JLSC

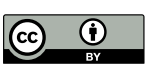
ISSN 2162-3309 | JLSC is published by the Pacific University Libraries | http://jlsc-pub.org

Volume 7, General Issue (2019)

Mapping the Scholarly Literature Found in Scopus on Research Data Management: A Bibliometric and Data Visualization Approach

Li Zhang & Nickoal Eichmann-Kalwara

Zhang, L. & Eichman Kalwara, N. (2019). Mapping the Scholarly Literature Found in Scopus on Research Data Management: A Bibliometric and Data Visualization Approach. *Journal of Librarianship and Scholarly Communication, 7*(General Issue), eP2226. https://doi.org/10.7710/2162-3309.2226



ISSN 2162-3309 10.7710/2162-3309.2226

RESEARCH

Mapping the Scholarly Literature Found in Scopus on Research Data Management: A Bibliometric and Data Visualization Approach

Li Zhang
Science & Engineering Librarian, Mississippi State University
Nickoal Eichmann-Kalwara
Digital Scholarship Librarian, University of Colorado Boulder

INTRODUCTION Since the 2000s, interest in research data management (RDM) has grown considerably. As a result, a large body of literature has discussed a broad variety of aspects related to data management. But, few studies have examined and also interpreted from visual perception the intellectual structure and progressive development of the existing literature on RDM. **METHODS** Guided by five research questions, this study employed bibliometric techniques and a visualization tool (CiteSpace) to identify and analyze the patterns of the scholarly publications about RDM. **RESULTS** Through CiteSpace's modeling and computing, the knowledge (or network) structures, significant studies, notable topics, and development trends in the literature of RDM were revealed. **DISCUSSION** The majority of the literature pertinent to RDM was published after 2002. Major research clusters within this interdisciplinary field include "scientific collaboration," "research support service," and "data literacy," while the "scientific collaboration" research cluster was the most active. Topics such as "digital curation" and "information processing" appeared most frequently in the RDM literature. Additionally, there was a sharp increase in several specific topics, such as "digital library," "big data," and "data sharing." **CONCLUSION** By looking into the "profile" of the literature on RDM, in terms of knowledge structure, evolving trends, and important topics in the domain, this work will add new information to current discussions about RDM, new service development, and future research focuses in this area.

Received: 06/05/2018 Accepted: 02/15/2019

Correspondence: Li Zhang, 395 Hardy Rd., Mississippi State, MS, 39762, lzhang@library.msstate.edu

IMPLICATIONS FOR PRACTICE

1. Addressing challenges in managing research data requires increasingly interdisciplinary collaboration. Academic libraries need to respond to the developments and identify new opportunities to facilitate the process of data management.
2. Exploration of relevant literature and keeping pace with the dynamic area of study will contribute to a better understanding of RDM.
3. Data literacy needs to gain increasing attention in scholarly literature, especially as librarians and others focus on statistical education and data visualization training.

INTRODUCTION

In recent years, management of research data has received notable attention in a wide array of disciplines. This situation is associated with several factors, of which the following two are probably the most prominent. First, scholarly research is becoming more data driven, and dealing with a vast amount of complex data poses challenges to researchers in analysis, storage, and many other areas. Second, new policies are in place. Researchers are required by funding agencies and major publishers to prepare data management plans (DMPs) and make their data/research results publicly accessible to improve transparency in the research and increase reproducibility. Big data and research data management (RDM) have been discussed extensively over the last several years, and the stakeholders involved are all striving to better understand this relatively new field. Being closely associated with information access, management, and dissemination, academic libraries are also actively exploring opportunities to play a role in the RDM landscape.

Not surprisingly, the new demands and issues connected with RDM have stimulated tremendous research interest in the topic. Scholars and practitioners have been sharing their findings, practices, and ideas in the literature across various disciplines. For example, a large-scale international survey examined scientists' data use patterns and their perceptions about data sharing and reuse (Tenopir et al., 2011). To understand data sharing practices in the social sciences, Gherghina and Katsanidou (2013) analyzed political science journals to uncover journal policies pertinent to data access and availability. Molloy (2014) conducted interviews among performing arts practitioners to learn artists' knowledge about digital curation and their preservation activities. A group of biomedical researchers analyzed general topics that should be covered in RDM plans (Williams, Bagwell, & Nahm Zozus, 2017). Likewise, a great deal of library and information science literature has discussed academic libraries' participation and engagement in RDM.

