

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

October 2021

Predicting LIS Scholarly Research Directions In The Era Of Data Science

Nazia Ismail Chohdary MPhil

MPhil Scholar, Department of Library and Information Science, The Islamia University of Bahawalpur,
nazia.ismael@gmail.com

Mahe Bushra Asghar PhD

Assistant Professor, Department of Library and Information Science, The Islamia University of Bahawalpur,
mahebushraasghar@gmail.com

Muhammad Ahmad Al Shaheer

ahmadshaheerc12@gmail.com

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>



Part of the [Library and Information Science Commons](#)

Chohdary, Nazia Ismail MPhil; Asghar, Mahe Bushra PhD; and Al Shaheer, Muhammad Ahmad, "Predicting LIS Scholarly Research Directions In The Era Of Data Science" (2021). *Library Philosophy and Practice (e-journal)*. 6328.

<https://digitalcommons.unl.edu/libphilprac/6328>

Predicting LIS scholarly research directions in the era of Data Science

¹Nazia Ismail

MPhil Scholar, Department of Library and Information Science, The Islamia University of
Bahawalpur nazia.ismael@gmail.com

²Mahe Bushra Asghar

Assistant Professor, Department of Library and Information Science, The Islamia University of
Bahawalpur mahebushraasghar@gmail.com

³Engr. Muhammad Ahmad Al Shaheer

PhD Scholar, Department of Electrical Engineering, Military College of Signals National
University of Sciences and Technology Islamabad ahmadshaheerc12@gmail.com

Abstract

Purpose –The purpose of this paper is to present the quantitative analysis of scholarly literature published worldwide on Data Science in the field of Library and Information Science (LIS).

Design/methodology/approach –The Web of Science (WoS) Core Collection database was used as the main source for extracting data from the published documents during the period 2010-2019. The analysis of the published documents was based on the following important trends: Number of Data Science publications in all fields, Number of Data Science publications in the LIS field, the most cited document type on Data Science in LIS perspective, the most prolific author, the most productive publication year, the most productive country, annual citations, highly cited journal and research article on Data Science in LIS perspective, highly contributing research organizations and group authors on Data Science in LIS field and highly contributing universities on Data Science in LIS field during the period 2010-2019. The types of documents considered in this study were research articles, conference proceeding papers, book chapters, book reviews and editorial material.

Findings –The findings of this study showed that the USA possessed the highest number of publications on Data Science in the field of LIS. The total number of publications on Data Science including all fields was 106,028 and the number of publications on Data Science in LIS perspective was 3,799 during the period 2010-2019. The year 2019 was identified as the most productive year as the research output in this year was maximum. IEEE was identified as the most productive research organization/group author, Wuhan University China as the most productive university and the *Journal of Scientometrics* was declared as the most prolific journal on Data Science in LIS perspective during the period 2010-2019. Bornmann Lutz, from Germany, was identified as the most prolific author with a total of 69 publications and the article titled “*Business Intelligence and Analytics: From Big Data to Big Impact*” was the highly cited article on Data Science in LIS perspective during the period 2010-2019.

Originality/value –The study will be of great importance for students, scholars, educationists, professionals and researchers who intend to perform Scientometrics research studies in the LIS field.

Keywords: Web of Science, Data Science, Big Data, Library and Information Science, Scientometrics Analysis, Bibliometrics Analysis, Data Analysis, ISI WoS, Data scientists and Altmetrics.

Paper type: Research paper

1. Introduction

Science is defined as a process of giving knowledge through a systematic study. A large amount of data is available nowadays due to several gadgets, sensor devices and the use of social media (i.e. “the Internet of Things”) and thus data analysis is a basic requirement. The branch of science which deals with the analysis of data and the extraction of useful information from raw data is called Data Science. The role of Data Science is very important in Library and Information Science field because of “Big Data” (huge amount of data) (Virkus, Mandre & Pals, 2018).

The field of LIS has faced a lot of challenges recently due to emerging technologies, globalization, demographic trends, internationalization etc. It has always been a challenge to adopt innovative ways of collecting data and achieve the required data analysis goals to survive and extract maximum output from the collected data, for this reason, research in Data Science is necessary (Virkus, 2015). Data Science is an emerging field that provides a lot of opportunities for LIS professionals, educationists and researchers to fulfil new social requirements that include e-science and research data management (Garounfallou et al., 2008; Antell et al., 2014).

Several studies have evaluated the scientific output of various subjects in different parts of the world by using bibliometrics. This study uses a quantitative method of research to analyze the published literature in the field of LIS. This study is conducted by using quantitative methods to investigate the Data Science publications patterns in the field of LIS. The documents were analyzed based on the following indicators: Number of Data Science publications in all fields including LIS, Number of Data Science publications only in LIS field, the most cited document type on Data Science in LIS perspective, the most prolific author, the most productive publication year, the most productive country, highly cited journal and research article on Data Science in LIS perspective, highly contributing research organizations and group authors on Data Science in LIS field and highly contributing universities on Data Science in LIS field during the period 2010-2019.

This paper enlists the history and background of Data Science. It shows a strong connection of Data Science with the LIS field. It elaborates the importance and significance of Data Science strategies and processes in the LIS field. The paper analysis various Data Science trends in LIS perspective by analyzing the published literature on Data Science indexed in WoS database Core Collection during the period 2010-2019. The main objective of this study is to emphasize the importance of different Data Sciences skills, practices and processes that how they can improve the quality of research, the productivity of results and the better implementation of these results for the benefits of LIS society. Based on the results future directions are given to the LIS researchers, professionals and data scientists to improve their Data Science skills to enhance planning and decision-making strategies. The selected time frame of ten years was considered sufficient to identify current LIS trends. The results of this study revealed a wide range of data related to selected aspects of LIS research.

2. Literature Review

The published literature shows that LIS professionals are involved in a discipline in which many research methods and approaches are applied in literature analysis. It is mostly due to the outcome of the previous contribution and current efforts made by the LIS scholars (Chu, 2015). The progress in LIS has increased rapidly in the last decade due to the integration of the LIS field with new technologies. The number of LIS schools and libraries is also increasing with time. They are improved with the inventions of newer technologies (Abdullahi et al., 2007). The

contribution of LIS professionals in research productivity is very important to the development of the LIS profession (Maa & Lee, 2017; Uzun & Ozel, 1996). The review of the literature shows that many researchers have conducted Bibliometric studies to analyze LIS research publications.

In recent years, the literature on Big Data and Data Science is also increasing. The field of science responsible for the management of Big Data is called Data Science (Manyika, 2011). Data Science establishes the relationship and developments of human behaviour with nature as a part of natural science and social science and is the strategic decision making using data (Zhang et al. 2017). The requirement for more powerful networking, algorithms, potential data analysis techniques and highly skilled professional for analysis of data is rapidly increasing (Provost & Fawcett, 2013, p.51). Over the years, Data Science has gained tremendous importance due to its potential and strategic need (Manyika, 2011; Provost & Fawcett, 2013). Recently Data Science has been used in many fields such as public policy, health care, industry, marketing business, management, agriculture, economics, physics, education, public transportation, Library and Information Science etc. (Voulgaris, 2014).

