

A Bibliometric Analysis of Data Governance Research: Trends, Collaborations, and Future Directions

South Asian Journal of
Business Insights
2023,3(1), 70-92.
ISSN 2773-7012(print)
ISSN 2773-6997(online)
© 2023, Faculty of
Management and Finance
University of Ruhuna,
Sri Lanka

B.B.D.S. Abeykoon *

Faculty of Management and Finance, University of Ruhuna
dilhani@mgt.ruh.ac.lk



A.B. Sirisena

Faculty of Management and Finance, University of Ruhuna
amila@badm.ruh.ac.lk

Abstract

Research on data governance has been shaped by an excessive number of articles published in recent decades. The research study used bibliometric techniques to examine 757 journal articles in the Scopus database and it shows a dynamic growth of the topic in the selected years from 2012 to 2022. The methodology is based on the performance analysis technique and science mapping technique in bibliometric analysis. The results explore the contribution to the domain in terms of authors, sources, and countries, disclose the network between the countries and scholars relevant to the publications, and provide research areas that need more attention which drives data governance. Data protection, data access, artificial intelligence, data management, cloud computing, blockchain, and data transformation are among the areas where there is a lack of information, therefore academics should pay closer attention to them in the future. The authors observed that a wide range of disciplines have been used to address the topic of data governance. The study lays the foundation for academics, practitioners, and policymakers to conduct more future research and leads to advancing data governance research.

Keywords: Bibliometric Analysis, Data Governance, Performance Analysis, Science Mapping

* Corresponding Author - dilhani@mgt.ruh.ac.lk

Introduction

Data governance implies the management and control of data, along with its authority and control in an organization. Also, the term 'Governance' in general refers to the way an organization goes about ensuring that strategies are set, monitored, and achieved. The researchers are specifying their publications on fragmented foci on the topic of "data governance". Historically, data was considered a by-product of legacy systems and had little value which was not a valuable shared asset. The environment for data assets started to become prominent after the early 1990s, with business decisions based on data analysis (Al-Ruithe et al., 2019). Currently, data governance possesses a higher level of importance for enterprises and government institutions due to several key trends. Numerous diverse sources generate data in massive volumes, which leads to inconsistencies. Hence, decisions should not be made with incorrect data before processing it into a consistent source. Enterprises have grown fast with the reality of everlasting data complexity, and handling the data requires different kinds of techniques that drive the new demands. Organizations seeking future opportunities eagerly believe that data governance is the ultimate solution to the arising complexity of data (Abraham et al., 2019).

As part of data governance, formal policies and procedures are defined, implemented, and enforced for the generation, storage, usage, and maintenance of information. A collection of uniform, coherent perspectives of the enterprise can be achieved by organizations through effective data governance by rationalizing data from numerous platforms. According to studies, the quality of data can be improved thanks to structured data governance systems (Were & Moturi, 2017). Data governance should offer rewards and penalties to encourage the right behavior in those who are responsible for gathering, managing, and using data. Collaboration between the individuals and organizations that make up the system, outside of a single company, is essential for data governance. To ensure that the correct data is securely and reliably transmitted across participating companies while complying with the General Data Protection Regulation (GDPR), trustworthy frameworks are necessary in this multi-organizational setting (Janssen et al., 2020). Both data governance and data literacy, which address data quality and research data management, are crucial building blocks in the knowledge base of information workers supporting data-intensive research. Further, determining decision domains and decision-making responsibilities is made easier by applying data governance to research data management procedures and data literacy training. Since data governance is a service based on standardized, repeatable processes and is intended to facilitate cost savings and transparency on data-related processes, adopting it is advantageous. It is also helpful because it discusses laws, regulations, and standards, as well as decision-making authority, accountability, and enforcement strategies (Koltay, 2016).

The "data-driven transformation" has a growing impact on both the economy and society. The so-called "transformation phase" toward a global "data economy" is represented by this development. As a result, new technical trends, including the Internet of Things, automated manufacturing, artificial neural networks, big data analytics, autonomous networked systems, and reference architectures for smart cities are being built upon. These new technologies are built on the top of digital data and information based on data governance (Cuno et al., 2019). One of the growing topics in the information systems (IS) community is data governance. The amount of data used within organizations has drastically expanded in recent years, playing a

crucial role in company operations. The amount of data used by organizations has rapidly expanded in recent years, and this has a major impact on how well enterprises operate. Particularly, statistics have an impact on both operational and strategic choices. For data to be viewed as a valuable asset, understanding how to govern it has also become crucial. Data governance is now seen as an emerging topic and has grown in popularity quickly. Thus, practitioners view data governance as a potential strategy for businesses to preserve and improve the usability and quality of their data (Alhassan et al., 2016).

Numerous research studies worldwide have discussed the application of data governance (Al-Ruithe et al., 2019; Benfeldt et al., 2020; Gupta et al., 2020; Kroll, 2018; Kuziemski & Misuraca, 2020). Also, both private and public organizations are using them. Wise decisions made through data governance prevent the development of new problems that can jeopardize an organization's life (Oudat et al., 2021). This study mainly focused on the literature published during the 2012-2022 period. To the best of our knowledge, this is the most recent study on data governance using bibliometrics. Even though there are few studies on reviewing data governance, they have not analyzed the bibliometrics with recent data and have not revealed the dynamics between the various foci in the knowledge domain. The identified research gap is filled, and a vast domain of data governance is covered with this quantitative approach. The main objectives of this study were: (1) to uncover the contribution to the domain in terms of authors, sources, and countries. (2) To disclose the network between the countries and scholars relevant to the publications. (3) To provide research areas that need more attention this drives data governance. This study considerably contributes to the fast-growing research on data governance by laying the foundation for future research and investigations.

This study focused on the literature published during the 2012-2022 period which includes 757 journal articles. Researchers can assess whether the study's paper count is sufficient to support the use of bibliometric analysis. The research field can be considered large enough if the number of papers in the database is in the hundreds or thousands (e.g., 500 or more). If there are only tens (e.g., 50) or low hundreds (e.g., 100-300) of papers, then it is considered a small database and does not warrant bibliometric analysis. However, the application of bibliometric analysis in business research is relatively new and underdeveloped. It has gained high popularity in recent years, and scholars use it for various purposes, such as to uncover journal and article performance, collaboration patterns, and the intellectual structure of a specific domain in the literature (Donthu et al., 2021). The application of quantitative techniques to bibliometric data describes the bibliometric methodology, and some of the databases, such as Scopus and Web of Science, have made the acquisition of large volumes of bibliometric data relatively easy (Burton et al., 2020). The bibliometric software, such as VOSviewer and Gephi, enables the analysis of such data in an interesting way for scholars (Kumar et al., 2021).

The remainder of the article describes the research methodology and discusses the results of the investigation. The methodology section explains the systematic process used to conduct and analyze the literature. The findings are described in the results section. The paper marks its end by addressing several implications, limitations, and potential future directions of data governance.

