

Big data adoption in academic libraries: a literature review

by Nove Eka Variant Anna

Submission date: 10-Feb-2023 10:44AM (UTC+0800)

Submission ID: 2010578647

File name: Big_data_adoption_in_academic_libraries_a_literature_review.pdf (98.68K)

Word count: 3619

Character count: 19444

LIBRARY HI TECH NEWS

VOLUME 37 NUMBER 4 2020
MAY

AN EMERALD PUBLICATION

incorporating Online and CD Notes

LIBRARY HI TECH NEWS NUMBER 4 2020

CONTENTS

Feature article

2 Big data adoption in academic libraries: a literature review1

Colum

2 Getting acquainted with social networks and apps: it is time to talk about TikTok7

Professional literature

2 A new full-text finder tool for linking to scientific articles13

Colum

Incorporating online privacy skills into one-shot sessions.15

Through the looking glass

Books online: e-books, e-paper, and e-readers19

Library Link

www.emeraldinsight.com/librarylink

Big data adoption in academic libraries: a literature review

Nove E. Variant Anna and Endang Fitriyah Mannan

Introduction

Bibliometric studies provide a map of knowledge that is useful for researchers to see which research themes have been widely studied or not, making it easier for researchers to determine the direction of research conducted. Bibliometric studies have been carried out in various fields of research and until now using special software to analyse the bibliographic data. Many scientists and practitioners have reviewed the possibility of implementing big data in libraries. Before discussing big data in the library, there were many bibliometrics researches with the topic of the library, among others [Barik and Jena \(2019\)](#) who analysed publications from ten open access journals in library science. The purpose of this research is to find out since big data in library articles were written, which countries frequently write about big data and keywords related to data in the library.

Big Data, according to [Dumbill \(2013\)](#) in [Ali \(2015\)](#), is data that exceeds the capacity of conventional database system processing. The data is big, moves fast or does not fit the structure of a database architecture. To get value from this data, alternative way must be chosen to process it. While another definition is mentioned by [Heer and Kandel \(in Lemieux, 2014\)](#):

what volume the data becomes big remains an open question, however, with some suggesting that it comprises data at

the scale of exabytes, while others argue for zettabytes or yottabytes.

[St. Nath \(2015\)](#) in [Ali \(2015\)](#) also defines big data as follows:

Big data is an evolving term that describes any voluminous amount of structured, semi-structured and unstructured data that has the potential to be mined for information.

Big Data has characteristics as follows: 3V; volume, variety, velocity and other V elements such as veracity and value. Volume (data capacity) is related to the size of data storage media, which is very large or may be unlimited to bytes or zettabytes; variety (diversity of data) is related to the type of data that can be processed, starting from structured data to unstructured data; while velocity (speed) is related to the speed of data processing generated from various sources, ranging from batch data to real time, while the characteristics of veracity (truth) and value are related to the uncertainty of the data and the value of the benefits of the information generated ([Sirait, 2016](#)).

Big data is a phenomenon caused by the swift flow of information. The issue of big data in libraries has begun to be widely discussed. While its application in the public or government sector, seems to be still limited. According to [Ali \(2015\)](#) big data requires libraries to improve, among other things, library organizations and internal data improvements that cover both electronic and printed collections. One

of the important issues is the improvement of librarians with special skills. In addition, according to Jin *et al.* (2015) every organisation, including libraries, that wants to succeed in big-data projects must be clearly related to their requirements and work efficiently with big data. Some librarian jobs that have evolved from the basic librarian work include data management consultant, data mining consultant, data research scientist, librarian data services, librarian data design, digital archivist, digital collections, strategist and architecture librarian, digital humanities design consultant, digital records archivist manager, data management services, librarian data research, research data management coordinator, scientific data curation, librarian specialist/metadata, scientific data curator and social science data consultant. Changes in the competence of librarians in the digital age require them to update their knowledge and skills so that they are in line with the changes that occur.

Researching big data in libraries in the literature

The Scopus database was accessed on 24 May 2019 by using the keyword “big data” AND “library” in the search box. There were 74 papers, however, 1 article was dropped because of it not meeting the criteria (affiliation and abstract were not available). The papers consist of journal articles, conference papers, book chapters, editorial and review. Then the data were extracted into excel and analysed as follows (by the year, by the author/s’s country, by the theme and by the publisher). Following that the collected data were analysed using VOX viewer software to see the relationship between big data terminology and library, terminology clustering, keywords that often appear, countries that publish big data, number of big data authors, year of publication and name of journals that publish big data and library articles (Alagu and Thanuskodi, 2019).