In 2016 an annual report by Jameskurose and keithmarzulo co-chairs of the subcommittee on Networking and Information Technology Research and Development (NTRD) claimed that the development of mathematics and computer science combined with a huge amount of data has led to a new ecosystem called Data Science. Data Science has many definitions but it can be defined as: “generating insights from data to inform decision making”. Federal Big Data Research and the Development Plan recently released by Big Data and Research Initiative, clearly presents curators, librarians, and data scientists as core competitors to help need the growing demands for data analysis.

Chu, (2015) described that the literature which has been published so far on Data Science in the field of LIS shows that LIS professionals are interested in the discipline of Data Science. He has to propose the different methods and approaches that can be applied to investigate the published literature. He shows that a lot of contribution has been made by LIS scholars in the field of Data Science the literature revealed that several Bibliometrics studies have been carried out by many researchers so far in the field of LIS.

Han et al., (2014) performed the Scientometrics analysis to evaluate several trends, network structure and the major groups of collaborative authors in the literature on LIS at both national and international levels. He observed 8,570 research papers taken from the world's top 15 LIS journals indexed in the WoS Core Collection database from the year 2000-2011. The finding showed that out of the total LIS publications 66% were published with a mutual collaboration of various authors. In the collaborative pattern the two trends i.e. two country and two institution papers were the basic collaboration patterns at both the country and institutional level. The position of two countries USA and UK were concluded to be quite good and satiable.

Park, (2008) performed a Bibliometrics analysis considering the world's top LIS journals indexed in the WoS and he analyzed the pattern of authorship perspective concerning the Asia-Pacific Region. Results concluded that 1,317 articles were published during the year 1967-2005. The countries like Singapore, Japan, New Zealand, The Phillipian, Malaysia, Australia, China, South coria Taiwan and Thailand were proved the most prolific and productive countries in terms of LIS research. 73.1% of the articles were published by the collaborative various authors in LIS journals and 50% of the research papers were published in the top 20 journals. The collaborative efforts among different countries in Asia and Pacific Region were Australia & China, Australia & New Zealand, and China & Singapore respectively.

Ivanovic and Ho, (2016) carried out an analysis to analyze the various properties for the

highly cited publications in the discipline of Library and Information Science. They conducted a study using the specific category of “Journal Citation Reports (JCR)” and “Social Science Citation Index (SSCI)”. Their results showed that there were nearly 501 highly cited research articles that were published during the period 1956-2009, in 37 high impact LIS journals indexed in WoS. Out of the 26% highly cited articles most of them were published in the “Management Information Systems (MIS) Quarterly Journal”. Canada and the USA were considered to be most productive with 13 highly productive research institutions. Harvard University USA was proved as the most productive among the other most productive institution and a total of 67% of most cited articles were from the USA.

Rafiq, Jabeen and Yun (2015) conducted a multivariate analysis of the progress in the LIS field and various trends of publications globally. They selected 40 LIS journals identified in the “Journal Citation Report (JCR)” 2010 edition for review. Their results showed that 18,371 research papers were published in the LIS field during the period 2003-2012. A gradual growth rate of 11.37% was observed in 2009. The trends which increase significantly were the trend of self-citation at an average rate of 38.56%, while the trend of citation in the LIS field decreased. The “Journal Article” was concluded at the most popular publication types among the LIS scholars and researchers. The authors from the USA wrote greater than 43% of LIS literature. The USA was the most productive country in the number of publications, and it also participated in a collaborative effort with other countries which include Canada, Australia and the UK. The most productive institution was the Victoria University of New Zealand with a total of 317 (1.73%) research publications. The “Journal of Academic Librarianship” was the journal that possessed the highest number of citations (1,401), whereas the Asian countries i.e. India, Iran, Taiwan and China each produce only 1% of all LIS publications. In another research, Rafiq, Jabeen, Tahir and Yun, (2015) analyzed the outcome of research in the LIS field during the period 2003-2012 through the WoS in 40 LIS journals. Results indicated that most of the authors from 2003-2009 in the LIS field were single authors with 12,847 publications published i.e. 69.9% respectively. With time this trend decreased gradually and the number of collaborative publications increased during the period 2010-2012. They showed that the University of Illinois, USA was the leading institution in the publication of LIS and produced 95 articles (about 0.52%) during this time. It was concluded that the Asian countries Iran, India, Taiwan and China lag far behind in the publication of LIS literature.

Jabeen et al., (2016) conducted a Bibliometrics analysis to find out the collaboration pattern of the Asian countries. They considered articles in the WoS database published during the period 1994-2013. Their results showed that the Asian countries did not have good and effective cooperation. From the viewpoint of LIS, the cooperation of authors in a continent and within the continent was very low at both author and institutional levels. Their study showed that although the countries from Asia produced a large number of LIS papers the USA and UK were always above the Asian countries. The researcher emphasizes the reason for low research output and the limited cooperation is that in Asia there is a shortage of funds and few incentives are provided to the researchers as in the western countries. Researchers indicated that there is a need to develop policies, sign MOUs and participate in joint research programs, collective research plans and short spam research exchange programs.

Deses in 2015 performed a study to investigate the role of tenure track librarians at the University of Mississippi. For this purpose, he investigated the articles published from 2008-2013 in the LIS field. The survey indicated that the majority (59%) of the authors produced collaborative publications and these were the most cited publications.

In 2016 Malone and Burke performed an analysis to determine the opinions and understanding of academic librarians about Altmetrics and Bibliometrics. Results showed that the understanding of academic librarians about Bibliometrics was greater than Altmetrics. Their research concluded that knowledge about Data Science is important, not only for the use of Bibliometrics and its tools effectively but also for librarians, educationists, scholars and researchers.

Maharana and Das in 2014 analyzed the progress of LIS research publications in WoS, which were published by the Asian authors from 1999 through 2013. The results showed that the authors from India published 140 research articles in the LIS field. The annual range of publications by the Indian authors was 9-10 papers per annum and most of the publications were generally research articles.

The present study will contribute towards the knowledge about Data Science by presenting some facts-based results about the current trends of Data Science in the LIS field shown by the LIS scholars globally. The review of the literature shows that many researchers have conducted Bibliometrics and Scientometrics studies to measure the research productivity annually, research productivity of institution, authors' collaboration in research work, most reputed author and highly cited articles in various disciplines. This study not only provides an updated insight into the research productivity of the world on Data Science in the field of LIS, but it also will help the LIS scholars to stay informed of the most productive research institution, highly cited journals and most prolific authors on Data Science in the field of LIS. It will eventually facilitate the LIS researcher and scholar to make informed decisions related to Data Science in selecting the LIS journals for both reading and publishing articles and choosing the institutions or universities for education. Similarly, the list of highly cited authors on Data Science will help the researcher in identifying those whose research is highly valued on Data Science in the field of LIS.

