BIG DATA RESEARCH PUBLICATIONS IN LIBRARY AND INFORMATION SCIENCE: A SCIENTOMETRIC ANALYSIS

MD SAFIQUR RAHAMAN
Research Scholar,
Department of Library and Information Science,
DSB Campus, Kumaun University,
Nainital-263001
Uttarakhand
E-mail: safirahaman2007@gmail.com

Dr. YOUGAL JOSHI
Co-ordinator and Convenor,
Department of Library and Information Science,
DSB Campus, Kumaun University,
Nainital-263001
Uttarakhand
E-mail: ycjoshi@yahoo.com

ABSTRACT

The present study evaluates the research productivity on big data in the field of library and information science between 2005 and 2022. It investigated the annual growth pattern, top countries, organizations, document types. Language, productive institutions, productive country, top most journals, prolific authors, authorship pattern, key words used, author key words, most cited papers and country collaboration. Scientometric methods were employed for investigation. Web of Science database was used for data retrieval. Total 1780 papers published on big data. Results showed that there was an upward trend in the growth and development of publications on big data, Journal articles are the dominant documents, East China Jiaotong University produced more publications, USA most productive country, International Journal of Information Management published more articles, Liu of Hong Kong Polytechnic University the most prolific author, two author papers more, multi authorship pattern predominates, author key words, most cited paper and new research areas emerged recently are identified. The USA and China (n=69) were found as the leading collaborative country. This is the first comprehensive Scientometric study that gives a holistic view of big data research in Library and Information Science to the best of our knowledge. The study is useful to LIS and Big Data researchers, faculty members and computer scientists.

Keywords: ***Big Data, Library and Information Science, Scientometrics, Biblio metrics, Research productivity, Web of Science, Citation Analysis, Authorship Pattern, Prolific Authors, Country Collaboration, Highly Cited Papers.***

1. Introduction

Big data is a collection of structured, semi-structured, and unstructured data that may be mined for Information and utilized in machine learning, predictive modeling, and advanced analytics activities. The volume of data is expected to grow dramatically in the coming year due to modern technology such as faster broadband connections, mobile technologies, more powerful computers, bigger memory, cloud computing and the Internet of Things, etc. For example, market research firm IDC predicts that the worldwide Big Data market is expected to increase from USD 138.9 billion in 2020 to USD 229.4 billion in 2025. Academicians, researchers, and others are increasingly using internet platforms like Google, LinkedIn, Twitter, and Facebook to research, buy books, and learn about new library services. Libraries are currently experiencing an unavoidable boom because of information digitization and in this era of big data, library and information science plays a critical role. Big data has inherent problems in terms of getting Information out and keeping librarians and users up to date. Libraries need big data applications that provide predictive

analytics of user reading patterns and online services and apps that better comprehend the user's demands and expectations. As a result, big data provides more accurate predictions for information-seeking behavior trends, utilization of the system, library resources, and future library planning. Librarians can be trained in the fundamentals of using big data technologies and software to improve information services, faceted classification, collection development, and the establishment of controlled index words and ontology.

The main aim of the present study is to examine big data research output in LIS, research growth, and trends using scientometric indices. Big data even though fascinated researchers in Library science yet no recent research studies have been conducted on big data in the library and information science context, especially from the viewpoints of Scientometrics and visualization. Therefore, we need to make a comprehensive overview in this research direction and find out some basic patterns of big data research in LIS. The study would support new policies for LIS researchers, academicians, and computer scientists.

2. Review of Literature

A total of 36,000 big data papers were downloaded from the web of science between 2012 and 2017 by Mohammadi and Karami (2022) and studied the modeling and word co-occurrence analysis to discover significant themes. Baskaran (2022) examined worldwide trends in big data publications using the Web of Science database between 2011 and 2020.

Using bibliometric method, Nobanee et al. (2021) highlighted big data literature in the banking sector between 2012 and 2020. Zhang et al. (2021) conducted a bibliometric study on big data analytics and machine learning research between 2006 and 2020 based on a smple of 2160 Scopus papers.

Bibliometric and co-citation analyses on Big data analytics in health between 2000 and 2016 has been conducted by Galetsi and Katsaliaki (2020) in which the data extracted from Web of Science. Collecting data from Web of Science, Ajibade and Mutula (2020) evaluated the research productivity of south Africa in Library science on big data between 1992 and 2019. Liu et al. (2020) evaluated big data research between 2013 and 2018 using Scopus database. A bibliometric analysis on the evolution of data science and big data research has been studied by Raban and Gordon (2020). Global big data research in the agriculture field has been visualized by Trivedi (2019). A total of 379 research papers were downloaded from the web of science database and analyzed with the help of Microsoft excel, Vosviewer, and biblioshiny.

Xu and Yu (2019) highlighted big data research between 2009 and 2018. This research systematically studied and analyzes the big data publications in the Science Citation Index (SCI) and Social Science Citation Index (SSCI) databases using bibliometric and visual analysis approaches. A bibliometric study on big data research trends was studied by Kalantari et al. (2017). Even though there are several studies, only selected few are reviewed here.

3. Objectives of the Study

The core objectives of this study are:

- To investigate Big Data publication growth and citation impact in the LIS field between 2005 and 2022
- To identify the most relevant sources and language of documents
- To assess authorship pattern and most prolific authors on BD in LIS
- To know the most productive organizations and countries

- To examine the top journals publishing papers on BD as well as the prolific authors
- To find out the authorship pattern and to visualize the author key words
- To evaluate research themes and to find out the highly cited research papers
- To know the country collaboration pattern in BD research

4. Methodology

This science mapping study employed the scientometric method to assess big data (BD) research productivity in library and information science (LIS). This method, also known as science mapping, represents the relationship between disciplines, domains, specialties, documents, and authors. The present study focused on scientometric indicators such as yearly growth of literature, productive country and organization, prolific authors, significant sources, author keywords, collaborative countries and most cited research papers, exploring research themes, etc.

