, Open Science and libraries–The Data Deluge Column. *Library Hi Tech News, 33*(8), 11-16.

Ge, M., Bangui, H., & Buhnova, B. (2018). Big data for internet of things: A survey. *Future Generation Computer Systems, 87*, 601-614.

Gilmore, R. O. (2016). From big data to deep insight in developmental science. *Wiley Interdisciplinary Reviews: Cognitive Science, 7*(2), 112-126.

Heidorn, P. B. (2011). The emerging role of libraries in data curation and e-science. *Journal of Library Administration, 51*(7-8), 662-672.

Henty, M. (2008). Dreaming of data: the library's role in supporting e-research and data management.

Koltay, T. (2017). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science, 49*(1), 3-14.

Lyon, L. (2012). The informatics transform: Re-engineering libraries for the data decade. *International Journal of Digital Curation, 7*(1), 126-138.

Lyon, L., & Mattern, E. (2017). Education for Real-World Data Science Roles (Part 2): A Translational Approach to Curriculum Development. *International Journal of Digital Curation, 11*(2), 13-26.

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and e-Business Management, 16*(3), 547-578.

Oyelude, A. A. (2017). Data, data everywhere, but little to use! *Library Hi Tech News, 34*(8), 24-25.

Peyne, B., & Chan, A. (2017). Data-driven decision making in Marketing: A theoretical approach.

Pinfield, S., Cox, A. M., & Smith, J. (2014). Research data management and libraries: relationships, activities, drivers and influences. *PLoS One, 9*(12), e114734.

Ramkumar, P. (2018). Initiatives of Research Data Literacy for Data Management in Libraries. *Journal of Advances in Library and Information Science, 7*(1), 128-131.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal, 27*(3), 379-423.

Tukey, J. W. (1962). The future of data analysis. *The annals of mathematical statistics, 33*(1), 1-67.

Tukey, J. W. (1977). *Exploratory Data Analysis: Limited Preliminary Ed*: Addison-Wesley Publishing Company.

Van Der Aalst, W. (2016). Data science in action *Process Mining* (pp. 3-23): Springer.

Walek, A. (2017). Is data management a new “digitisation”? A change of the role of librarians in the context of changing academic libraries’ tasks.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics, 34*(2), 77-84.

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

Yoon, A., & Schultz, T. (2017). Research data management services in academic libraries in the US: a content analysis of libraries’ websites.

Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American society for information science and technology, 58*(4), 479-493.

Zuo, Z., Zhao, K., & Eichmann, D. (2017). The state and evolution of US iSchools: From talent acquisitions to research outcome. *Journal of the Association for Information Science and Technology, 68*(5), 1266-1277.