3. Research Questions

- RQ1.* How many publications on the term Data Science in the LIS field were indexed in the WoS Core Collection database during the period 2010 to 2019?
- RQ2.* What were the main document types (i.e. Article, Proceeding Papers, Review, Book Chapters and Editorial Materials) on Data Science indexed in the WoS Core Collection database during the period 2010 to 2019?
- RQ3.* Which was the most cited publication on the term Data Science in the LIS field on annual basis among those publications that were indexed in the WoS Core Collection database during the period 2010 to 2019?
- RQ4.* Which was the most prolific and highly cited author/group authors and journal on the term Data Science in the LIS field indexed in WoS Core Collection database during the period 2010 to 2019?
- RQ5.* Which was the most productive country, university and research organization on the term Data Science in the LIS field indexed in WoS Core Collection during the period 2010 to 2019?

4. Research Method

The basic purpose of the study was to present a Scientometrics analysis of the published literature on Data Science in the LIS field during the period 2010-2019 worldwide. The survey method was used in the study, and the WoS Core Collection database was used for the collection of facts, a well-known and reliable database that is often used for Scientometrics analysis and scientific research (Herther, 2009). WoS is a comprehensive citation data source that can be used to measure research productivity (Tripathi et al., 2018). The WoS provides wide coverage of data about the author, journal, subject and country's contribution (Herther, 2009). In this study the WoS database Core Collection was selected because for the decades, this database has been used widely for measuring the quality of peer-reviewed literature internationally (Abrizah et al., 2013; Meho, 2007). Paul-Hus (2016) reported that the WoS provides a multidisciplinary, comprehensive, and authoritative coverage of more than 13,605 international research journals.

In this study, the researcher only focuses on the publications titled Data Science in the LIS field. The period selected for the analysis of published literature was from 2010-2019. The type of documents included in the literature was limited to research articles, conference proceeding papers, book chapters, book reviews and editorial material. After choosing these parameters in the WoS database, a report was formed. The total number of documents retrieved from the WoS database based on the search term Data Science was 106,028 publications including different fields like IT, Computer Science, Chemistry, Biology, Agriculture, and Medicine etc. After refining the searched results by selecting specific document types and the period of 2010-2019, the number was reduced to 3,799 publications on Data Science in the field of LIS during the period 2010-2019. The results were saved on the computer for a detailed analysis and to report them in the future. The researcher analyzed the results based on the following trends: Number of Data Science publications in all fields including LIS, Number of Data Science publications only in LIS field, the most cited document type on Data Science in LIS perspective, the most prolific author, the most productive publication year, the most productive country, annual citations, highly cited journal and research article on Data Science in LIS perspective, highly contributing research organizations and group authors on Data Science in LIS field and highly contributing universities on Data Science in LIS field during the period 2010-2019. In this regard, Garg and Kumar, (2019) researched the same pattern to know the global output of research in the LIS field.

5. Research Limitations

This study is limited to only those documents, which were indexed in WoS database Core Collection and published from 2010-2019. Five specific document types were selected i.e. research articles, conference proceeding papers, book chapters, book reviews and editorial material.

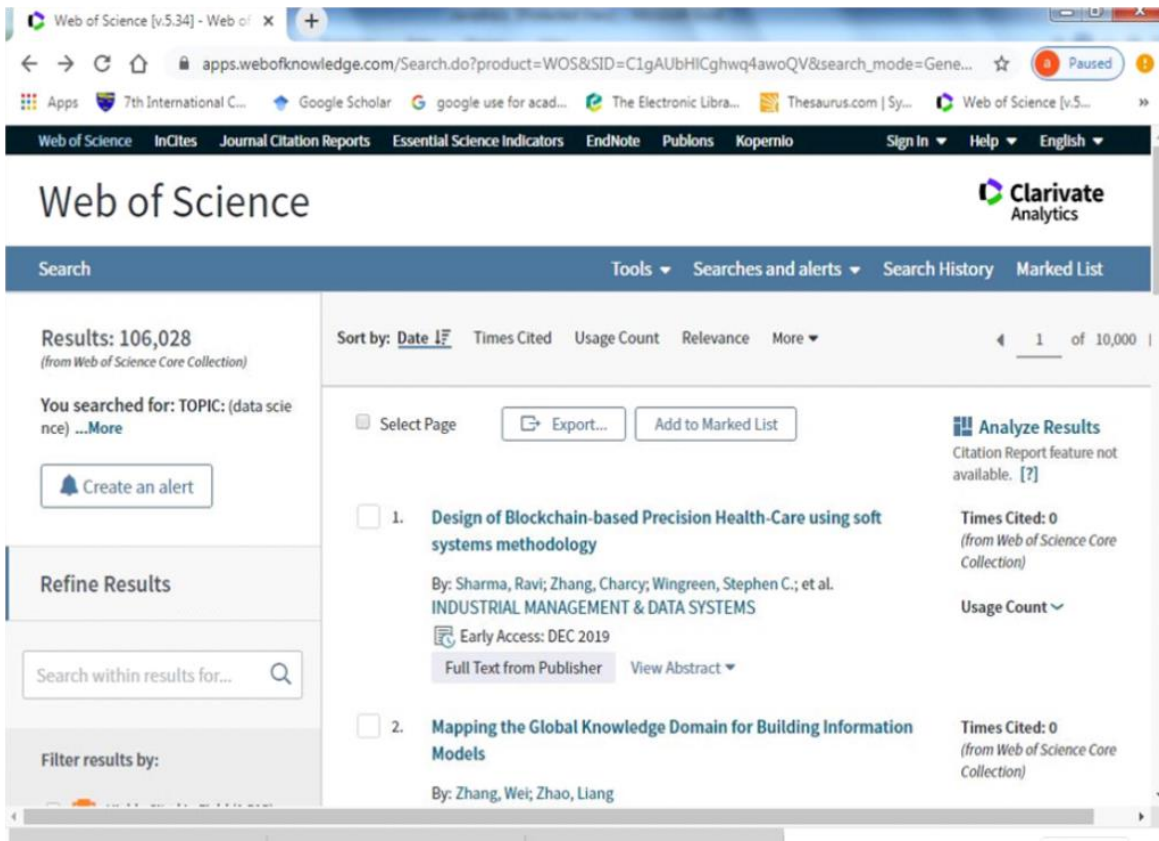


Figure 1 Results on Data Science in all fields including LIS during the Period 2010-2019

6. Results

This section presents the results of this study. The results can be divided into the following sections:

6.1 RQ1. How many publications were on Data Science in the LIS field in WoS database Core Collection during the period 2010 to 2019?