A visualization using VOXviewer software including important words in abstracts of big data articles and libraries. The results of the visualization show the relationship of words or terms that appear in big data in library articles.

There are four clusters of terminology or keywords that are interrelated with big data and libraries, which are distinguished by green, red, blue and yellow colour. In the green cluster, there are the words “university”, “tool”, “role”, “knowledge”, “researcher”, “data set” and “government”. According to the papers, big data analytics is implemented in academic libraries, in which in academic libraries the data are very complex and huge. Academic libraries are recognized as a research library that provides many data and knowledge for their users. There are several clusters the first of which includes “big data era”, “innovation”, “change”, “internet”, “model” and “library service”. This cluster shows that in the big data era is related to innovation and rapid change and it is supported by the internet to connect data, information and people. The second cluster focuses on the words “algorithm”, “collection”, “fact”, “project”, “hadoop”, “volume” and “variety”. The next cluster focus on big data technology such as software they use, algorithm to manipulate the data, the volume and variety of big data. Finally there is a cluster that is represented by the words “information science”, “design methodology approach” and “practical implication”. The last cluster shows the information science, practical implication of big data and methodology that they used. The four clusters are tangent to each other. It is said that the big data in library comprise of the academic libraries and library services and researchers, the software that they used, it strong relation to innovation and change and implementation in the library.

Keywords that often appear based on the results of terminology visualization are marked by thickening of the circles and text in the Voxviewer, such as “big data”, “libraries”, “libraries”, “data handling”, “data mining” and “university libraries”, “digital libraries”, “academic libraries”, “big data application” and “data management”. Writing about big data in libraries related to data handling, data mining, data management, big data applications where the article discusses use of data for management purposes in the library. Implementation of big data is found in many college or academic libraries but rarely discussed in public libraries. Thus, with the era of big data,

there are many opportunities for libraries solve problems that arise by processing data with the help of technology. The authors of articles on big data in libraries that published at least three papers were China (32), Poland (3), UK (3) and the USA (14).

Current research trends refer to data-based research. This provides an opportunity for librarians to develop their skills in the field of information technology (Farida, 2018). One example of research on big data in a library was conducted by Teets and Goldner (2013), focusing on WorldCat OCLC database from a global library collection and includes works by other libraries so that they may expose their data using big data concepts and standards.

In terms of publication, writings on the topic of big data in the library that are indexed by Scopus are written in five different forms, namely articles, conference papers, book chapters, editorials and reviews. The majority of writings are in the form of articles (47.94 per cent) and conference paper (31 per cent). Not many books, editorials and reviews discuss the topic of big data in the library. Some of the conference titles containing big data topics are as follows: International Conference on Data Engineering, International conference on Computational Science and Engineering 2013, International Conference on High Performance Computing and Simulation, International Conference on Fuzzy Systems, International Conference on Cloud Computing, International Conference on Semantics, Knowledge and Grids, and many more. There are various international conferences focusing on computer science and engineering, the results of the data also show that the publication of big data from the international conference results in the field of library has yet to be done before. This is one of the gaps in the field of library science where a scientific communication forum is needed to discuss big data and its implementation in the library.

There are currently many conferences and call for paper on the topic of big data, for example, the ALIEP conference and the 2019 ICADL conference. In the conference, one of the sub-themes was big data and data science. Universitas Sebelas Maret Surakarta library held an International

Seminar with the theme “Big Data: Building and Disseminating Research in Higher Education” in 2019. This activity was held with the aim of providing insights on the importance of research in an educational institution, especially tertiary institutions, so as to create a sustainable and quality research atmosphere through the role of librarians in the industrial revolution era 4.0. The seminar was attended by lecturers, students, librarians and researchers from various regions in Indonesia. However, the international seminar did not specifically target libraries and big data.

One article on the topic of big data in the library first appeared in 2012. Then every year the number of writing about big data in the library increases and reaches its peak in 2018, which were the last year studies. This shows that the big data issue has only been popular in the past five years. It can be predicted that in 2019 onwards, the number of articles about big data in libraries will continue to grow.