4.1. Search Query

The following search query was framed in the advance search box of the Web of Science database to retrieve the bibliographic data.

- TS= ("Big data" OR "bigdata")
- Refined by: Research Areas: Information Science, Library Science.

4.2. Date of data extraction:

The search query was run on 23rd March 2022 at Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia where the first author is working. A total of 71,423 research papers were found by using the mentioned search query.

4.3. Inclusion criteria

Inclusion criteria have been applied in the initial search results of 71,423 documents. Sixty-nine thousand six hundred forty-three research papers were excluded from the initial search results by applying research areas (Information Science, Library Science). Finally, 1,780 research papers have been selected for final analysis.

4.4. Method of analysis

All the 1780 research papers have been downloaded from Web of Science in different file formats and analyzed with bibliometric analysis tools such as VOSviewer, (van Eck & Waltman, 2010), Biblioshiny (Aria and Cuccurullo, 2019), Bibexcel (Persson, 2016), and Microsoft Excel.

5. Results and Discussion

5.1. Main Information about Research Data

The present study has finalized 1780 research papers on BD in LIS, which appeared in 232 sources between 2005 and 2022. Seventy-two thousand three hundred one references and 5292 author keywords were used to produce 1780 documents. The average number of citations per document was 16.75. Total authors are 3897. A single author produced 414 papers, while multi- authors contributed 1366 papers. The research paper per author was 0.457, while the collaboration index on BD in LIS was 2.58, implying reasonable collaboration among the researcher of BD in LIS. According to the research objectives, the result and analysis are given in the below sections.

Table 1

Main Information About Data

Description	Results
Timespan	2005 - 2022
Sources (Journals, Books, etc)	232
Documents	1780
Average years from publication	3.79
Average citations per document	16.95
Average citations per year per document	3.409
References	72301
Document Contents	
Keywords Plus (ID)	2134
Author's Keywords (DE)	5292
Authorship Details	
Total Authors	3897
Author Appearances	5057
Authors of single-authored documents	377
Authors of multi-authored documents	3520
Author Collaboration	
Single-authored documents	414
Documents per author	0.457
Authors per document	2.19
Co-Authors per document	2.84
Collaboration Index	2.58

5.2. Publication Growth and Citation Trends on Big Data

It is seen from the results shown in fig.1 that the first research paper on BD in LIS was reported in 2005. Unfortunately, there were no publication reported between 2006 and 2010. After 2011, research reported every year, and the year 2020 emerged as the leading year with 276 publications and 2926 citations. This was followed by 2019 with 271 publications and 4094 citations, 2021 with 260 publications and 1142 citations, and 2018 with 230 publications and 3877 citations. As far as the most cited year, 2015 received the highest number of citations for 158 publications, followed by 2016 with 4545 citations and 2019 with 4094 citations. The year 2012 was identified as the leading year in terms of average citation per publication (TC/TP=125.72), followed by 2016 (TC/TP=34.70)

and 2015 (TC/TP=31.54). The year 2022 produced 140 publications with 55 citations, an average citation per publication was 0.39. Complete documents in the year 2022 not available and in the near future it is expected to receive more citations and publications, as the real impact of citation is effected only after two to three years. This analysis reveals an increasing trend of BD publications in LIS during recent years, while the impact of citation was higher before 2019 compared to recent years.

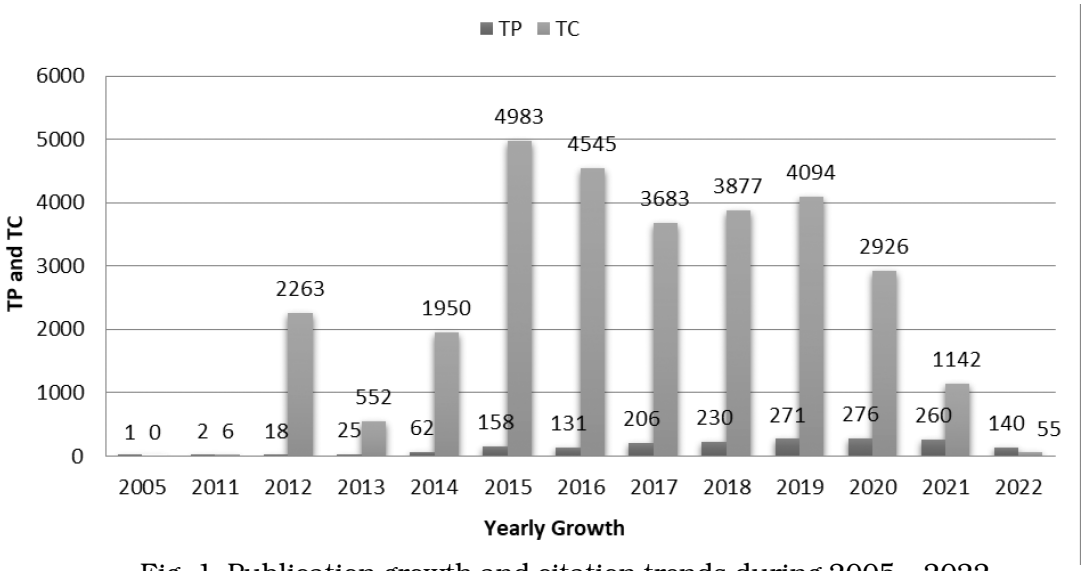


Fig. 1. Publication growth and citation trends during 2005 – 2022

5.3. Type of Documents on BD

It was found that there were eight forms of research produced during the study period. Journal articles are the major form of research on BD in LIS with 1212 publications and 24369 citations. This is followed by the conference papers with 358 publications and 698 citations. There are 86 reviews with 2906 citations, and editorial materials with 82 publications and 2169 citations. News items (n=3), letters, and corrections (n=2 each) were the least type of research in the fields. As far as the citation impact on the form of research, editorials received the highest number of average citations per publication (TC/TP=26.45), followed by a review paper with 33.79 TC/TP and article with 20.11 TC/TP.