Methodology

Study Selection Process and Methods

This study used the Bibliometric Analysis methodology which is a popular and rigorous method for analyzing large volumes of scientific data that enables researchers to reveal emerging areas in a specific field. The four-stage process pertains to bibliometric analysis, which includes the steps of defining the aims and scope of the bibliometric study, choosing techniques for bibliometric analysis, collecting the data for bibliometric analysis, and running the bibliometric analysis and reporting the findings (Donthu et al., 2021). The scope of the study should generally be large enough before defining the research's scope and the goals should be established as bibliometric analysis is intended to handle massive volumes of bibliometric data (Ramos-Rodríguez & Ruíz-Navarro, 2004).

The two techniques, *the performance analysis technique* which accounts for the contributions of research constituents and *the science mapping technique* which focuses on the relationships between research constituents were used as analysis techniques. They were used to address the first, second, and third objectives. The first objective was (1) to uncover the contribution to the domain in terms of authors, sources, and countries. The second objective was (2) to disclose the network between the countries and scholars relevant to the publications. And, the third objective was (3) to provide research areas that need more attention which drives data governance. Thus, the first and the second steps of bibliometric analysis were completed. The third step of the research study is to collect the data for the selected bibliometric analysis techniques. The Scopus database was used in the study to identify the publications on data governance that were published from 2012 to 2022. This database was selected based on its usage in the majority of literature review studies, which have comprehensive coverage. The search term for the current study was “*data govern**” as the term can be included in the research articles with different suffixes. The other criteria for the selection of the database were the inclusion of “Final Journal Articles” in “English” during the period. Moreover, data cleaning should be done essentially because the database was not merely designed for bibliometric analysis. The duplicate data records and the erroneous data entries were removed from the corpus, leading to a high-quality bibliometric data analysis. Figure 1 shows the procedure for database selection in the research study with the inclusion and exclusion criteria.

Method of Analysis

The Biblioshiny and VOSviewer software were used to load the data from the Excel sheet and provide results for the bibliometrics study. The statistical software RStudio “Cherry Blossom” Release for Windows and VOSviewer were used for visualization purposes in the study. Biblioshiny combines the functionality of the bibliometric package with the ease of use of web apps using the Shiny package environment. Bibliometric analysis can be performed using a variety of algorithms provided by the Bibliometrix package, and data matrices for co-citation, coupling, scientific collaboration analysis, and co-word analysis can be built using these routines. Additionally, new knowledge consistently appears at points where structural and temporal evolution converges, such as in network analysis, factorial analysis, and thematic mapping. And the findings are described in the results section (Huang et al., 2021).

The VOSviewer was primarily used to create the visualization of the keyword co-occurrence network. Among the most frequently utilized units of analysis of bibliometric networks are the keywords that denote an article's main substance. Using the co-occurrence of the keywords in an article, different links can be made (Aparicio et al., 2019). In a map known as "keyword co-occurrence network visualization," the VOSviewer depicts these connections (Andrić et al., 2023; Dogra & Priyashantha, 2023; Priyashantha, 2023; Priyashantha, De Alwis, et al., 2022). The term co-occurrence analysis is essential to identifying the term used in article abstracts. The term co-occurrence network visualization was created using the term co-occurrence data that the VOSviewer had retrieved. The phrase is divided into numerous clusters using this technique (van Eck & Waltman, 2014). The analysis conducted through the software helps to achieve the three objectives of this study by uncovering the contribution to the domain in terms of authors, sources, and countries, by disclosing the network between the countries and scholars relevant to the publications, and by providing research areas that need more attention which drives data governance.

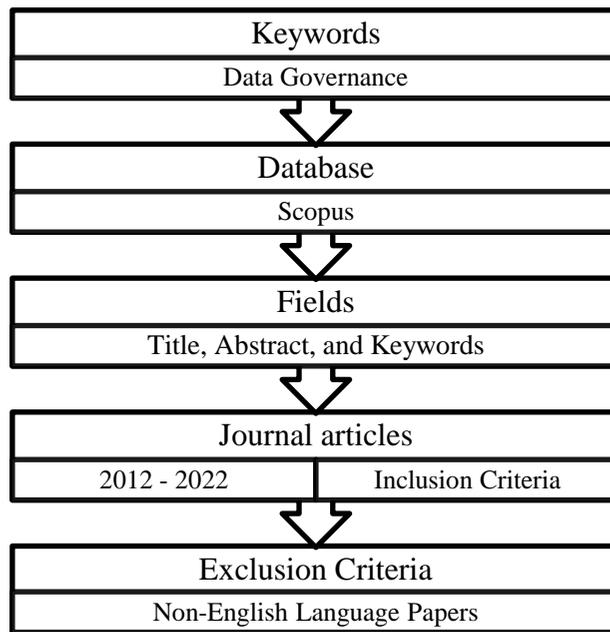


Figure 1: Database selection criteria

Results

The fourth and final step of the procedure for bibliometric analysis is to run the analysis and report the findings. This study contains trends and equivalent rationales that craft insightful discussions rather than simply reporting a summary of the bibliometric corpus. The figures and the tables are used for visualization purposes over descriptive discussions.

Descriptive Analysis

The study reviewed 757 research publications published between 2012 and 2022 in 485 journals. 2959 authors have published them in respective journals. Since the bibliometric analysis was used to review the articles, 2221 keywords, and 35354 references were used. According to the information retrieved, the percentage of the annual growth rate of publications is 32.82%, and the average number of citations per document is 13.29. The number of authors of single-authored documents is 138, while there are 149 single-authored documents in the corpus. Moreover, the value for co-authors per document is 4.36, and the percentage of international co-authorship is 24.83. This information is depicted in Table 1.

Table 1: Preliminary details of the corpus

Description	Results
MAIN INFORMATION ABOUT DATA	
Timespan	2012:2022
Sources (Journals)	485
Documents	757
Annual Growth Rate %	32.82
Average citations per doc	13.29
References	35354
DOCUMENT CONTENTS	
Author's Keywords	2221
AUTHORS	
Authors	2959
Authors of single-authored docs	138
AUTHORS COLLABORATION	
Single-authored docs	149
Co-Authors per Doc	4.36
International co-authorships %	24.83
DOCUMENT TYPES	
article	757

Main Analysis

Figure 2 represents year-by-year research articles production and shows that the number of articles published during the 2012-2022 period has increased. Accordingly, it shows that scholars are becoming increasingly concerned about data governance during this period. A highly inclined area is visible after the year 2016, and the trend is much more progressive after the year 2020.

Recent years have seen a tremendous increase in the scholarly literature on data governance, and Figure 3 depicts the regional distribution of scholars in this topic area, with the USA, United Kingdom, China, Canada, and the Netherlands at the top. The figure is composed of both single-country publications in green and multiple-country publications in brown. In addition to the top five countries, countries such as Italy, Australia, Germany, Switzerland, India, Korea, Brazil, Spain, etc. have significantly contributed to the research stream on data governance. Among all the top countries listed in the figure, Finland is geared towards single-country publications only.