Looking at the history of the development of big data, around 2005, people began to realize that there was a number of user-generated data through Facebook, YouTube and other online services. Hadoop, an open-source software for collecting and analysing data was developed in the same year. NoSQL has also begun to gain popularity so far. The development of social media also has a strong influence on the existence of big data.

The data shows the top five journals indexed by Scopus, which contain writings about big data in the library. First is Library Hi Tech with five articles (6.84 per cent), followed by the advanced in intelligent systems and computers journal containing four titles (5.47 per cent). Hi-tech Agro food industry journal and advanced material research journal with three articles each (4.1 per cent) and Journal of Advanced Oxidation Technology with two articles (2.73 per cent).

The results of big data research in the library still involves around the discussion on the concept of big data, the essence of big data for libraries, the view of the head of the library related to big data and its implementation in the library. Many library leaders do not understand the importance of big data

for the library, whereas experts say that big data is very important for library development. Furthermore, Golub and Hansson (2017) provide an argument that research related to big data in libraries is a great opportunities for academics to develop libraries, especially the need for open access to information for the public. There are still many research gaps that can be investigated. One of the results of implementing big data in a library is to study user patterns in using the library so that an appropriate library service will be obtained and Jantti (2016) analyses the use of the value cube to integrate all library data at the University of Wollongong. The era of big data brings new challenges for libraries, especially in providing data and information for service development. The use of data processing tools and data processing techniques with new methods is also a serious challenge for libraries (Chuífeng, 2013). With big data, libraries can develop various services including knowledge services, user failure analysis, value analysis, intelligent-aided decision-making and forecast resource faults (Weihong *et al.*, 2012).

Libraries can also use data originating from outside the library. Sun and Ma (2014) have a concern on the security and privacy of library data in the era of big data. Big data also encompasses library services with the concept of big service and smart library in tertiary institutions proposed by Zhao *et al.* (2018). However, information technology in libraries are unable to handle large amounts of data, both structured and unstructured (Ye, 2017). Some big data analytics software can also be used by libraries to process data from social media such as social networks analysis, statistical analysis and geographic coding that can be used to assess the effectiveness of using social media in a library (Cervone, 2017). In its implementation, the use of data for decision-making is also rarely used, there are still subjectivity factors that determine the decision-making policies in the library (Johnson and Cassady, 2018). The role of librarians in implementing big data in libraries is considered quite important, in collaboration with data analysts, data architects and IT specialists, to build library databases. The job of the

librarian here is to determine metadata, create a thesaurus and create search parameters (Johnson, 2017).

Research on big data is very closely related to digital libraries. Libraries are expected to be able to mine data that is from outside the library (Yingying, 2018). Bhat, 2018, emphasizes that the current condition of many digital libraries is that they do not have data storage that is able to play a digital preservation function in the library. Llinin (2017) once mentioned that changes in information processing in libraries will occur in the era of big data, where everything can be done virtually from virtual training, virtual classrooms and virtual experiments.

Big data and public libraries have not been researched until now. Big data is still widely associated with university libraries. There are also a number of librarians who do not understand the concept of big data-related libraries (Zhan and Widen, 2018). However, Kim and Cooke (2017) have conducted research using big data analysis on public library operations in London and Seoul, and learned that big data software to analyse data from both libraries provides a new breakthrough in terms of analysis and offers a new way of evaluating library services quickly and effectively.

Concluding remarks

The implementation of big data in libraries is still in an early stage, it is shown from the limited number of in the library literature as of 2018. Not many libraries use big data to support innovation and services since there have been lack of skills. Librarians can enhance their skills and knowledge by following some training in big data analytics or research data management. The information technology infrastructure also needs to be upgraded since big data need big IT capacity. Finally, a big data management policy should be made to ensure implementation and sustainability.

REFERENCES

Alagu, A. and Thanuskodi, S. (2019), *Bibliometric Analysis of Digital Literacy Research Output: A Global Perspective*, Library Philosophy & Practice.