Table 2

Type of Documents on Big Data in LIS

Rank	Type of Documents	TP	TC	TC/TP
1	Journal article	1212	24369	20.11
2	Conference papers	358	698	1.95
3	Reviews	86	2906	33.79
4	Editorials	82	2169	26.45
5	Book review	35	18	0.51
6	News item	3	0	0.00
7	Letter	2	12	6.00
8	Correction	2	2	1.00
*TP = Total Publication, ** TC = Total Citation, ***TC/TP = Average citation per publication				

5.4. Language-wise Productivity of Big Data

The analysis (table 3) revealed that English is the dominant language with 1696 publications with 29902 citations. Next in order is Spanish with 38 publications and 240 citations. Rest of the documents are in eight languages and are in less number. The average citation per publication, English, and Spanish, are again on the top list with 17.63 and 6.32, respectively.

Table 3

Language-wise Distribution of Documents on Big Data

Rank	Language	TP	TC	TC/TP
1	English	1696	29902	17.63
2	Spanish	38	240	6.32
3	Portuguese	18	14	0.78
4	German	10	3	0.30
5	Hungarian	6	4	0.67
6	Russian	5	5	1.00
7	French	3	4	1.33
8	Croatian	2	1	0.50
9	Turkish	1	1	1.00
10	Chinese	1	0	0.00

*TP = Total Publication, ** TC = Total Citation,
***TC/TP = Average citation per publication

5.5. Productive Organizations

The analysis on the most productive organizations (table 4) showed that East China Jiaotong University (Jianxi, China) produced more with 48 publications and 230 citations. Unfortunately, this organization lacks average citation per publication (TC/TP = 4.79) compared with other top ten organizations. Wuhan University, again in China is next with 32 publications and 542 citations. Next in the list is Delft University of Technology (Netherlands) and Nanjing University of Science and Technology (China) with 19 publications each and 512, 218 citations, respectively. Other universities are given in table 4. Further it is revealed that the average citation per publication; Sun Yat-Sen University, Delft University of Technology, and City University of Hong Kong were the top three most impactful organizations with the highest TC/TP (42.69, 26.95, 26.13, respectively. The results also showed that half of the top ten productive organizations belong to China (n=5), USA (n=2), and Netherland, Australia, and Hong Kong one organization from each country.

Table 4

Top Ten Most Productive Organization

Rank	Organization	Country	TP	TC	TC/TP
1	East China Jiaotong University	China	48	230	4.79
2	Wuhan University	China	32	542	16.94
3	Delft University of Technology	Netherlands	19	512	26.95
4	Nanjing University of Science & Technology	China	19	218	11.47
5	University of Wollongong	Australia	18	504	28.00
6	University of Illinois	USA	16	312	19.50
7	City University of Hong Kong	Hong Kong	15	392	26.13
8	Chinese Academy of Sciences	China	13	103	7.92
9	Sun Yat-sen University	China	13	555	42.69
10	Arizona State University	USA	12	202	16.83

*TP = Total Publication, ** TC = Total Citation,
***TC/TP = Average citation per publication

5.6. Most Productive Countries

USA was identified as the most productive country in Big Data (n=478) and attracted the highest number of citations (n=12955). It also has the highest average of 27.10 TC/TP. A similar result was reported by Rahaman et al. while evaluating traditional knowledge research (Rahaman et al., 2021). China followed this with 404 publications and 4554 citations, England with 137 publications and 3853 citations. Then comes India with 131 publications and 2188 citations, and Spain with 91 and 805 citations. The analysis in table 5 also revealed the top four productive countries and the highest cited countries (USA, China, England, and India) with 12955, 4554, 3853, and 2188 citations. Canadian publications identified as the best average citation per publication with 38.71 TC/TP, followed by England with 28.12 TC/TP and The USA with 27.10 TC/TP. South Korea was the least productive country among the top ten listed countries producing BD research in LIS (n=67) with 1227 citations. The publication of Spain recorded as the least cited with an average of 8.85 citations per publication in the top ten list.

Table 5

Top Leading Countries on BD Research in LIS

Rank	Country	TP	TC	TC/TP
1	USA	478	12955	27.10
2	Peoples R China	404	4554	11.27
3	England	137	3853	28.12
4	India	131	2188	16.70
5	Spain	91	805	8.85
6	Australia	88	1588	18.05
7	Germany	83	1625	19.58
8	Canada	70	2710	38.71
9	Netherlands	70	1314	18.77
10	South Korea	67	1227	18.31

*TP = Total Publication, ** TC = Total Citations,
***TC/TP = Average citations per publication

5.7. Top Ten Journals

International Journal of Information Management, published from UK, was the most relevant source, publishing 139 papers, as well as it got highest number of total citations (TC=7712). This is followed by the *Journal of The American Medical Informatics Association* with 55 publications (from USA), *International Journal of Geographical Information Science (UK) and Scientometrics*, (Netherlands) produced 49 publications each and *Journal of Enterprise Information Management* (UK) with 42 publications. *Information Processing & Management* was the least productive among the top ten listed sources, with 32 publications. The research paper published in *Information & Management* attracted 2nd highest number of citations (n=1778). Regarding the impact of citation per publication, two journals (*International Journal of Information Management* and *Information & Management*) were on the top list with 55.48 and 48.05, respectively. Among the top ten listed sources, six have more than 5.00 journal impact factors, and the *International Journal of Information Management* has the highest journal impact factor (14.09).