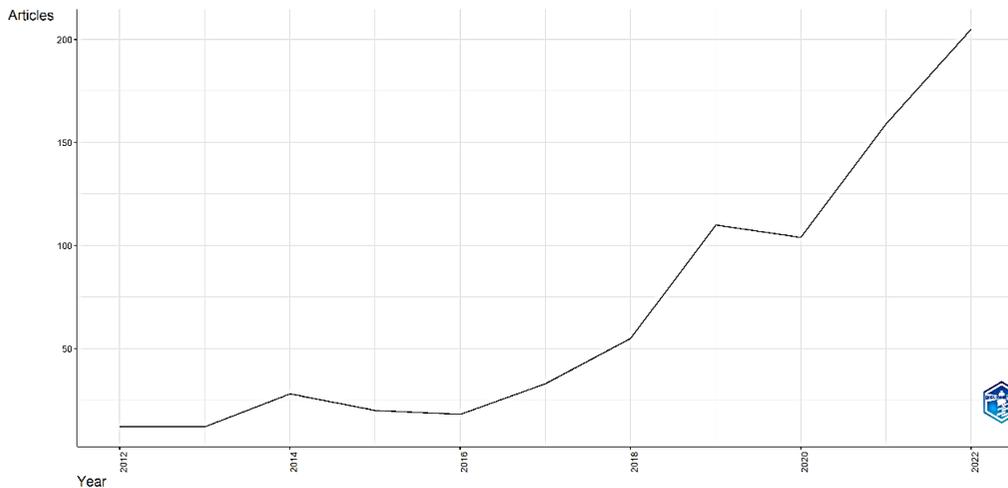


Figure 2: Year-wise publications

Figure 4 shows the research cooperation on the world map, and as a whole, this collaboration diagram shows that the research has been mostly restricted to developed countries and mostly inexistent in other countries. Hence, the researchers in developing countries should be motivated to conduct their research studies on data governance further.

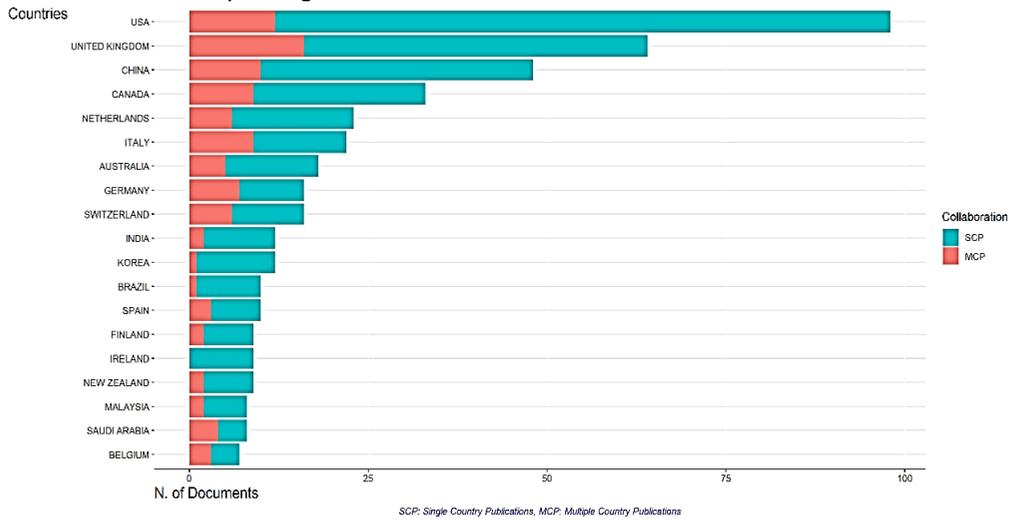


Figure 3: Distribution of scholars world-wide

Table 2 lists the top 25 most relevant sources which have published the greatest number of research articles in the period of 2012-2022 on data governance. The journals *Sustainability* (Switzerland), *Journal of the American Medical Informatics Association*, *Big Data and Society*, *Internet Policy Review*, and *Computer Law and Security Review* at the top have published 18, 14, 12, 11, and 9 research publications so far during the considered period.

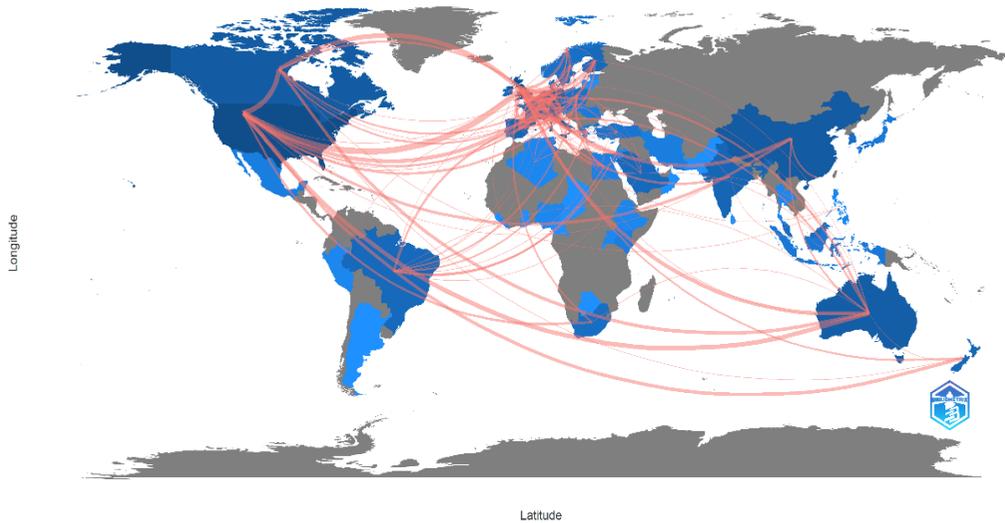


Figure 4: Country collaboration map

Table 2: Most relevant sources

Sources	Articles
Sustainability (Switzerland)	18
Journal of The American Medical Informatics Association	14
Big Data and Society	12
Internet Policy Review	11
Computer Law and Security Review	9
BMJ Open	8
Data and Policy	8
Journal of Decision Systems	8
Applied Marketing Analytics	7
IEEE Access	7
Computer Fraud and Security	5
Frontiers in Genetics	5
Information Communication and Society	5
International Journal of Advanced Computer Science and Applications	5
International Journal of Population Data Science	5
Journal of Data Protection and Privacy	5
Journal of Medical Internet Research	5
PLOS One	5
Spectroscopy (Santa Monica)	5
Yearbook of Medical Informatics	5
Future Internet	4
Information (Switzerland)	4
International Journal of Digital Earth	4
International Journal of Information Management	4
Journal of Intellectual Property, Information Technology and E-Commerce Law	4

Figure 5 shows the production of publications over time by the five most relevant journals that publish data governance research: Big Data and Society, Computer Law and Security Review, Internet Policy Review, Journal of the American Informatics Association, and Sustainability

(Switzerland). The Journal of the American Informatics Association shows a significantly higher growth of production in the initial stage of the time frame than the other sources. Also, the Computer Law and Security Review Journal, and the Big Data and Society Journal represent increases in production starting from the year 2013. However, overall publication production after 2017 has shown continuous rapid growth for all the sources mentioned.