- Ali, I. (2015), "Big data: Apa dan pengaruhnya pada perpustakaan?", *Jurnal Media Pustakawan*, Vol. 22 No. 4, pp. 19-23.
- Barik, N. and Jena, P. (2019), "Bibliometric portrait of select open access journals in the field of library and information science: a scopus based analysis", *Library Hi Tech News*, pp. 1-18.
- Bhat, W.A. (2018), "Long-term preservation of big data: prospects of current storage technologies in digital libraries", *Library Hi Tech*, Vol. 36 No. 3, pp. 539-555.
- Cervone, H.S. (2017), "Evaluating social media presence: a practical application of big data analytics in information organization", *Digital Library Perspectives*, Vol. 33 No. 1, pp. 2-7.
- Farida, U. (2018), "Pengelolaan big data pada perpustakaan: tantangan bagi pustakawan di era perpustakaan digital", *Journal Net. Library and Information*, Vol. 1 No. 1, pp. 19-29, available at: <http://e-journal.sari-mutiara.ac.id>
- Golub, K. and Hansson, J. (2017), "(Big) data in library and information science: a brief overview of some important problem areas", *Journal of Universal Computer Science*, Vol. 23 No. 11, pp. 1098-1108.
- Jantti, M. (2016), "Libraries and big data", *Quality and the Academic Library*, pp. 267-273.
- Jin, X., Wah, B.W., Cheng, X. and Wang, Y. (2015), "Significance and challenges of big data research", *Big Data Research*, Vol. 2 No. 2, pp. 59-64.
- Johnson, C. and Cassady, S. (2018), "How librarians make decisions: the interplay of subjective and quantitative factors in the cancellation of big deals", *Collection and Curation*, doi: [10.1108/CC-05-2018-0013](https://doi.org/10.1108/CC-05-2018-0013).
- Johnson, V. (2017), "Leveraging technical library expertise for big data management", *Journal of the Australian Library and Information Association*, Vol. 66 No. 3, pp. 271-286.
- Kim, Y.S. and Cooke, L. (2017), "Big data analysis of public library operations and services by using the chernoff face method", *Journal of Documentation*, Vol. 73 No. 3, pp. 466-480.
- Linlin, Z. (2017), "Analysis of the information processing technology of university libraries in the big data era", *Agro Food Industry Hi-Tech*, Vol. 28 No. 1, pp. 2036-2040.
- Sirait, E.R.E. (2016), "Implementasi teknologi big data di lembaga pemerintahan Indonesia", *Jurnal Penelitian Pos Dan Informatika*, Vol. 6 No. 2, pp. 113-136.
- Sun, N.N. and Ma, L.H. (2014), "Research on information security and privacy of libraries in big data era", *Advanced Material Research*, Vols 1049/1050.
- Teets, M. and Goldner, M. (2013), "Libraries' role in curating and exposing big data", *Future Internet*, Vol. 5 No. 3, pp. 429-438.
- Weihong, F., Chenhui, L., Xingwang, Z., Xiaozhu, Q. and Zikuan, G. (2012), "How do libraries need 'big data'?", *Library Journal*, Vol. 11.
- Ye, C. (2017), "Research on the key technology of big data service in university library", *13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, IEEE, pp. 2573-2578.
- Yingying, W. (2018), "Research on the construction of digital library under the background of the era of big data", *Journal of Advanced Oxidation Technologies*, Vol. 21 No. 2.
- Zhan, M. and Widen, G. (2018), "Public libraries: roles in big data", *The Electronic Library*, Vol. 36 No. 1, pp. 133-145.
- Zhao, J., Cai, W. and Zhu, X. (2018), "Research on smart library big service application in big data environment", *International Symposium on Computational Science and Computing*, Springer, Cham, pp. 238-245.
- and Practice, 1-8.08-2012. *Library Philosophy and Practice*, p. 1020.
- Darmandeh, M., Noruzi, A. and Givi, M.E. (2019), "Opportunities of big data management in libraries and information centers: structural-interpretive analysis and finding a solution", *Iranian Journal of Information Processing Management*, Vol. 34 No. 2, pp. 841-870.
- Li, J., Lu, M., Dou, G. and Wang, S. (2017), "Big data application framework and its feasibility analysis in library", *Information Discovery and Delivery*, Vol. 45 No. 4, pp. 161-168.
- Li, S., Jiao, F., Zhang, Y. and Xu, X. (2019), "Problems and changes in digital libraries in the age of big data from the perspective of user services", *The Journal of Academic Librarianship*, Vol. 45 No. 1, pp. 22-30.
- Quan, Y. and Lei, C. (2019), "Use of the internet technology in the intelligent library services of colleges and universities library in the big data era", *Advances in Intelligent Systems and Computing*, Conference Paper.
- Rufaidah, V.W. (2008), "Kolaborasi dan graf komunikasi artikel ilmiah peneliti bidang pertanian: Studi kasus pada jurnal penelitian dan pengembangan pertanian", *Jurnal Perpustakaan Pertanian*, Vol. 17 No. 1, pp. 10-21.
- Xiaodan, D. and Wei, W. (2017), "Discussion on university library service pattern in big data era", *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, IEEE, pp. 597-600.
- Xu, S., Du, W., Wang, C. and Liu, D. (2017), "The library big data research: status and directions", *International Journal of Software Innovation*, Vol. 5 No. 3.
- Zhang, J. (2018), "Design of university library and information management system based on big data fusion", *MATEC Web of Conferences*, Vol. 232, EDP Sciences, p. 1010.