Table 6

Top Ten Most Relevant Journals on BD in LIS

Rank	Name of Journal	TP	TC	TC/TP	JIF	Country
1	*International Journal of Information Management*	139	7712	55.48	14.09	United Kingdom
2	*Journal of The American Medical Informatics Association*	55	1151	20.93	4.50	USA
3	*International Journal of Geographical Information Science*	49	1399	28.55	4.18	United Kingdom
4	*Scientometrics*	49	886	18.08	3.23	Netherlands
5	*Journal of Enterprise Information Management*	42	465	11.07	5.39	United Kingdom
6	*Government Information Quarterly*	38	1045	27.50	7.27	United Kingdom
7	*Information & Management*	37	1778	48.05	7.55	Netherlands
8	*Professional De La Informacion*	37	284	7.68	2.25	Span
9	*Journal of Knowledge Management*	33	872	26.42	8.18	United Kingdom
10	*Information Processing & Management*	32	818	25.56	6.22	United Kingdom

*TP = Total Publication, ** TC = Total Citation,
***TC/TP = Average citation per publication

5.8. Prolific Authors in Big Data

Researchers affiliated with Chinese institutions were the most prolific authors according to this study. Liu X., affiliated with The Hong Kong Polytechnic University, (Hong Kong, China) was the most prolific author in the field with 15 publications and 700 citations. This is followed by Zhang, Y. (Nanjing University of Information Science and Technology, China) with 13 publications and 237 citations. Next in order is Li, J. of Central South University (Hunan, China) with 11 publications and 203 citations. Janssen, M. (Delft University of Technology) and Liu, Y. (Peking University, China) with 10 publications each, and 307, 174 citations respectively. Wang S, Wang Y, and Xu X were the least productive authors, with eight publications each. Dwivedi YK, affiliated with Swansea University, UK, was the most cited author in the top ten list with 1025 citations, also has the highest average citation per publication (TC/TP=113.89).

Table 7

Top Ten Prolific Authors and Affiliation in Big Data

Rank	Author	Affiliation	Country	TP	TC	TC/TP
1	Liu, X	The Hong Kong Polytechnic University	China	15	700	46.67
2	Zhang, Y	Nanjing University of Information Science and Technology	China	13	237	18.23
3	Li, J	Central South University	China	11	203	18.45
4	Janssen, M	Delft University of Technology	Netherlands	10	307	30.70
5	Liu, Y	Peking University	China	10	174	17.40
6	Dwivedi, Y.K.	Swansea University	UK	9	1025	113.89
7	Zhang, J.	Chinese Academy of Science and Technology for Development	China	9	327	36.33
8	Wang, S.	University of Illinois	USA	8	62	7.75
9	Wang, Y.	Shandong Agriculture and Engineering University	China	8	337	42.13
10	Xu, X.	South China University of Technology	China	8	162	20.25

*TP = Total Publication, ** TC = Total Citation,
***TC/TP = Average citation per publication.

5.9. Authorship Pattern

The authorship of the publications ranged from single to 20 authors. On further analysis, it is found that two authorship has been the most preferred authorship pattern on BD research in Library and Information Science with 457 publications. This is followed by single-authorship publications (414 publications). Three authored documents are 410, followed by a collaborative effort of four-authorship and five-authorship trends with 281 and 111 publications respectively. The results depicted in fig.2 reveals that the maximum number of total citations were received by the four-authorship pattern with 8500 citations, followed by publications prepared by two-authorship with 7805 citations, and four-authorship with 4654 citations. Single-authorship publications were cited with an average of 9.35 citations per publication. Two-authorship, three-authorship, four-authorship, and five-authorship were cited with an average of 17.08, 20.73, 16.56, and 18.90 citations per publication. Fifteen-authorship received the highest average citation per publication with 117 citations for a single publication, Eight-authorship with 47.47 TC/TP, and Nine-authorship with 43.50 TC/TP.

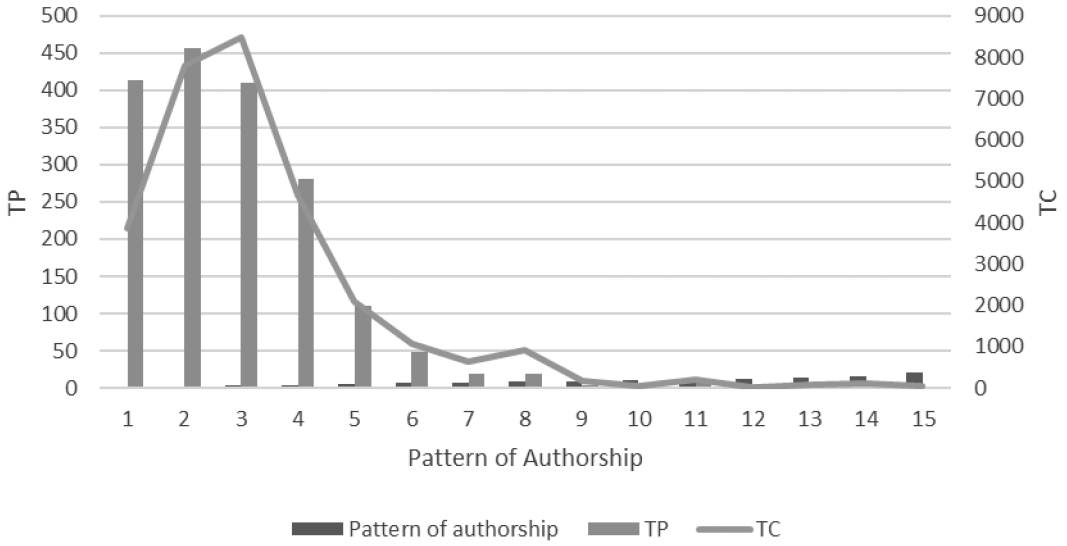


Fig. 2. Pattern of Authorship

5.10. Keyword analysis of Big Data Publications

To visualize the keywords of BD research in LIS, the co-occurrence selected from "types of analysis" and all keywords selected from a "unit of analysis, a full counting method has been employed to evaluate the keywords. A minimum of 25 occurrences of keywords are considered for analysis. Out of the total (6661) keywords, 68 meet the thresholds. For each of (the 68) keywords, the total strength of the co-occurrence links with the other keywords will be calculated. Thus keywords with the greatest total link strength will be selected. By applying the above criteria, the author found 68 keywords, 5 clusters, 1651 links, and 8756 total link strengths. As shown in the figure, all the 68 keywords were grouped in five clusters based on their likeliness and similar content. Various color of the map signifies various clusters. The map is made up of five clusters that reflect varied research themes.