The study analyzed that how many LIS research articles were published and what was the number of citations of the research articles during the period 2010-2019. Figure 4 demonstrates that 2019 was the most productive year in terms of LIS publications. The results showed that there were 3,799 publications on Data Science in the field of LIS during the period 2010 to 2019 considering all document types as shown in figure 2 and when the search was limited using five specific document types i.e. article, proceeding paper, review, book chapters and editorial materials there were 3,766 publications as shown in figure 3.

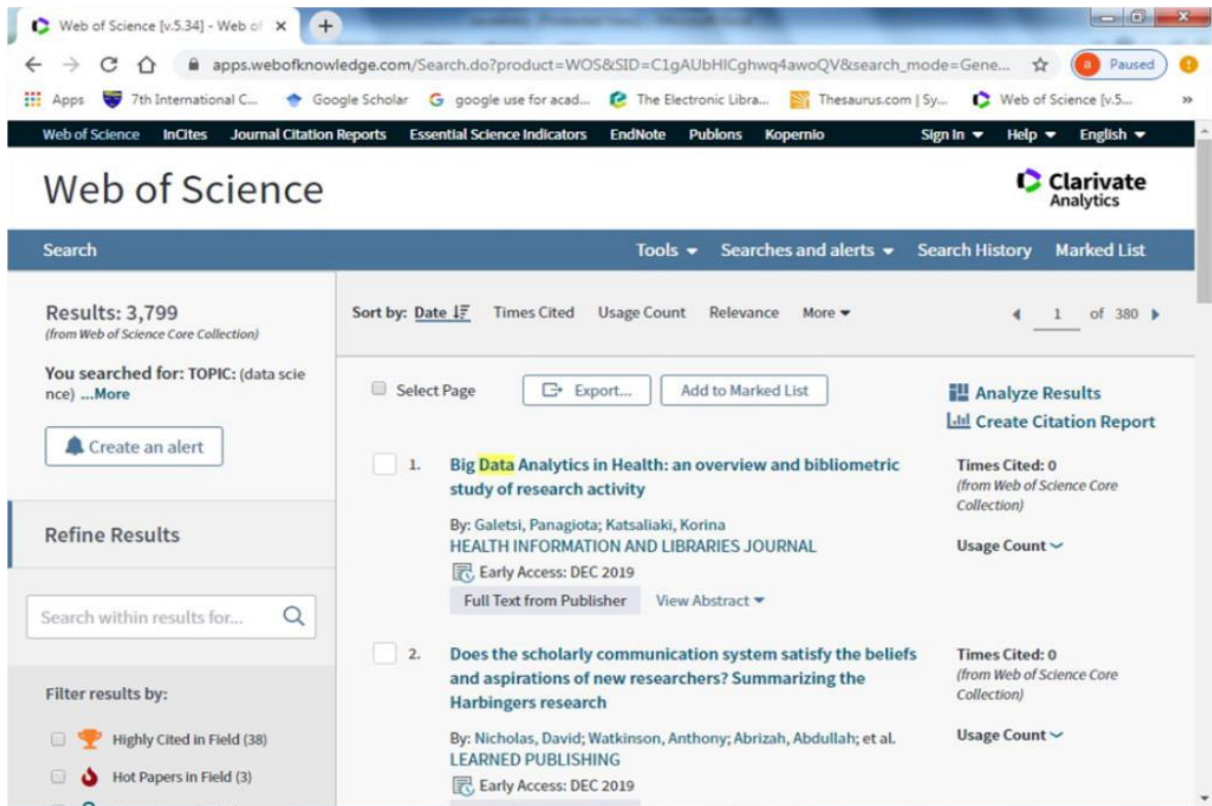


Figure 2 Results on Data Science in LIS Perspective during the Period 2010-2019

6.2 RQ2. What was the main document type (i.e. Article, Proceeding Papers, Review, Book Chapters and Editorial Materials) on Data Science in WoS database Core Collection during the period 2010 to 2019?

The findings of the study showed that there were 3799 publications indexed in the WoS Core Collection database during the period 2010 to 2019 for all document types including articles, proceeding papers, review articles, book chapters, editorial materials, early access and retracted publications. When the search was carried out considering five specific document types i.e. article, proceeding paper, review, book chapters and editorial materials the number of publications was reduced to 3,776 publications as shown in figure 3. Table I indicates the individual results of each document type with their respective percentage which shows that the highest contribution was offered by the articles by the document type (article).

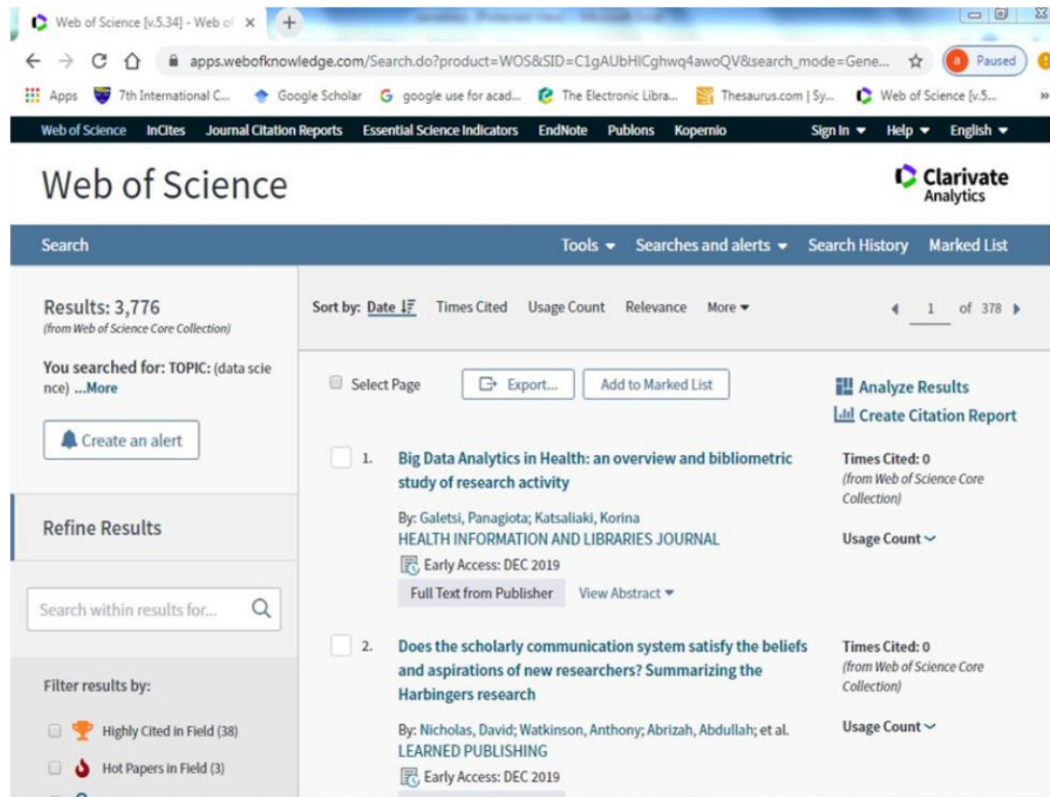


Figure 3 Collective Results of Specific Document Types

The following table summarizes the results taken by limiting the search to five specific document types

Rank	Document Types	Number of Publications	Percentage
1 st	Article	3,234	85.646%
2 nd	Proceedings Paper	458	12.129%
3 rd	Review	132	3.496%
4 th	Book Chapter	66	1.748%
5 th	Editorial Material	55	1.457%

Table I Individual Results of Each Document Type Used with their Respective Percentage Searched during 2010-2019

6.3 RQ3. Which was the most cited publication on Data Science in the LIS field on annual basis among the publications in WoS database Core Collection during the period 2010 to 2019?