Figure 6 depicts the most productive scholars over the study period, along with the number of publications they have produced. The most productive scholars were determined using an author's dominance diagram. The top 10 are displayed in the figure. The five most influential scholars in data governance in the period of 2012-2022 are Carroll SR., Sammon D., Hudson M., Knoppers BM., and Mcdowall RD. Also, the figure shows the number of articles produced by the authors and the total citations per year. The study's conclusions show how important the mentioned authors to the advancement of the field's research agenda.

A Sankey diagram was used to study the flow and the relationship between authors, keywords, and sources in data governance research. It is also called the three-field plot, which is shown in Figure 7. Accordingly, this diagram is used to illustrate the flow between the mentioned entities, and the width of the arrows and the boxes show the frequency of publications (Soundararajan et al., 2014). The word *data governance* is the most prominent keyword, as the study is based on that. In addition, *data sovereignty*, *data protection*, *privacy*, *big data*, *artificial intelligence*, *COVID-19*, *cloud computing*, *blockchain*, *data sharing*, *data quality*, *public health*, *machine learning*, *ethics*, *data management*, etc., are some of the most prominent keywords in the diagram. Moreover, Carroll SR., Hudson M., Floridi I., Vayena E., and Benkhelifa E. are a few of the authors who have utilized the most diverse terms out of all the authors.

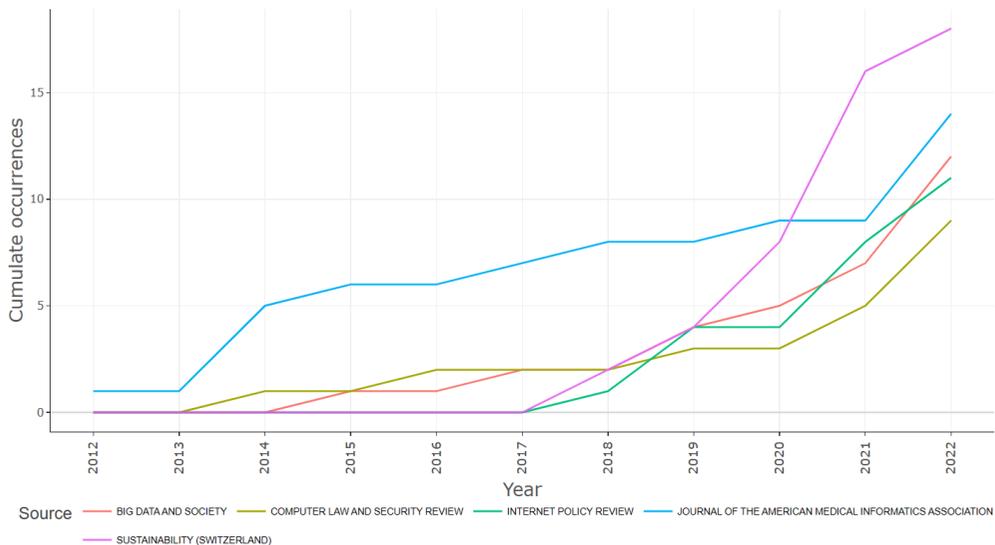


Figure 5: Source's production over time

The use of co-citation analysis in science mapping implies that works frequently cited together have similar themes (Hjørland, 2013). The co-citation analysis of the authors of the articles resulted in Figure 8. Accordingly, the most influential researchers for data governance research are Otto, b., Janssen, m., Wang, y., Zhang, y., Kithcin, r., Vayena, e., and Floridi, i. typically; some scholarly papers link distinct clusters together and receive citations from various clusters. The national collaboration network for data governance research is shown in Figure 9. This illustration demonstrates that the United States of America is the country that contributes the most to data governance research. Also, it represents strong collaborations with other nations such as the United Kingdom, China, Canada, Australia, and the Netherlands. Minor collaborations can be seen between South Korea, India, Iran, Taiwan, South Africa, etc. These results, notwithstanding the existence of some degree of cooperation between surrounding countries, demonstrate that global collaborative networks are not determined by geographical proximity or linguistic closeness.