Cluster # 1: Includes a total of 23 keywords and appeared 1357 times on BD research in LIS (Big Data Analytics, Business Analytics, Business Intelligence, Business Value, Capabilities, Competitive Advantage, Data Analytics, Decision-Making, Dynamic Capabilities, Firm Performance, Impact, Information-Technology, Innovation, Intelligence, Knowledge Management, Performance, Perspective, Predictive Analytics, Resource-Based View, Supply Chain, Supply Chain Management, Systems, and Value Creation).

Cluster # 2: Comprises 18 keywords, and all of these are seen 1906 times (Big Data, Classification, Data Mining, Deep Learning, Information, Internet, Machine Learning, Media, Model, Networks, Online, Quality, Sentiment Analysis, Social Media, System, Text Mining, Trust, and Twitter)

Cluster # 3: Consist of 14 keywords and give the impression of 671 times on BD research in LIS, 671 (Adoption, Artificial Intelligence, Artificial-Intelligence, Big-Data, Challenges, E-Government, Ethics, Framework, Governance, Government, Open Data, Policy, Privacy, and Strategy)

Cluster # 4: Includes a total of 8 keywords and this appeared 504 times (Analytics, Business, Data Quality, Data Science,

Information-Systems, Knowledge, Science, and Technology)

Cluster # 5: Comprises of 5 keywords and it appeared a total of 302 times 1780 publications (Cloud Computing, Design, Future, Internet of Things and Management)

The size of the nod denotes the occurrence of the keywords in the BD research in the field of library and information science. The keywords Big Data (n=1003) occurred highest number times. This is followed by Management (n=162), Big Data Analytics (n=144), Social Media (n=130), Analytics (n=125), Impact (n=121), Information (n=112), Challenges (n=100), Systems (n=100) and Innovation (n=97).

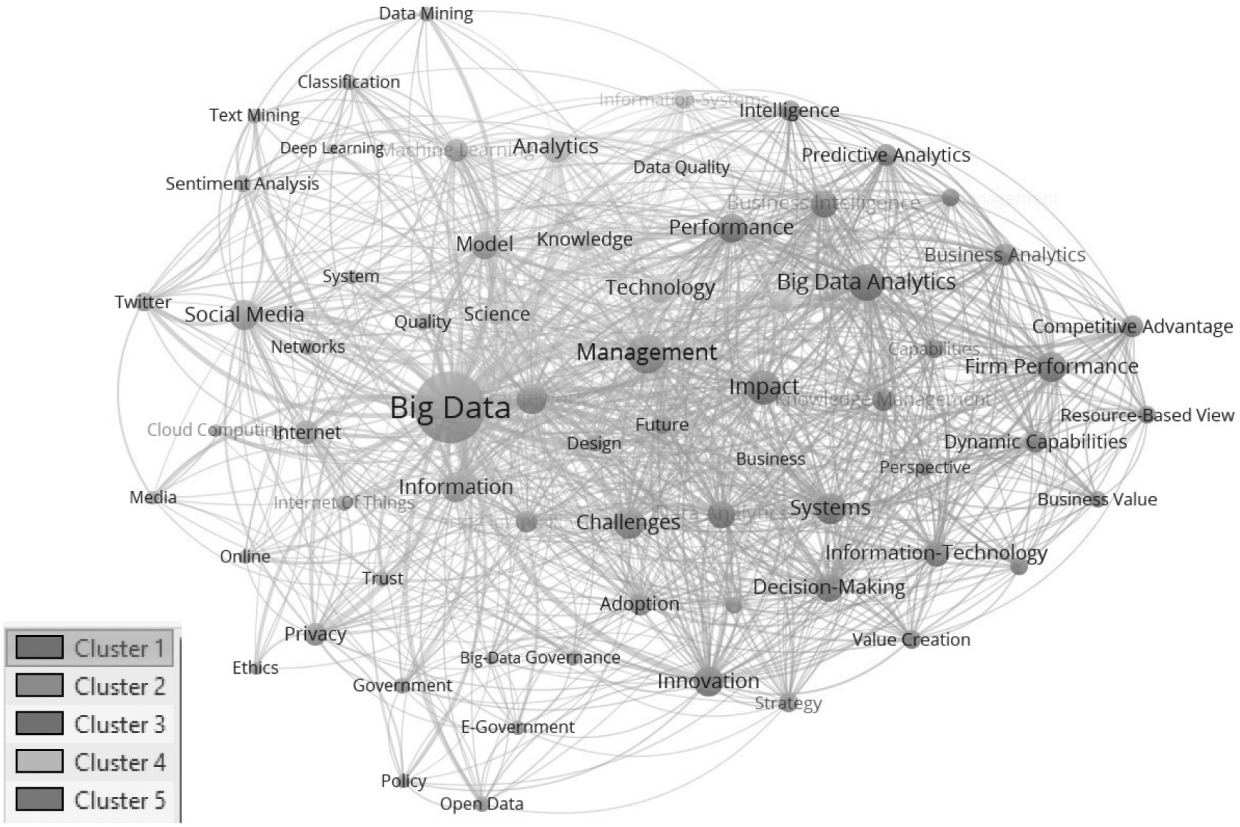


Fig. 3. Visualizing keywords by VOS viewer software
(A minimum of 25 occurrences of keywords were selected)

5.11. Thematic Evaluation of Author Keywords

We have grouped the whole study periods in three different times slice to notice the usage and appearance of different author keywords in different periods. Electronic health records, Business intelligence, Data mining, Big data, Research, Data Analysis, Social media, Bibliometrics, and Data were the most popular author keywords of LIS researchers in BD publications between 2005 and 2017. New author keywords of Open Data, Machine learning, Big data analytics, and Data science emerged during 2018-2020, suggesting the shift in the research themes. Artificial Intelligence, Digital technologies, and classification are three new themes that have emerged in the last two years (2021-2022). Big data, Data, Bibliometrics, and Social media

were the most famous author keywords from 2005-to 2022. Big data analytics, Open data, Data Science, Classification, Artificial intelligence, and Digital technologies emerged as the new research themes between 2018- and 2022.