Table II illustrates the list of top ten highly cited Library and Information Science articles published on Data Science by the LIS authors in WoS journals globally. The highly cited article published during the period 2010 to 2019 on Data Science was “*Business Intelligence and Analytics: From Big Data to Big Impact*” authored by Chen, Hsinchun; Chiang, Roger H. L.; Storey, Veda C. published in December 2012 in MIS Quarterly journal with 1,387 citations, the publication on 2nd rank was “*Negative Results are Disappearing from Most Disciplines and Countries*” authored by Fanelli and Daniele published in March 2012 in the “*Journal of*

Scientometrics” with 377 citations. At the 3rd rank was the article titled “ESP: A Tool to Estimate Scale Parameter for Multi-Resolution Image Segmentation of Remotely Sensed Data” authored by Dragut, Lucian; Tiede, Dirk and Levick, Shaun R. published in March 2010 with 374 citations.

Rank	Title	Authors	Source Journal	Publication Month and Year	Volume	Issue	Total Citations
1 st	BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT	Chen, Hsinchun; Chiang, Roger H. L.; Storey, Veda C.	MIS QUARTERLY	Dec 2012	36	4	1387
2 nd	Negative results are disappearing from most disciplines and countries	Fanelli, Daniele	SCIENTOMETRICS	March 2012	90	3	377
3 rd	ESP: a tool to estimate scale parameters for multi-resolution image segmentation of remotely sensed data	Dragut, Lucian; Tiede, Dirk; Levick, Shaun R.	INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE	March 2010	24	6	374
4 th	The rate of growth in scientific publication and the decline in coverage provided by the Science Citation Index	Larsen, PederOlesen; von Ins, Markus	SCIENTOMETRICS	Sep 2010	84	3	348
5 th	The journal coverage of Web of Science and Scopus: a comparative analysis	Mongeon, Philippe; Paul-Hus, Adele	SCIENTOMETRICS	Jan 2016	106	1	302
6 th	The Conundrum of Sharing Research Data	Borgman, Christine L.	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	June 2012	63	6	246
7 th	Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison	Harzing, Anne-Wil; Alakangas, Satu	SCIENTOMETRICS	Feb 2016	106	2	217
8 th	Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references	Bornmann, Lutz; Mutz, Ruediger	JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY	Nov 2015	66	11	202
9 th	Taiwan's National Health Insurance Research Database: administrative health care database as study object in bibliometrics	Chen, Yu-Chun; Yeh, Hsiao-Yun; Wu, Jau-Ching; Haschler, Ingo; Chen, Tzeng-Ji; Wetter, Thomas	SCIENTOMETRICS	Feb 2011	86	2	189
10 th	The Leiden ranking 2011/2012: Data collection, indicators, and interpretation	Waltman, Ludo; Calero-Medina, Clara; Kosten,	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	Dec 2012	63	12	188

Table II Top Ten Highly Cited Articles Published on Data Science in the LIS field During the Period 2010 to 2019

The following figure shows that the number of citations was maximum in the year 2019.

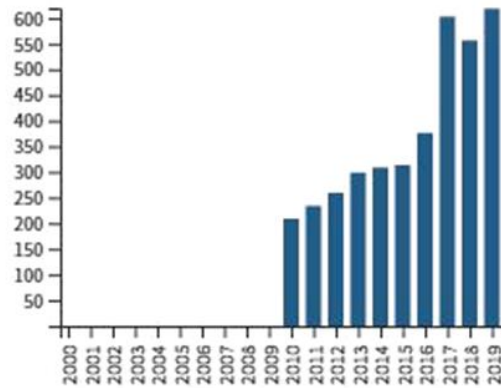


Figure 4 Number of citations per year

6.4 RQ4. Which was the most prolific and highly cited Data Science author/group authors and journal in the LIS field in WoS database Core Collection during the period 2010 to 2019?

Table III demonstrates the results of the study that the most prolific author on Data Science in the LIS field was Bormann Lutz with a total of 69 publications during the period 2010 to 2019 on Data Science in the LIS field. Thelwall M. was at the 2nd rank with 49 publications, however, he was declared as the most productive author in a study conducted by Ahmad, Sheikh and Rafi in 2019. Table IV summarizes that the Institute of Electrical and Electronics Engineers (IEEE) was the highly cited research organization/group author with a total of 17 publications on Data Science in the LIS field during the period 2010 to 2019. The highly cited journal was the “*Journal of Scientometrics*” during the period 2010 to 2019 on Data Science in the LIS field as shown in Table V

Rank	Authors	Record Count	Percentage Contribution
1 st	Bornmann L.	69	1.827%
2 nd	Thelwall M.	49	1.298%
3 rd	Leydesdorff L.	42	1.112%
4 th	Sugimoto CR.	18	0.477%
5 th	Costas R.	17	0.450%
6 th	Onyancha OB.	16	0.424%
7 th	Mutz R.	15	0.397%
8 th	Robinson-Garcia N.	15	0.397%
9 th	Haunchild R.	14	0.371%
10 th	Rousseau R.	14	0.371%

Table III Contribution of Various Authors during 2010-2019 on Data Science in LIS Field

Rank	Group Authors	Number of Publications	Percentage
1 st	IEEE	17	0.450%
2 nd	ACM	10	0.265%
3 rd	ASSOC COMP MACHINERY	5	0.132%
4 th	IST	4	0.106%
5 th	SOC IMAGING SCI TECHNOL	3	0.079%
6 th	ACAD CONF LTD	1	0.026%
7 th	CTR CAUSAL DISCOVERY TEAM	1	0.026%
8 th	DIRECT2EXPERTS COLLABORATION	1	0.026%
9 th	INFORMATION RESOURCES MANAGEMENT ASSOCIATION	1	0.026%
10 th	N4U CONSORTIUM	1	0.026%

Table IV Highly Contributing Research Organizations on Data Science in LIS Field During 2010 to 2019

Rank	Journal Title	Number of Citations	Percentage
1 st	<i>Journal of Scientometrics</i>	715	18.935%
2 nd	<i>Journal of Informetrics</i>	189	5.005%
3 rd	<i>Journal of the Association for Information Science and Technology</i>	144	3.814%
4 th	<i>Proceedings of the International Conference on Scientometrics and Informetrics</i>	90	2.383%
5 th	<i>Electronic Library</i>	83	2.198%
6 th	<i>International Journal of Geographical Information Science</i>	83	2.198%
7 th	<i>Information Research and International Electronic Journal</i>	76	2.013%
8 th	<i>Journal of the American Medical Informatics Association</i>	76	2.013%
9 th	<i>Journal of the American Society for Information Science and Technology</i>	73	1.933%
10 th	<i>Journal of Documentation</i>	69	1.827%

Table V Top Ten Highly cited journals in the LIS field

6.5 RQ5. Which were the most productive country and universities on Data Science in the LIS field in WoS database Core Collection during the period 2010 to 2019?