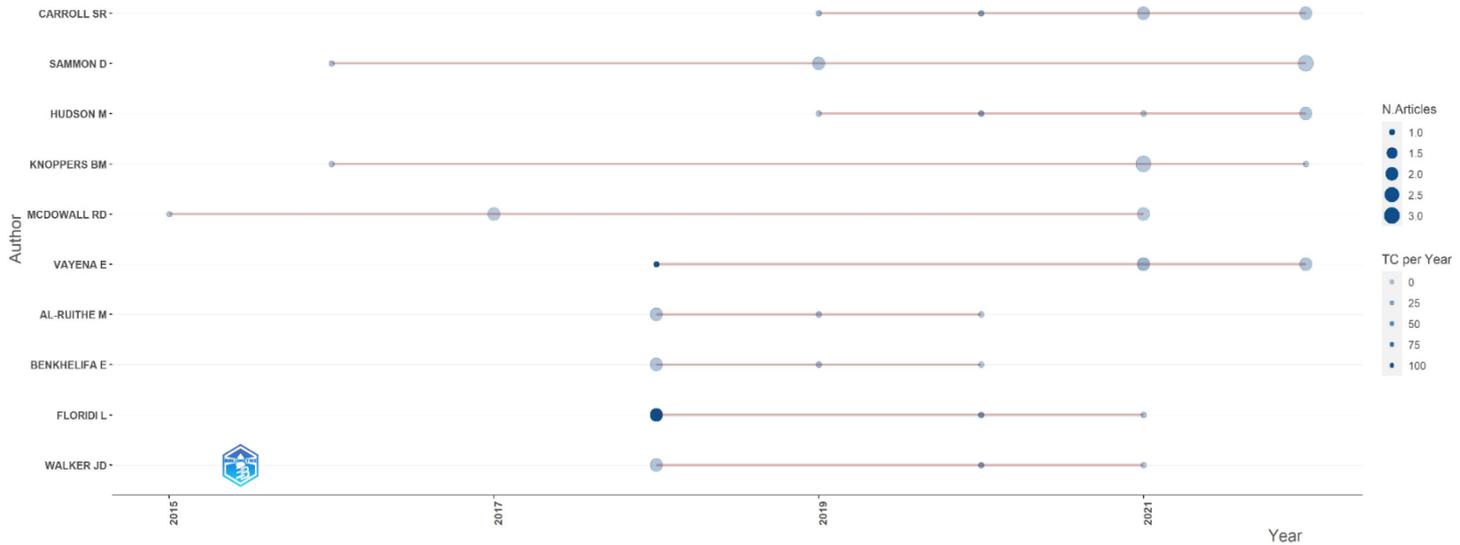


Figure 6: Author's production over time

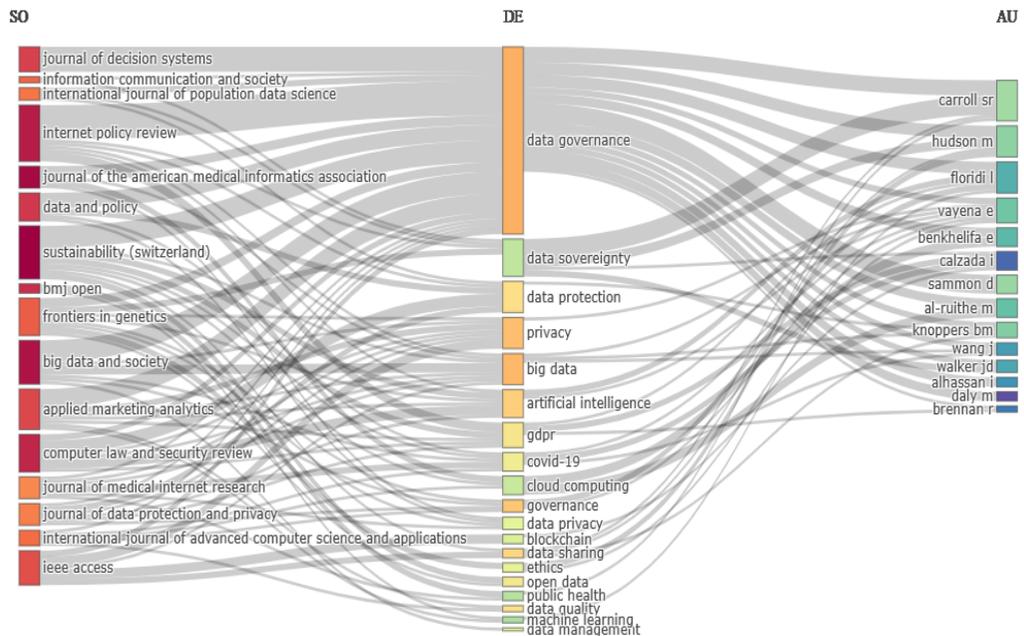


Figure 7: Sankey diagram of data governance research (source-keyword-author linkages)

A keyword co-occurrence analysis was conducted with a word cloud in complement to the Sankey diagram, as shown in Figure 10. Keywords represent the studies to the scientific community as they are one of the main units of analysis in a bibliometric study (Corbet et al., 2019). The VOSviewer software gives output for the dataset provided as depicted, and the clusters have been formed in the word cloud, which is shown in different colors. The clusters are formed by the frequency of the words. Additionally, the breadth of the linkages reflects the number of related term co-occurrences, and the node sizes are proportionate to their frequency (Neff & Corley, 2009).

A single or multiple relationship can be linked with nodes in a network diagram. A simple line is used to depict the link between two nodes. A pair of nodes can be developed into a cluster with the accumulation of more data and relationships. Thus, network analysis can reveal interesting patterns and future research areas, along with the development of theories (Khan & Wood, 2016). Network analysis techniques such as keyword co-occurrence, collaboration networks, and co-citation networks were used in the study. The generated word cloud has the most prominent term, *data governance*, as the searched keyword was the same. It contains clusters, and these clusters consist of other *noteworthy words such as big data, data quality, data protection, data ethics, data management, electronic health records, data security, artificial intelligence, machine learning, data ownership*, etc. Ten clusters are available in the generated word cloud for the study, and the clusters in red, green, yellow, blue, purple, and yellow are significant.