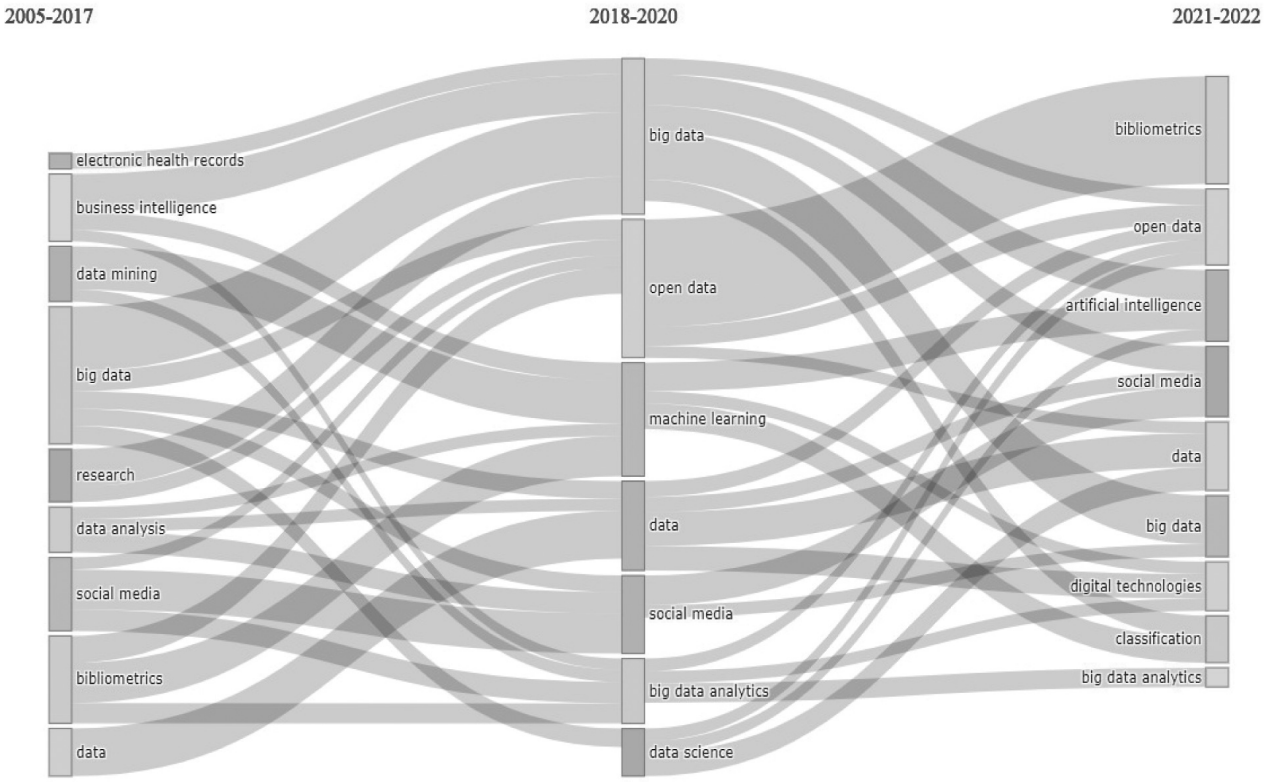


Fig. 4. The matic Evaluation of Author Keywords

5.12. Most Cited Research Papers on Big Data

The article entitled "Business Intelligence and Analytics: From Big Data to Big Impact" by Chen et al. (2012), published in *MIS Quarterly*, was the most cited article with 2164 total citations. This is followed by "Beyond the hype: Big data concepts, methods, and analytics" by (Gandomi & Haider, (2015) with 1511 citations published in *International Journal of Information Management*. "Big other: Surveillance Capitalism and the Prospects of an Information Civilization" by Zuboff, (2015) with 761 citations, "The role of big data in smart city" by Hashem et al., (2016) with 402 citations, and "A systematic literature review of blockchain-based applications: Current status, classification, and open issues" by Casino F (2019) with 385 citations.

The article "Data quality management, data usage experience and acquisition intention of big data analytics" by Kwon et al. (2014) was the least cited among the top ten most-cited research papers. The article "Business Intelligence and Analytics: From Big Data to Big Impact" and "Beyond the hype: Big data concepts, methods, and analytics" recorded the highest total citation per year with 196.73 and 188.88, respectively. It is also noticeable that most of the top ten most-cited research papers appeared in the *International Journal of Information Management* (n=4); all other journals contributed single papers in the top ten list.