One of the most important findings of this study included the most productive country and university on Data Science during the period 2010 to 2019. This study showed that the most productive country was the United States of America (USA) with a total number of 969 publications, Peoples Republic of China was at the 2nd rank with 365 publications, England, Spain and Germany were at 3rd 4th and 5th ranks with 306, 285 and 257 publications, Brazil and Canada were at 6th and 7th ranks with 240 and 165 publications, unfortunately, India was at the 3rd last position with 155 publications, Netherlands and Australia were at the bottom of the list with 154 and 116 publications each respectively during the period 2010 to 2019 on Data Science in LIS field.

Table VII shows that the most productive university was the Wuhan University, China with a total of 67 publications, University of Amsterdam, USA was at the 2nd rank with 54 publications, Indiana University, the UK with 52 publications was at 3rd rank with 52 publications, Leden University, the UK was at 4th rank with 51 publications, Max Planck Gessel Institute was at 5th rank with 47 publications, Center for Scientific and Information Communication, Canada was at 6th rank with 44 publications and at the bottom of the list were Drexel University, 44 publications, University of Illinois 43 publications, University Granada 42 publications and at the last Wolverhampton University with 40 publications on Data Science in LIS field during the period 2010 to 2019. The results showed that the contribution from individual American and European universities decreased during this period there was only one institute from the USA while most of the universities to the UK. The greatest contribution was offered by the Chines University Wuhan University which shows that work in the field of Data Science is increasing in China rapidly.

Countries	Number of Publications	Percentage
USA	969	25.662%
Peoples R China	365	9.666%
England	306	8.104%
Spain	285	7.548%
Germany	257	6.806%
Brazil	240	6.356%
Canada	165	4.370%
India	155	4.105%
Netherlands	154	4.078%
Australia	116	3.072%

Table VI Top Ten Most Productive Countries in LIS Field

Rank	Name of University	Number of Publications	Percentage
1 st	Wuhan University	67	1774%
2 nd	University of Amsterdam	54	1.430%
3 rd	Indiana University	52	1.377%
4 th	Leden University	51	1.351%
5 th	Max Planck Gessel Institute	47	1.245%
6 th	CSIC	44	1.165%
7 th	Drexel University	44	1.165%
8 th	University of Illinois	43	1.139%
9 th	University Granada	42	1.112%
10 th	Wolverhampton University	40	1.059%

Table VII Top Ten Most Productive Universities in LIS Field

7. Discussions

In this study, the Scientometrics research method was used to analyze several trends of Data Science in the LIS field. Scientometrics analysis refers to the scientific analysis of facts, figures, or any form of available data. Scientometrics analysis deals with the extraction of knowledge from data (both structured and unstructured form). Erfanmanesh *et al.* in 2017 also conducted the study and performed Scientometrics analysis in the field of LIS. They analyzed various trends related to multiple disciplines, researchers, institutions; countries etc. and find out their impact on society. They highlighted the importance of Scientometrics analysis in the field of LIS.

Due to a large amount of data producing recently from social media, digital resources and gadgets etc. a new branch of science has gained much importance recently. The branch of science which deals with the newer approaches, algorithms, methods and skills to collect, store and manage the collected data is called Data Science. A similar study was conducted by Virkus in 2019 to know the importance of Data Science as a new discipline of science. According to Virkus, (2019) Data Science is an emerging field, therefore excitement and confusion for this field are always present. Recently many studies have been performed on Data Science. Data Science has given much importance recently in workshops and conferences, a lot of training has been conducted recently and in the past to educate the data scientists and equip them with latest Data Science skills. Data scientists are highly demanded by prominent companies and they hire data scientists with good data management skills.

Like other fields of science, Data Science has a fundamental importance in the field of LIS. In the present study, Scientometrics analysis was conducted using the WoS Core Collection database. The purpose of the study was to highlight the importance of Data Science in the LIS field. In this study, the literature indexed in the WoS Core Collection database was considered to highlight the importance of Data Science in the LIS field from 2010 to 2019. Several similar studies have been performed using the WoS Core Collection database recently in LIS fields like Rafiq, Jabeen, Tahir, Jabeen and Yun, in 2015 also Jabeen *et al.* in 2016 and Blessinger and Frasier in 2007. The difference between their studies and the present study is that they performed a bibliometric analysis of LIS literature but they did not focus on the specific area of Data Science from a LIS perspective.

In the present study, various trends related to Data Science in the perspective of LIS is

evaluated during the period 2010 to 2019. A similar study was conducted by Blessinger and Frasier in 2007 but they analyzed only LIS publications published during the period 1994 to 2004. The present study has shown one of the similar results to Blessinger and Frasier that the LIS researchers and professionals are discussing the practical issues in recent research and the problems faced by them in their professional life.

The present study showed that the progress in the field of Data Science was greatest during the last decade i.e. 2010 to 2019. Virkus in 2019 also reported similar results and he also showed that the maximum numbers of LIS publications on Data Science were during the period 1980 to 2019 and they fall in the last decade i.e. 2010 to 2019. He has also shown a gradual increase of LIS publications to 44.4% during this period.

This study shows that the contribution of western and European countries on Data Science in LIS perspective was greatest during the period 2010 to 2019. Rafiq, Jabeen, Tahir, Jabeen and Yun, in 2015 also Jabeen *et al.* in 2016 and Blessinger and Frasier in 2007 gave similar results i.e. the contribution of the western and European authors on Data Science in LIS perspective is greatest in the world. In the present study, the USA is declared as a highly productive country in the world, similar results were presented by Jabeen *et al.* in 2015 and 2016 respectively. They declared the USA as the most productive country and the USA produced (43%) of the publications indexed in the WoS Core Collection database during the period 1980 to 2009.

One of the interesting trends presented in the current study is that the collaboration pattern of American and European countries on Data Science in LIS perspective was maximum during the period 2010 to 2019. The contribution of Asian authors in collaborative research is meagre and not so appreciable. Similar results for a collaborative pattern of research were presented by Jabeen *et al.* 2015 respectively. In their study, they showed that the American authors have collaborated with Australia, Canada, and the UK in LIS publications during the period 1980 to 2019.