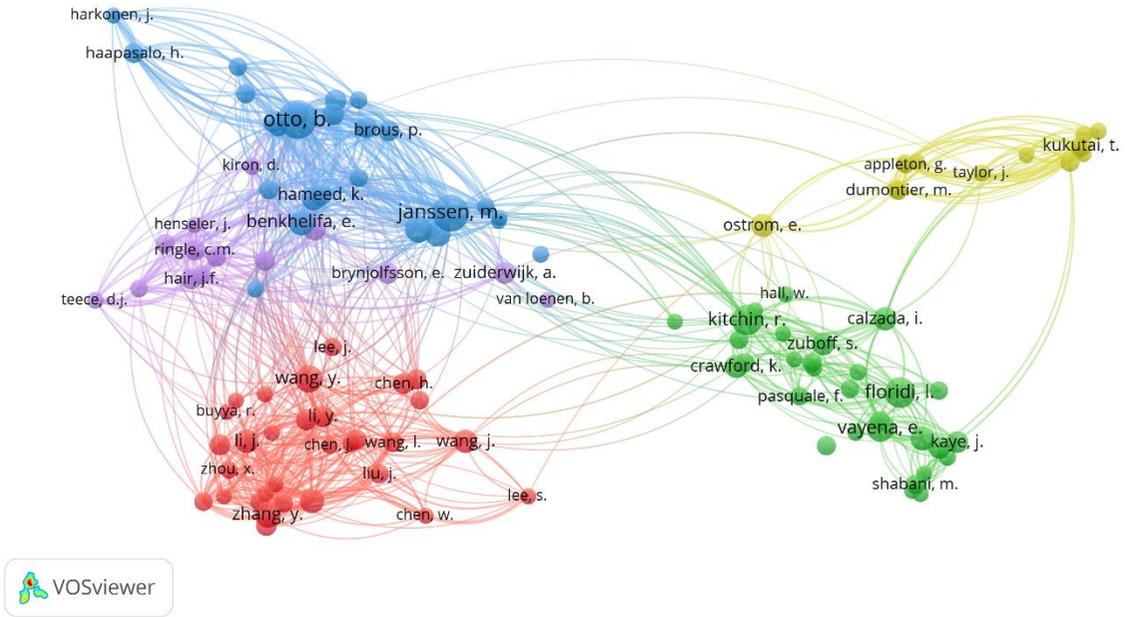


Figure 8: Collaboration network of authors in data governance research

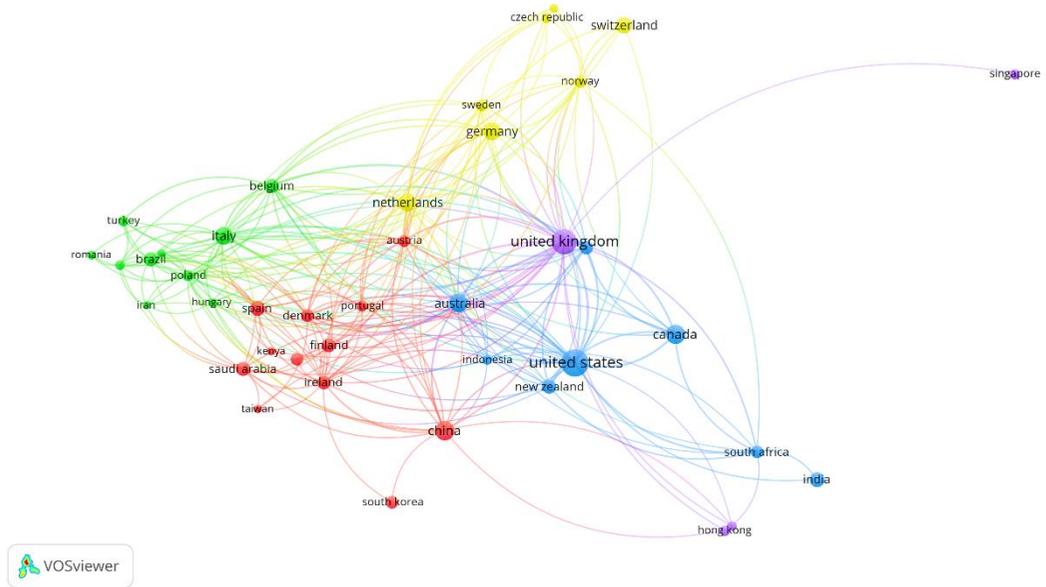


Figure 9: Collaboration network of countries in data governance research

The density visualization of keyword occurrences for the study is depicted in Figure 11, which was obtained again using VOSviewer software. Keywords falling into the red color area indicate established knowledge in the research area (van Eck & Waltman, 2014). While keywords found in the green area indicate that very little study has been done, those found in the yellow area imply a reasonable degree of investigation (Priyashantha, Dahanayake, et al., 2022). Accordingly, as in the network visualization diagrams, the author's searched keyword area is the densest area with the red and it shows us that established knowledge exists for data governance while moderate knowledge has been generated for the area *smart city*. However, insufficient knowledge is indicated on the topics of *data protection*, *data access*, *artificial intelligence*, *data management*, *cloud computing*, *blockchain*, *data transformation*, etc. Therefore, worldwide researchers should be more attentive on these topics in the future.

Following the suggestions from the literature, a multiple correspondence analysis (MCA) was used to investigate the conceptual framework of data governance research (Demiroz & Haase, 2019; Mostafa, 2020; Pazmiño-Sarango et al., 2022). As seen in Figure 12, this maps the proximity of the keywords inside the publications. On the map, the keywords with comparable distributions are located closer together (Mobin et al., 2021). The map has revealed two clusters, as in the MCA image. The red cluster, which is the largest, sheds light on the function of data governance with technology. The keywords such as *machine learning*, *internet of things*, and *metadata* imply that. Meanwhile, the blue cluster focuses on data governance related to governments with keywords like *data management*, *public sector*, *GDPR (General Data Protection Regulation)*, *trust*, *data sovereignty*, etc.

Thematic maps use the clusters in keyword co-occurrence analysis to discover their evolution and dynamics (Chakraborty et al., 2021). Density and centrality indicators provide the base for thematic maps and build research themes and relations (Stopar, 2016). Thematic maps have been used in several bibliometric studies because of their significance and simplicity (Armenta-Medina et al., 2020; Fortuna et al., 2020; Knapczyk et al., 2020).

The study's thematic map, shown in Figure 13, classifies the research concentrations according to their levels of centrality and density (Mobin et al., 2021). The centrality of the theme symbolizes the extent of the relationship among distinct topics, while density symbolizes the progress (Esfahani et al., 2019). The map's recurring keywords have clusters that establish themes and lay the groundwork for research in the quadrants (Cobo et al., 2011). The data governance-related themes are shown in the figure. The quadrants are divided with dotted lines based on the criteria mentioned. The number of publications where the keywords appear is shown by the size of the bubbles in the diagram. The motor themes, which are topics with high density and centrality, make up the first quadrant. The first quadrant's topics are well-developed and connected both inside and outside. The high-density and low-centrality themes belong to the second quadrant, called *niche themes*. The third quadrant has *emerging or declining themes* that have low density and centrality values. They reflect weak external and internal ties. The final quadrant has *basic themes* with sharp external ties and insignificant internal ties (Rejeb et al., 2022). Accordingly, the basic themes of *COVID-19*, *public health*, *blockchain*, *big data*, *smart cities*, *artificial intelligence*, etc. can be identified. Moreover, niche themes such as *data sovereignty* and *big data management* can also be shown.