Although a broad spectrum of literature has been published on RDM, very few studies have probed and interpreted from visual perception the intellectual structure and progressive development of the existing literature. As part of our continuing research, the current study uses bibliometric methods to investigate the "profile" of the scholarly literature concerning RDM. Employing a citation analysis and visualization tool, this study seeks to characterize the complex and enormous bibliographic information in a more intuitive and efficient way. By gaining insights into the shape of the literature on RDM, in terms of knowledge structure, evolving trends, and important topics in the domain, we hope this work will add new information to current discussions about RDM, new service development, and future research focuses in this field.

LITERATURE REVIEW

Since the last decade, RDM has stimulated an extensive discussion worldwide and pointed to a great need for new research directions (Cox, Kennan, Lyon, & Pinfield, 2017). The demand requires a better understanding of how existing research literature is organized in the RDM domain. In this preliminary study, we investigate and identify essential themes and dynamic aspects of the publications on the subject.

Bibliometric analysis is a research method applied in many subject fields. Alan Pritchard (1969) first introduced the word *bibliometrics*, defining it as "the application of mathematical and statistical methods to books and other media of communication" (p. 348). This quantitative technique is usually used to map scholarly literature, revealing patterns and trends. For example, White and McCain (1998) performed a bibliometric analysis of articles published in 12 journals in the field of information science and obtained interesting findings. Among them included the revealing of "the specialty structure of the discipline over 24 years" (p. 327) and the observation of fundamental changes in information science during those years. Kraus, Filser, Eggers, Hills, and Hultman (2012) conducted a citation and co-citation analysis to uncover the overall structure and development in the research of entrepreneurial marketing. Rodrigues, van Eck, Waltman, and Jansen (2014) combined the techniques of bibliometrics, text mining, and information visualization to show the architecture of the literature on patient safety in order to gain a macro-level view on this topic, because the traditional literature review approach could not provide sufficient understanding. Employing bibliometric keyword network analysis, Dotsika and Watkins (2017) examined articles on seven innovative technologies (3D printing, Bitcoin, Social media, Big data, Internet of things, Cloud, and MOOCs) to find structural and temporal developments within publications on these topics, and to predict potentially growing trends.

A variety of bibliometric indicators exist for assessing and tracking changes in scholarly communications. In their article discussing evaluation of scientific publications, Durieux and Gevenois (2010) summarized three types of indicators: quantity, quality, and structural. While the first two indicators are for measuring research productivity and performance, structural indicators are used to determine connections between different areas of research. In addition to gauging research and discovering historical developments in a field, bibliometric methods can reveal research gaps in the literature as well (Verbeek, Debackere, Luwel, & Zimmermann, 2002). By counting citation numbers or word frequencies, bibliometric analyses also reveal hot topics in a research field (Small, Boyack, & Klavans, 2014). Furthermore, bibliometrics involve visualizing a subject field. Börner, Chen, and Boyack (2003) point out that information visualizations can give "overviews about general patterns and trends" and uncover "relations otherwise not noticed" (p. 209). This approach of interpreting information contrasts with the individual text reviewing approach in that it analyzes content at a macro level. As a form of macroscopic or "distant" reading, data information visualization allows for presenting data, on either a large or small scale, quickly and easily using illustrative formats and engaging methods (Kirk, 2012). Thanks to technology, computational tools enable broader investigation of formal and informal intellectual networks and beyond, such as exploring the vast body of published scholarly literature. Of the bibliometric studies mentioned previously, all except one have used visualization tools to aid their analyses.

Information professionals aim to discover the development of knowledge. Accomplishing this involves the study of "scholarly communities and networks, the growth and evolution of fields, the pervasion of research topics, authors, etc." (Börner et al., 2003, p. 180). Chen, Ibekwe-SanJuan, and Hou (2010) claim that a network consisting of various clusters exhibits the intellectual structure of a knowledge domain. Therefore, cluster analysis (i.e., determining subject categories) is one method for visualizing knowledge domains.