Another result presented in the present study was the most productive university during the period 2010 to 2019 on Data Science in LIS perspective. It was concluded that the Wuhan University China was the most productive university on Data Science in the LIS field. Jabeen *et al.* presented a different result and showed that Victoria University, New Zealand was at the top of the list among the top ten universities of the world in LIS publications during the period 1980 to 2019. The reason for this contradiction was that Jabeen *et al.* considered all LIS publications indexed in the WoS Core Collection database (Data Science and non-Data Science) and the period for their study is much greater than the present study.

In this study, it was concluded that the most productive document type was “Article” during the period 2010 to 2019 on Data Science from the LIS perspective. A similar result was presented by Virkus and Garoufallou in 2019 that the “Article” was the most productive document type on Data Science in the LIS field during the period 1980 to 2019.

One of the other results presented in this study was the most prolific journal on Data Science in the LIS field during this period. The results showed that the journal “Journal of Scientometrics” with 715 citations was the most prolific. Virkus and Garoufallou in 2019 declared “Journal of the American Medical Informatics Association” with 11 publications as the most prolific journal on Data Science in the LIS field. The results of Sheikh, Ahmad and Rafi concluded “Journal of the Medical Library Association” as the most prolific journal with 218 citations. The similarity between all the three results was that all the journals belonged to the USA.

Among other results were the most productive year and the most prolific author on Data

Science from the LIS perspective. The results of this study showed that 2017 and 2019 were the most productive year in terms of publications on Data Science from a LIS perspective. Virkus and Garoufallou in 2019 reported a similar result i.e. the most productive publication years were 2015, 2017 and 2019 respectively. In the present study, the most prolific author was Bornmann Lutz with a total of 69 publications contrary to the Virkus and Garoufallou results i.e. Agarwal and Dhar (2014) with 112 citations.

8. Conclusions and Recommendations

The results of this study concluded, Data Science is a new field of the 21st century, the knowledge of this fundamental field of science is necessary for every LIS scholar, educationist, researcher, and every person who wants to or intends to progress in the field of LIS. The reviewed literature showed rapid progress in Data Science in the LIS field during the previous decade i.e. 2010 to 2019.

The contribution of the USA and the European countries is the greatest during this period. The contribution of China among the Asian countries is appreciable whereas the contribution of other Asian countries is meagre and limited. Unfortunately, very few studies have been performed by Pakistani researchers on Data Science in the LIS perspective during the period 2010 to 2019.

Since most of the data available currently are in unstructured form and heaps of data is generating from social media and electronic gadgets. The use of newer data technologies to manage, control, refine, and store available data is increasing day by day. Similarly, the need for analyzing this data by using newer algorithms, newer approaches, the latest systems and the latest methodologies to control the flow of data as a LIS professional is also increasing rapidly. Data Science must be a core subject of interest for every LIS scholar, researcher, educationist, and professional. Data Science must be added as a basic course in the curriculum of LIS students at both graduate level and postgraduate levels.

The main reason for conducting the present study was not only to evaluate the importance of Data Science in the LIS field but this study also gives some recommendations and future directions that how the research on Data Science in the LIS field can help the library professionals in Pakistan to uplift the LIS sector. It also enumerates that how LIS professionals can digitally manage and analyze Big Data.

In the light of research findings, the authors have suggested some recommendations for the prospective researcher:

- This study was limited to the WoS Core Collection database however the study should be expanded with Scopus and Google Scholar etc. in future.
- The scope of this study can be expanded by increasing the period of the study.
- The focus of the study may be increased in future by adding the areas and other disciplines of science like computer science and IT, humanities, medical sciences, applied sciences, human psychology and other social sciences etc.
- The researchers of the developing countries including Pakistan may develop collaboration with researchers from the developed countries like the USA and UK.
- The researchers from developing countries must have their academic profiles on social networking platforms related to Google Scholar, Research Gate, LinkedIn, and Mendeley etc. In this way, they can communicate and collaborate with their research fellows in the developed countries anytime and anywhere in the world. This will increase the research

productivity of their respective country and also increase their exposure to highly reputed international research projects.

- The LIS institutions in the developing countries should provide financial assistance and research funding to the LIS scholars, students, researchers and staff members based on the Data Science research work so that they can take more interest in research on Data Science in the LIS field.
- In developing countries including Pakistan, the government should focus on establishing more LIS schools and LIS research centres. The LIS schools must be fully equipped with the latest Data Science skills. Only in this way highly specialized “Data Scientists”, “Data Engineers”, “Data Curators” and “Data Managers” etc. can be produced in the LIS field in developing countries.
- Libraries may organize training and workshops on Data Science related to the LIS field both at college and university levels to create awareness about “Data Science Skills” among students, researchers and faculty members. Libraries must be equipped with the latest literature on Data Science in the form of books, brochures, flyers etc. so that the Data Science aspirants from all fields including LIS can use them as reading material for guidelines at any time.
- In developing countries including Pakistan, the course curriculum and a course outline of undergraduate, graduate and postgraduate classes must be revised and redesigned according to the new Data Science fundamentals and approaches. The MPhil and PhD scholars should be offered dual degree programs and should be given opportunities to collaborate with the American and European universities of high repute.
- The libraries in Pakistan may acquire the subscription of top citation databases including WoS database Core Collection & Scopus etc., digital libraries, e-learning institutes and digital platforms available on Data Science so that the LIS students, scholars, educationists, faculty members, researchers and professionals can use these digital resources any time to fulfil their research needs.

References

- Abdullahi, I., Kajberg, L. and Virkus, S. (2007). Internationalization of LIS education in Europe and North America. *New Library World*, Vol. 108 No's 1–2, pp. 7-24.
- Abrizah, A., Zainab, A. N., Kiran, K., & Raj, R. G. (2013). LIS journals scientific impact and subject categorization: A comparison between Web of Science and Scopus. *Scientometrics*, 94(2), 721–740. <https://doi.org/10.1007/s11192-012-0813-7>
- Agarwal, R., & Dhār, V. (2014). 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>
- Ahmad, K., Sheikh, A., & Rafi, M. (2019). Scholarly research in Library and Information Science: an analysis based on ISI Web of Science. *Performance Measurement and Metrics*. <https://doi.org/10.1108/PMM-06-2019-0024>