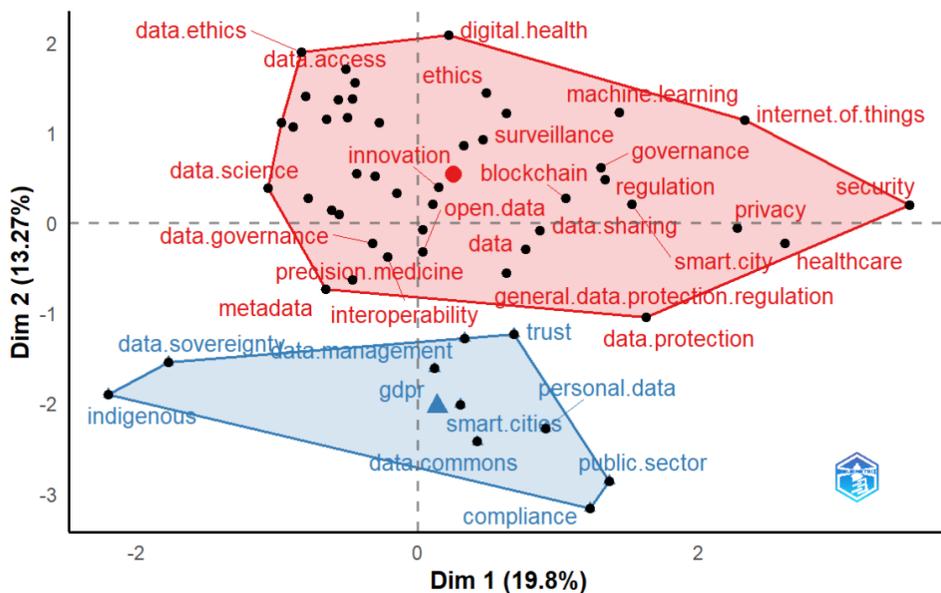


Figure 12: Conceptual map in data governance research

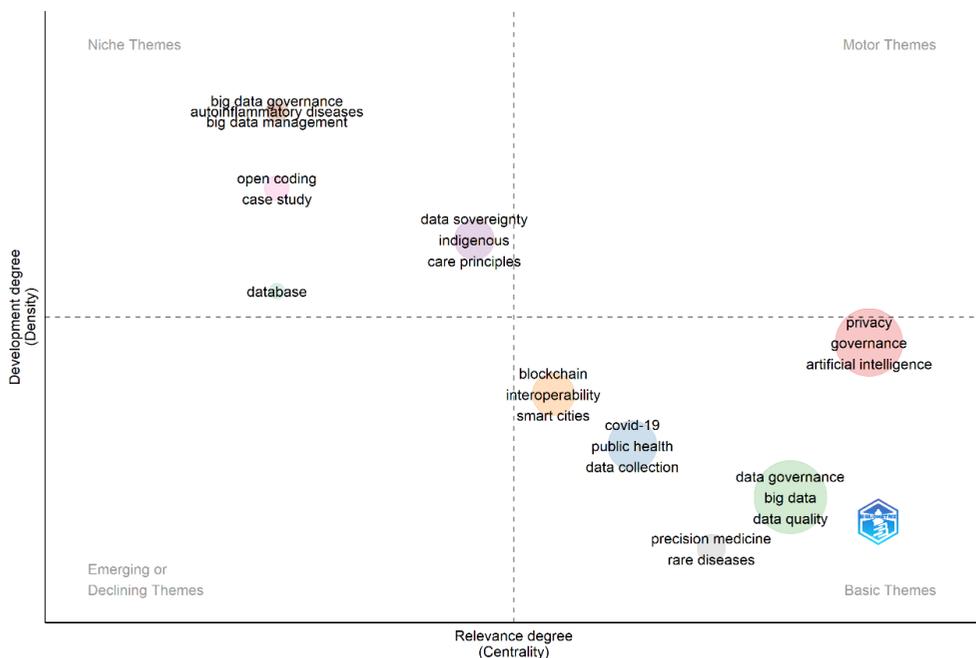


Figure 13: Thematic map in data governance research

Discussion

The bibliometric analysis was conducted to achieve the three main objectives. The first objective was: (1) to uncover the contribution to the domain in terms of authors, sources, and countries. Publication-related metrics under the performance analysis were used to address it. First, a graph of the year-by-year research articles was created to show the domain's trend, and it was discovered that production increases towards the end of the considered period. The top five nations that have made a substantial contribution to the data governance research stream are the United States, the United Kingdom, China, Canada, and the Netherlands. Finland is the only one of the top nations that only supports single-country publications. The journals Sustainability (Switzerland), Journal of the American Medical Informatics Association, Big Data and Society, Internet Policy Review, and Computer Law and Security Review are the most pertinent sources that have published the most research articles on data governance between 2012 and 2022. An author's dominance diagram was used to assess which students were the most prolific. The statistic also displays the authors' total number of papers written and annual citations. The study's findings demonstrate how crucial the cited authors are to the development of the field's research agenda. The five academics who will have the greatest impact on data governance between 2012 and 2022 are Carroll SR, Sammon D, Hudson M, Knoppers BM, and Mcdowall RD.

The second objective was: (2) to disclose the network between the countries and scholars relevant to the publications. To address this, the science mapping approaches were used. A research field's intellectual structure, including its underlying ideas, can be revealed through analysis. When two publications appear together in the reference list of another publication, they are linked in a co-citation network (Liu et al., 2015; Rossetto et al., 2018). Otto, b., Janssen, m., Wang, y., Zhang, Kithcin, r., Vayena, e., and Floridi, i. are the researchers who have had the largest impact on data governance research. Scholarly works frequently connect diverse clusters and receive citations from a variety of clusters. When evaluating the contributions of various nations, the collaboration diagram resulted for the countries demonstrate how the study has primarily been limited to developed nations and is mostly inexistent in other nations. As a result, academics in underdeveloped nations should be encouraged to continue their study on data governance.

Three enrichment paths based on network analysis in bibliometric research are presented, building on the core methods of bibliometric analysis: network metrics, clustering, and visualization (Donthu et al., 2021). This science mapping technique called co-word analysis and enrichment paths which were based on network analysis in bibliometric research were used to achieve the third objective. The third objective was: (3) to provide research areas that need more attention which drives data governance, and using a word cloud, the keyword co-occurrence analysis was done. Important terms including big data, data quality, data protection, data ethics, data management, electronic health records, data security, artificial intelligence, machine learning, and data ownership are included in the clusters. However, there is insufficient knowledge on the topics of data protection, data access, artificial intelligence, data management, cloud computing, block chain, and data transformation. These topics need to be more attentively studied by researchers worldwide in the future.

Future research may also consider sources like conference papers, proceedings, and books because the current study only used articles from peer-reviewed journals. Another shortcoming

of using only the Scopus database can be omitted by using other databases such as WoS (Web of Science), which might have led to avoiding critical data governance research. Furthermore, non-English publications can be added to the research database to increase the number of relevant publications.

Conclusion

The authors of the study have a broad perspective on data governance research and the bibliometric analysis methodology was adopted. The data governance studies conducted during 2012-2022 summarized the evolution and were considered for review. The main objectives were: (1) to uncover the contribution to the domain in terms of authors, sources, and countries. (2) to disclose the network between the countries and scholars relevant to the publications, and (3) to provide research areas that need more attention which drives data governance. The performance analysis techniques and the science mapping techniques were used to conduct the study. Regarding the first objective, the contribution of authors, sources, and countries was investigated. The study found the top nations that have made a substantial contribution, the most pertinent sources that have published the most research articles, and the top academics who will have the greatest impact on data governance between 2012 and 2022 to the data governance research stream. In addition, it is highlighted that developing countries are especially required to conduct more research on data governance, even though the academic community has conducted a considerable number of studies in recent years. Concerning the second objective, the study disclosed the network between the countries and scholars relevant to the publications. The collaborative network diagrams were used for the purpose and the results revealed the relationships between the countries and the scholars along with diverse clusters. Finally, the third objective was to provide research areas that need more attention which drives data governance. Accordingly, the authors found that the information on the subjects of data management, artificial intelligence, cloud computing, block chain, and data transformation is limited. Researchers from all over the world will need to pay more attention to these subjects in the future.

This study lays the foundation for academics, practitioners, and policymakers to conduct more future research and leads to advancing data governance research in regions that have a research gap. Also, the study directs the attention of scholars to new opportunities and trends. The broad spectrum of communities and the sources from which the publications were obtained capture the diversity among the researchers. The findings enhance the understanding of academics on data governance research and support finding value-adding collaborations and sources. The doctoral students and the researchers can benefit from this study by gaining knowledge of what is being researched in the literature on data governance-related research. Also, practitioners are capable of taking advantage of existing research findings by improving their knowledge and implementing the novel solutions suggested by academia. The bibliometric analysis findings conducted on data governance increase the understanding of the trends. Thus, the results of the study can be used to undertake a thorough study, such as content analysis, to gain a thorough knowledge of the underlying causes that hastened the rise in publications on data governance research.

References

- Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 49, 424–438. <https://doi.org/10.1016/j.ijinfomgt.2019.07.008>
- Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: An analysis of the literature. *Journal of Decision Systems*, 25(sup1), 64–75. <https://doi.org/10.1080/12460125.2016.1187397>
- Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2019). A systematic literature review of data governance and cloud data governance. *Personal and Ubiquitous Computing*, 23(5–6), 839–859. <https://doi.org/10.1007/s00779-017-1104-3>
- Andrić, B., Priyashantha, K. G., & De Alwis, A. C. (2023). Employee Engagement Management in the COVID-19 Pandemic: A Systematic Literature Review. *Sustainability*, 15(2), 987. <https://doi.org/10.3390/su15020987>
- Aparicio, G., Iturralde, T., & Maseda, A. (2019). Conceptual structure and perspectives on entrepreneurship education research: A bibliometric review. *European Research on Management and Business Economics*, 25(3), 105–113. <https://doi.org/10.1016/j.iemeen.2019.04.003>
- Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data Governance as a Collective Action Problem. *Information Systems Frontiers*, 22(2), 299–313. <https://doi.org/10.1007/s10796-019-09923-z>
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, 5(1), 146–166. <https://doi.org/10.1016/j.joi.2010.10.002>
- Corbet, S., Dowling, M., Gao, X., Huang, S., Lucey, B., & Vigne, S. A. (2019). An analysis of the intellectual structure of research on the financial economics of precious metals. *Resources Policy*, 63, 101–416. <https://doi.org/10.1016/j.resourpol.2019.101416>
- Cuno, S., Bruns, L., Tcholtchev, N., Lämmel, P., & Schieferdecker, I. (2019). Data Governance and Sovereignty in Urban Data Spaces Based on Standardized ICT Reference Architectures. *Data*, 4(1), 16. <https://doi.org/10.3390/data4010016>
- Demiroz, F., & Haase, T. W. (2019). The concept of resilience: A bibliometric analysis of the emergency and disaster management literature. *Local Government Studies*, 45(3), 308–327. <https://doi.org/10.1080/03003930.2018.1541796>
- Dogra, P., & Priyashantha, K. G. (2023). Review of work from home empirical research during Covid-19. *Asia Pacific Management Review*, S1029313223000283. <https://doi.org/10.1016/j.apmr.2023.04.003>

- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Esfahani, H., Tavasoli, K., & Jabbarzadeh, A. (2019). Big data and social media: A scientometrics analysis. *International Journal of Data and Network Science*, 145–164. <https://doi.org/10.5267/j.ijdns.2019.2.007>
- Gupta, N., Blair, S., & Nicholas, R. (2020). What We See, What We Don't See: Data Governance, Archaeological Spatial Databases and the Rights of Indigenous Peoples in an Age of Big Data. *Journal of Field Archaeology*, *45*(sup1), S39–S50. <https://doi.org/10.1080/00934690.2020.1713969>
- Hjørland, B. (2013). Facet analysis: The logical approach to knowledge organization. *Information Processing & Management*, *49*(2), 545–557. <https://doi.org/10.1016/j.ipm.2012.10.001>
- Huang, J.-H., Duan, X.-Y., He, F.-F., Wang, G.-J., & Hu, X.-Y. (2021). A historical review and Bibliometric analysis of research on Weak measurement research over the past decades based on Biblioshiny (arXiv:2108.11375). arXiv. <http://arxiv.org/abs/2108.11375>
- Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, *37*(3), 101493. <https://doi.org/10.1016/j.giq.2020.101493>
- Koltay, T. (2016). Data governance, data literacy and the management of data quality. *IFLA Journal*, *42*(4), 303–312. <https://doi.org/10.1177/0340035216672238>
- Kroll, J. A. (2018). Data Science Data Governance [AI Ethics]. *IEEE Security & Privacy*, *16*(6), 61–70. <https://doi.org/10.1109/MSEC.2018.2875329>
- Kumar, S., Sureka, R., Lim, W. M., Kumar Mangla, S., & Goyal, N. (2021). What do we know about business strategy and environmental research? Insights from *Business Strategy and the Environment*. *Business Strategy and the Environment*, *30*(8), 3454–3469. <https://doi.org/10.1002/bse.2813>
- Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy*, *44*(6), 101976. <https://doi.org/10.1016/j.telpol.2020.101976>
- Liu, Z., Yin, Y., Liu, W., & Dunford, M. (2015). Visualizing the intellectual structure and evolution of innovation systems research: A bibliometric analysis. *Scientometrics*, *103*(1), 135–158. <https://doi.org/10.1007/s11192-014-1517-y>
- Mobin, M. A., Mahi, M., Hassan, M. K., Habib, M., Akter, S., & Hassan, T. (2021). An analysis of COVID-19 and WHO global research roadmap: Knowledge mapping and future research agenda. *Eurasian Economic Review*. <https://doi.org/10.1007/s40822-021-00193-2>

- Mostafa, M. M. (2020). A knowledge domain visualization review of thirty years of halal food research: Themes, trends and knowledge structure. *Trends in Food Science & Technology*, 99, 660–677. <https://doi.org/10.1016/j.tifs.2020.03.022>
- Neff, M. W., & Corley, E. A. (2009). 35 years and 160,000 articles: A bibliometric exploration of the evolution of ecology. *Scientometrics*, 80(3), 657–682. <https://doi.org/10.1007/s11192-008-2099-3>
- Oudat, S., Qeshta, M., & Ali, B. J. A. (2021). Financial Performance and Audit Committee Characteristics: An Empirical Study on Bahrain Services Sector. *Journal of Contemporary Issues in Business and Government*, 27(02). <https://doi.org/10.47750/cibg.2021.27.02.453>
- Pazmiño-Sarango, M., Naranjo-Zolotov, M., & Cruz-Jesus, F. (2022). Assessing the drivers of the regional digital divide and their impact on eGovernment services: Evidence from a South American country. *Information Technology & People*, 35(7), 2002–2025. <https://doi.org/10.1108/ITP-09-2020-0628>
- Priyashantha, K. G. (2023). Disruptive technologies for human resource management: A conceptual framework development and research agenda. *Journal of Work-Applied Management*, 15(1), 21–36. <https://doi.org/10.1108/JWAM-10-2022-0069>
- Priyashantha, K. G., Dahanayake, W. E., & Maduwanthi, M. N. (2022). Career indecision: A systematic literature review. *Journal of Humanities and Applied Social Sciences*. <https://doi.org/10.1108/JHASS-06-2022-0083>
- Priyashantha, K. G., De Alwis, A. C., & Welmilla, I. (2022). Disruptive human resource management technologies: A systematic literature review. *European Journal of Management and Business Economics*. <https://doi.org/10.1108/EJMBE-01-2022-0018>
- Ramos-Rodríguez, A.-R., & Ruíz-Navarro, J. (2004). Changes in the intellectual structure of strategic management research: A bibliometric study of the Strategic Management Journal, 1980–2000. *Strategic Management Journal*, 25(10), 981–1004. <https://doi.org/10.1002/smj.397>
- Rejeb, A., Rejeb, K., Abdollahi, A., & Treiblmaier, H. (2022). The big picture on Instagram research: Insights from a bibliometric analysis. *Telematics and Informatics*, 73, 101–876. <https://doi.org/10.1016/j.tele.2022.101876>
- Rossetto, D. E., Bernardes, R. C., Borini, F. M., & Gattaz, C. C. (2018). Structure and evolution of innovation research in the last 60 years: Review and future trends in the field of business through the citations and co-citations analysis. *Scientometrics*, 115(3), 1329–1363. <https://doi.org/10.1007/s11192-018-2709-7>
- Soundararajan, K., Ho, H. K., & Su, B. (2014). Sankey diagram framework for energy and exergy flows. *Applied Energy*, 136, 1035–1042. <https://doi.org/10.1016/j.apenergy.2014.08.070>

- van Eck, N. J., & Waltman, L. (2014). Visualizing Bibliometric Networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring Scholarly Impact*. Springer International Publishing, 285–320. https://doi.org/10.1007/978-3-319-10377-8_13
- Were, V., & Moturi, C. (2017). Toward a data governance model for the Kenya health professional regulatory authorities. *The TQM Journal*, 29(4), 579–589. <https://doi.org/10.1108/TQM-10-2016-0092>