FURTHER READING

- Ahmad, K., JianMing, Z. and Rafi, M. (2019), "An analysis of academic librarians competencies and skills for implementation of big data analytics in libraries: a correlational study", *Data Technologies and Applications*, Vol. 53 No. 2.
- Ahmed, W. and Ameen, K. (2017), "Defining big data and measuring its associated trends in the field of information and library management", *Library Hi Tech News*, Vol. 34 No. 9, pp. 21-24.
- Anna, N.E.V. Mannan, E.F. Sri Rahayu, D.P. and Mutiai, F. (2018), "Library and information (LIS) research topics in Indonesia from 2006 to 2017", *Library Philosophy and Practice*, pp. 1-7.
- Bhakta, J. Kar, S. and Bhui, T. (2019), "Bibliometric mapping of LIS research in India: a study seen through the mirror of Indian citation index", *Library Philosophy*

Nove E. Variant Anna (nove.anna@vokasi.unair.ac.id) and **Endang Fitriyah Mannan** (endang.mannan@vokasi.unair.ac.id) both are based at the Universitas Airlangga, Surabaya, Indonesia.

Big data adoption in academic libraries: a literature review

ORIGINALITY REPORT

17%

SIMILARITY INDEX

14%

INTERNET SOURCES

13%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1	jurnal.darmajaya.ac.id Internet Source	2%
2	alibnet.org Internet Source	2%
3	"Big Data and Blockchain for Service Operations Management", Springer Science and Business Media LLC, 2022 Publication	1%
4	fliphtml5.com Internet Source	1%
5	"Big Data in Complex Systems", Springer Nature, 2015 Publication	1%
6	www.slideshare.net Internet Source	1%
7	gitbucket.inist.fr Internet Source	1%
8	fv.conference.unair.ac.id Internet Source	1%

9

Faten Hamad, Hussam Fakhuri, Sinaria Abdel Jabbar. "Big Data Opportunities and Challenges for Analytics Strategies in Jordanian Academic Libraries", *New Review of Academic Librarianship*, 2020

Publication

1 %

10

[dokument.pub](#)

Internet Source

1 %

11

Sukjin You, Soohyung Joo, Marie Katsurai. "Data mining topics in the discipline of library and information science: analysis of influential terms and Dirichlet multinomial regression topic model", *Aslib Journal of Information Management*, 2022

Publication

1 %

12

[www.emerald.com](#)

Internet Source

1 %

13

Nur Izzuddin, Ajib Setyo Arifin. "Analysis of Big Data Development in Jakarta Smart City", 2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA), 2018

Publication

1 %

14

Adeyinka Tella, Kehinde Khadijat Kadri. "Big data and academic libraries: is it big for something or big for nothing?", *Library Hi Tech News*, 2021

Publication

1 %

15	www.doria.fi Internet Source	1 %
16	uns.ac.id Internet Source	<1 %
17	Chunlei Ye. "Research on the key technology of big data service in university library", 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2017 Publication	<1 %
18	crl.acrl.org Internet Source	<1 %
19	ir.inflibnet.ac.in Internet Source	<1 %
20	ssl.du.ac.bd Internet Source	<1 %

Exclude quotes Off

Exclude matches < 10 words

Exclude bibliography On