- Antell, K., Foote, J. B., Turner, J., & Shults, B. (2014). Dealing with data: Science librarians' participation in data management at association of research libraries institutions. *College and Research Libraries*, 75(4), 557–574. <https://doi.org/10.5860/crl.75.4.557>
- Blessinger, K., & Frasier, M. (2007). Analysis of a decade in library literature: 1994-2004. *College and Research Libraries*, 68(2), 155–169. <https://doi.org/10.5860/crl.68.2.155>
- Chu, H. (2015). Library & Information Science Research methods in library and information science : A content analysis. *Library and Information Science Research*, 37(1), 36–41. <https://doi.org/10.1016/j.lisr.2014.09.003>
- Chu, H. (2015). Research methods in library and information science: A content analysis. *Library and Information Science Research*, 37(1), 36–41. <https://doi.org/10.1016/j.lisr.2014.09.003>
- Dees, A.S. (2015). A bibliometric analysis of the scholarly publications of Librarians at the University of Mississippi 2008–2013. *The Journal of Academic Librarianship*, Vol. 41 No. 3, pp. 241-245.
- Garg, K.C. and Kumar, S. (2019). Bibliometric assessment of the global research output in *Jatropha curcas* Linn as reflected by papers indexed in science citation index-expanded. *Performance Measurement and Metrics*, Vol. 20 No. 1, pp. 17-26, available at: <https://doi.org/10.1108/PMM-07-2018-0019>
- Garoufallou, E., Balatsoukas, P., Siatri, R., Zafeiriou, G., Asderi, S. & Ekizoglou, P. (2008). Greek academic librarians' perceptions of the impact of Google on their role as information providers. *Education for Information*, Vol. 26 No. 2, pp. 133-145
- Han, P., Shi, J., Li, X., Wang, D., Shen, S., & Su, X. (2014). International collaboration in LIS: Global trends and networks at the country and institution level. *Scientometrics*, 98(1), 53–72. <https://doi.org/10.1007/s11192-013-1146-x>
- Herther, N.K. (2009). Research evaluation and citation analysis: key issues and implications. *The Electronic Library*, Vol. 27 No. 3, pp. 361-375.
- Ivanović, D. & Ho, Y.-S. (2016). Highly cited articles in the information science and library science category in social science citation index: A bibliometric analysis. *Journal of Librarianship and Information Science*, Vol. 48 No. 1, pp. 36-46, Retrieved from <https://doi.org/10.1177/0961000614537514>
- Jabeen, M., Yun, L., Rafiq, M. & Jabeen, M. (2015). Research productivity of library scholars: Bibliometric analysis of growth and trends of LIS publications. *New Library World*, Vol. 116 Nos. 7-8, pp. 433-454.

- Jabeen, M., Yun, L., Rafiq, M., Jabeen, M. & Tahir, M.A. (2015). Scientometric analysis of Library and Information Science journals 2003–2012 using Web of Science. *International Information & Library Review*, Vol. 47 Nos. 3–4, pp. 71-82.
- Jabeen, M., Yun, L., Wang, X., Rafiq, M., Mazher, A., Tahir, M. A., & Jabeen, M. (2016). A study to analyze collaboration patterns for Asian Library and Information Science (LIS) scholars on author, institutional and country levels. *Serials Review*. <https://doi.org/10.1080/00987913.2016.1139526>
- Ma, T.J. and Lee, G.G. (2017). Open access journals: a bibliometric study in SSCI. *Information Discovery and Delivery*, Vol. 45 No. 2, pp. 79-86.
- Maharana, R.K. & Das, A.K. (2014). Growth and development of LIS research in India during 1999–2013: A bibliometric analysis. *Chinese Librarianship: An International Electronic Journal*.
- Malone, T. & Burke, S. (2016). Academic librarians' knowledge of bibliometrics and altmetrics. *Evidence Based Library and Information Practice*, Vol. 11 No. 3, pp. 34-49, Retrieved from <https://doi.org/10.18438/B85G9Jpp>. 35-46.
- Meho, L.I. (2007), "The rise and rise of citation analysis", *Physics World*, Vol. 20 No. 1, pp. 32-36.
- Mongeon, P. and Paul-Hus, A. (2016), "The journal coverage of Web of Science and Scopus: a comparative analysis", *Scientometrics*, Vol. 106 No. 1, pp. 213-228, available at: <https://doi.org/10.1007/s11192-015-1765-5>
- National Science Board (2005). Long-lived digital data collections: enabling research and education in the 21st century. Retrieved from www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf (Retrieve January 16, 2019).
- NIST (2018). National institute of standards and technology (NIST) special publication 1500-1r1. NIST Big Data Interoperability Framework: Volume 1, Definitions. Version 2. NIST Big Data Public Working Group (NBD-PWG). Retrieved from [www.nvlpubs.nist.gov/nistpubs/Special Publications/NIST.SP.1500-1r1.pdf](http://www.nvlpubs.nist.gov/nistpubs/Special%20Publications/NIST.SP.1500-1r1.pdf) (Retrieved January, 2020).
- Park, T.K. (2008). Asian and pacific region authorship characteristics in leading Library and Information Science journals. *Serials Review*, Vol. 34 No. 4, pp. 243-251.
- Provost, F., & Fawcett, T. (2013). Data science and its relationship data-driven decision making. (1). <https://doi.org/10.1089/big.2013.1508>
- Thelwall, M. (2008). Bibliometrics to webometrics. *Journal of Information Science*, Vol. 34 No. 4, pp. 605-621, retrieved from <http://jis.sagepub.com/content/34/4/605.abstract>

- Tripathi, M., Jeevan, V.K.J., Babbar, P. and Mahemei, L.K. (2018). Library and Information Science research in BRICS countries. *Information and Learning Science*, Vol. 119 No's 3–4, pp. 183-202.
- Uzun, A. and Özel, M. (1996). Publication patterns of Turkish astronomers. *Scientometrics*, Vol. 37 No. 1, pp. 159-169, doi: 10.1007/BF02093492
- Virkus, S. (2015). Change and innovation in European library and information science education. Bid: *Textos universitaris de biblioteconomia i documentación* 35 (deseembre), Retrieved from [http:// bid.ub.edu/en/35/virkus.htm](http://bid.ub.edu/en/35/virkus.htm) (Retrieved January 12, 2020).
- Virkus, S., & Garoufallou, E. (2019). Data science from a library and information science perspective. *Data Technologies and Applications*, 53(4), 422–441. <https://doi.org/10.1108/DTA-05-2019-0076>
- Virkus, S., Mandre, S. & Pals, E. (2018). Information overload in a disciplinary context. In Kurbanoglu, S., Boustany, J., Špiranec, S., Grassian, E., Mizrachi, D. and Roy, L. (Eds), *Information Literacy in the Workplace*, Springer, Cham, pp. 615-624.
- Voulgaris, Z. (2014). Data scientist: The definitive guide to becoming a data scientist. *Technics Publications*, Westfield, NJ.
- Zhang, J., Fu, A., Wang, H., & Yin, S. (2017). The Development of Data Science Education in China from the LIS Perspective 2, 3–17.

Further Reading

- Kelleher, J.D. & Tierney, B. (2018), *Data Science*, MIT Press, Cambridge, MA.
- Ratner, B. (2017). Statistical and machine-learning data mining: Techniques for better predictive modeling and analysis of big data. *Chapman and Hall/CRC Press*, Boca Raton.