Table 8

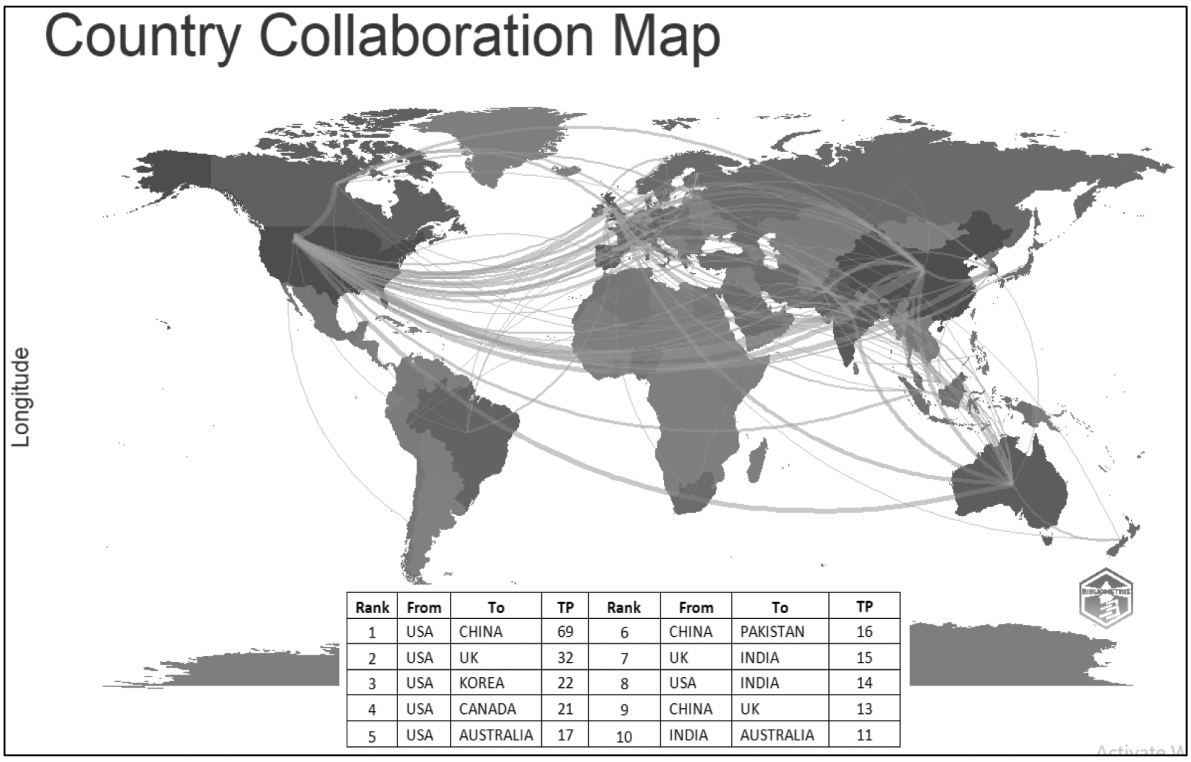
Top Ten Most Cited Research Papers

Rank	Title	Author	Year	Journal	TC	TC/Y
1	Business Intelligence and Analytics: From Big Data to Big Impact (Chen et al., 2012)	Chen, H.	2012	*MIS Quarterly*	2164	196.73
2	Beyond the hype: Big data concepts, methods, and analytics(Gandomi & Haider, 2015)	Gandomi, A	2015	*International Journal of Information Management*	1511	188.88
3	Big other: Surveillance Capitalism and the Prospects of an Information Civilization (Zuboff, 2015)	Zuboff, S.	2015	*Journal of Information Technology*	761	95.13
4	The role of big data in smart city (Hashem et al., 2016)	Hashem, IAT	2016	*International Journal of Information Management*	402	57.43
5	A systematic literature review of blockchain-based applicati-ons: Current status, classifica-tion and open issues (Casino et al., 2019)	Casino, F.	2019	*Telematics and Informatics*	385	96.25
6	Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda (Duan et al., 2019)	Duan, Y.	2019	*International Journal of Information Management*	314	78.50
7	Toward the development of a big data analytics capability (Gupta & George, 2016)	Gupta, M.	2016	*Information & Management*	312	44.57
8	Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda(Loebbecke & Picot, 2015)	Loebbecke, C	2015	*The Journal of Strategic Information Systems*	291	36.38
9	Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research (Agarwal & Dhar, 2014)	Agarwal, R.	2014	*Information system research*	283	31.44
10	Data quality management, data usage experience and acquisition intention of big data analytics (Kwon et al., 2014)	Kwon, O	2014	*International Journal of Information Management*	278	30.89

*TC/Y = Total Citation per Year

5.13. Country Collaboration

The USA and China (n=69) was the leading collaborative country in BD research by LIS researcher. This is followed by The USA and United Kingdom with 32 publications, The USA and Korea with 22 publications, The USA and Canada with 21 publications, and the USA and Australia with 17 publications. India and Austria were the least collaborative countries in the top ten most collaborative countries with 11 publications. Interestingly, the USA collaborated a maximum (n=6) times with other countries among the top ten list, followed by India with three counties, the United Kingdom, China, and Australia, each with two countries. A similar type of highest research collaboration (the USA and China) was reported by Ansari et al. (2021) in evaluating anthropometric measurement research.



Rank	From	To	TP	Rank	From	To	TP
1	USA	CHINA	69	6	CHINA	PAKISTAN	16
2	USA	UK	32	7	UK	INDIA	15
3	USA	KOREA	22	8	USA	INDIA	14
4	USA	CANADA	21	9	CHINA	UK	13
5	USA	AUSTRALIA	17	10	INDIA	AUSTRALIA	11

Fig. 5. Top Ten Collaborative Countries

6. Findings of the Study

The analysis revealed some interesting findings from this study, which are mentioned below:

- The year (2020) was identified as the most productive year with 276 publications and 2926 citations on BD research in the field of LIS.
- The year (2015) contributed the highest number of citations (TC = 4983).
- The year (2012) managed the highest average citation per publication (TC/TP = 125.72).
- The majority of LIS researchers published their big data publications as journal articles (TP=1212).
- Editorial materials received the highest average number of citations per publication (TC/TP=26.45).

- Most of the researchers preferred to write their research work in English language (TP=1696).
- East China Jiaotong University was the leading producer of BD research in LIS (TP=48).
- Sun Yat-sen University was the most cited organization (TC = 555) as well as having the highest average citation per publication (TC/TP = 42.69).
- The USA was the most productive country (TP = 478) and also attracted the highest number of citations (n = 12955).
- Canadian publications received the best average citation per publication (TC/TP=38.71).
- *The International Journal of Information Management*, published in the UK, was the most relevant source with 139 papers, as well as the highest number of total citations (TC = 7712) and the highest average citation per publication (TC/TP = 55.48).
- Liu X was the most prolific author of BD research in the LIS field, with 15 publications and 700 citations.
- Dwivedi YK was the most cited author (TC =1025) and also has the highest average citation per publication (TC/TP =113.89).
- Two-authorship has been the most preferred authorship pattern in BD research in library and information science with 457 publications.
- Multiple authorship predominates in BD research.
- The most important keywords in BD research publications in LIS are Big Data (n = 1003), Management (n = 162), Big Data Analytics (n = 144), Social Media (n = 130), Analytics (n = 125), Impact (n = 121), Information (n = 112), Challenges (n = 100), Systems (n = 100), and Innovation (n = 97).
- Big data, data, bibliometrics, and social media were the most popular author keywords between 2005 and 2022.
- Big data analytics, open data, data science, classification, artificial intelligence, and digital technologies emerged as the new research themes between 2018-and 2022.
- The journal article titled 'Business intelligence and analytics', by Chen et al. (2012) published in MIS Quarterly is the most cited article. It has the highest total citations.
- The USA and China (n = 69) were the leading collaborative countries in BD research by LIS researchers.