METHODS

By analyzing bibliographic records, this study attempts to explore and map current scholarly communications in the field of RDM. The questions guiding this study include the following:

1. What are the main areas of research explored and reported in the literature with regard to managing research data?
2. Are there any connections between these main areas? If so, which articles connect them?

3. Which articles are highly cited or important turning points in the study of research data management?
4. What are the hot topics or emerging trends in the data management domain?
5. What journals actively publish RDM-related articles?

The data for this present analysis was generated from Elsevier Scopus, a citation database of journal articles, books, and conference proceedings. Scopus was selected because of its interdisciplinary feature, as we are looking at RDM across multiple fields of study. RDM covers a great deal of topics and subtopics, such as data architecture, data security, data documentation, metadata, metadata schemas, data sharing, data access, and workflow. In this early exploratory study, we used simple synonyms of the phrase *research data management* and the words most frequently associated with "research data management" to capture articles on RDM. The following search queries were entered:

(TITLE-ABS-KEY ("research data manag*") OR TITLE-ABS-KEY ("responsible data manag*") OR TITLE-ABS-KEY ("data lifecycle manag*") OR TITLE-ABS-KEY ("data resource manag*") OR TITLE-ABS-KEY ("research data admin*") OR TITLE-ABS-KEY ("digital curat*") OR TITLE-ABS-KEY ("digital data manag*") OR TITLE-ABS-KEY ("data steward*") OR TITLE-ABS-KEY ("data curat*") OR TITLE-ABS-KEY ("research reposit*") OR TITLE-ABS-KEY ("data management plan*"))

The broader term *data management* was excluded from the search queries, because an initial test search in Scopus retrieved an extraordinary number of data management–related articles that were about programming, data mining, machine learning, and implementing databases. From the above searches, publications mentioning any of the search phrases in their titles, abstracts, or keyword fields were considered relevant and collected for further examination. There were 1,913 relevant documents from all types of materials (journals, conference proceedings, books, reviews, editorials, etc.), covering widespread subject fields. The searches did not specify any range of publication years. All records attached to the documents, including bibliographic information, citation information, and so forth, were then exported from Scopus in RIS format and saved in a folder on a local computer for visualization. In total, the 1,913 documents contained 23,402 cited references.

Several visualization tools are available for analyzing bibliographic data and generating citation networks, for example, BibExcel, CiteSpace, Sci2, and VOSViewer. This study used CiteSpace to read the pertinent documents and their cited references, because it offers both graph-based and timeline-based visualizations. As Chen (2004) explains, CiteSpace is a Java application that combines bibliometrics and network visualization to discern trends

and patterns in progressive knowledge domains. A main function of this tool is to conduct document co-citation analysis for extraction of subject clusters in citation data. (Co-citation analysis measures the frequency of *jointly* cited documents, providing assessment on document similarity).

To run this software, Java (JRE) 8 was downloaded as instructed. Within CiteSpace, the exported RIS file from Scopus was converted to Clarivate Web of Science format (WoS), which is the required file format. The conversion rate this study obtained was 91%.

RESULTS

Data analysis identified general characteristics of the existing research on RDM. Through CiteSpace's modeling, knowledge/network structures, significant studies, salient topics, and development trends in the literature of RDM were computationally detected.

Publication Distribution

The retrieved documents related to RDM were published between 1945 and 2018 (one article from 2018 was indexed in Scopus at the time the search was performed). A careful review of the "1945" article found that the correct publication date should be 1980. Thus, the earliest item was an editorial article from 1962, mentioning a growing need for journals to serve as repositories of experimental findings. As shown in Figure 1, publications on RDM appeared sporadically before 2000 and gradually increased until 2006. Since 2007, the number of research papers on the topic has shown exponential growth. Overall, the majority (96%) of the documents relevant to RDM were published after 2002.