7. Conclusion

The present study analyzed 18 years (2005-2022) of research articles related to big data (BD) in the field of Library and Information science (LIS). Web of science data have been utilized to retrieve the data and analyzed using the scientometric methodology. Even though the data was confined to the Web of Science database, the research showed several interesting findings, including a growing trend in publications and citations on BD in the field of LIS. The finding shows that most LIS researcher published their BD research in article form, and their preferred language was English. East China Jiaotong University, Wuhan University, Delft University of Technology, Nanjing University of Science and Technology, and the University of Wollongong were top-producing institutions, with East China Jiaotong University at the top. USA, China, England, India, and Spain were leading productive countries, and The USA

ranked first. *International Journal of Information Management* emerged as the most relevant journal in the field.. Liu X, Zhang Y, Li J, Janssen M, and Liu Y were the most prolific author in the field. Two-authorship is preferred and a trend of multiple authorship is seen. Big Data, Management, Big Data Analytics, and Social Media were LIS researchers' most preferred author keywords. The USA and China were the leading collaborative countries in LIS researchers' BD research. Big data, data, Bibliometrics, and Social media were the most famous author keywords throughout the study period. Big data analytics, open data, Data Science, Classification, artificial intelligence, and digital technologies emerged as the recent research themes. Furthermore, this research might serve as a reference for academics, information professionals, and computer scientists aiming to fill gaps and exploit potential in future research on the subject.

References

1. **Agarwal, R. and Dhar, V.** (2014). Editorial —Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. *Information Systems Research*, *25*, 3, 443–448. https://doi.org/10.1287/isre.2014.0546

2. **Aria, Massimo and Cuccurullo, Corrado** (2019). *bibliometrix 3.0*. Https://Www.Bibliometrix.Org/About.Html. https://www.bibliometrix. org/index.html

3. **Ajibade, P. and Mutula, S. M.** (2020). Big Data Research Outputs in the Library and Information Science: South African's Contribution using Bibliometric Study of Knowledge Production. *African Journal of Library, Archives and Information Science*, *30*, 1.

4. **Ansari, K. M. N. et al.** (2021). Assessment of literature growth in Anthropometric measurement research/ : A bibliometric analyses of Scopus indexed publications. *Library Philosophy and Practice*, 1–27. https://digitalcommons.unl.edu/libphilprac/5901/

5. **Baskaran, S.** (2022). Publications trends in big data: A scientometric analysis. *IP Indian Journal of Library Science and Information Technology*, *6*, 2, 102–107. https://doi.org/10.18231/j.ijlsit.2021.021

6. **Chen, Chiang and Storey.** (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, *36*, 4, 1165. https://doi.org/10.2307/4170350

7. **Eck, N. J. van and Waltman, L.** (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*, 2, 523–538. https://doi.org/10.1007/s11192-009-0146-3

8. **Galetsi, P. and Katsaliaki, K.** (2020). Big data analytics in health: an overview and bibliometric study of research activity. *Health Information & Libraries Journal*, *37*, 1, 5–25. https://doi.org/https://doi.org/10.1111/hir.12286

9. **Gandomi, A. and Haider, M.** (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*, 2, 137–144. https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2014.10.007

10. **Kalantari, A. et al.** (2017). A bibliometric approach to tracking big data research trends. *Journal of Big Data*, 4, 30.

11. **Liu, X. et al.** (2020). The research landscape of big data: a bibliometric analysis. *Library Hi Tech*, *38*, 2, 367–384. https://doi.org/10.1108/LHT-01-2019-0024

12. **Loebbecke, C. and Picot, A.** (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *The Journal of Strategic Information Systems*, *24*, 3, 149–157. https://doi.org/https://doi.org/10.1016/j.jsis.2015.08.002

13. **Mohammadi, E. and Karami, A.** (2022). Exploring research trends in big data across disciplines: A text mining analysis. *Journal of Information Science*, *48*, 1, 44–56. https://doi.org/10.1177/0165551520932855

14. **Nobanee, H. et al.** (2021). Big Data Applications the Banking Sector: A Bibliometric Analysis Approach. *SAGE Open*, *11*, 4. https://doi.org/10.1177/21582440211067234

15. **Persson, O.** (2016). Bibexcel is a versatile bibliometric toolbox developed. *Inflibnet*, *23*, 4.

16. **Raban, Daphne R. and Gordon, Avishag (**2020). The evolution of data science and big data research: A bibliometric analysis. *Scientometrics*, 122, 1563-1581.

17. **Rahaman et al.** (2021). A scientometric assessment of global research productivity in traditional knowledge: Evidence from scopus database. *KELPRO Bulletin*, *25*, June, 15–29.

18. **Trivedi, G.** (2019). Visualization and Scientometric Mapping of Global Agriculture Big Data Research. *Library Philosophy and Practice*.

19. **Xu, Z. and Yu, D.** (2019). A Bibliometrics analysis on big data research (2009–2018). *Journal of Data, Information and Management*, *1*, 1, 3–15. https://doi.org/10.1007/s42488-019-00001-2

20. **Zhang, J. Z. et al.** (2021). Big data analytics and machine learning: A retrospective overview and bibliometric analysis. *Expert Systems with Applications*, *184*, 115561. https://doi.org/https://doi.org/10.1016/j.eswa.2021.115561