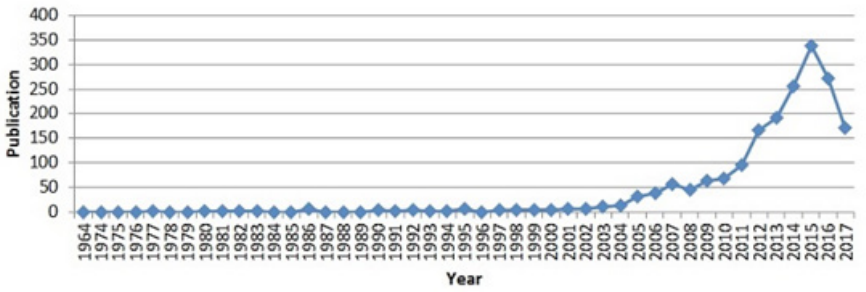


Figure 1. Distribution of annual publications on research data management

Major Research Areas on RDM

Research areas (clusters) that focus on various aspects of RDM were found through analyzing the bibliographic records. When placing parameters inside CiteSpace, the study set the time frame of publications between 2000 and 2017 and selected the top 100 most-cited articles in each of the years as samples for analysis. The tool discovered 130 clusters on RDM. Figure 2 displays the seven largest interconnected clusters, which account for 51% of the entire network. The nodes and links represent cited references and co-citation relationships. Cluster labels depict the general theme of a cluster; they are extracted in CiteSpace from noun phrases in article titles. As Figure 2 shows, the top three clusters in the literature network about RDM include: #0 scientific collaboration, #1 research support service, and #2 data literacy.

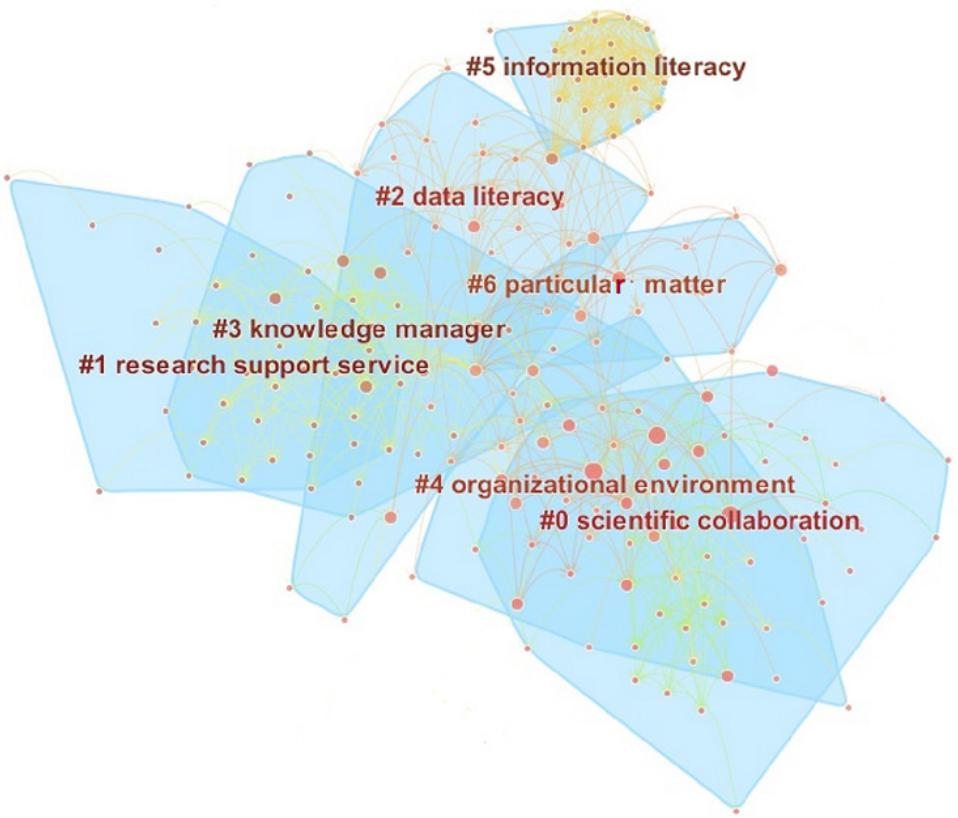


Figure 2. Main research clusters

The average publication year of the documents in the "scientific collaboration" cluster is 2009, and the median is the same. Some of the main topics covered in this cluster include data management requirement, data sharing, ongoing gap, technology collaboration, knowledge infrastructure, future priorities, institutional issues, shared repositories, and interdisciplinary approach.

While the median publication year of the documents in the "research support service" cluster is 2011, the mean year is 2010. Documents in this cluster are related to workflow service, academic library, bibliometrics, data science, digital curation, building professional development opportunities, data curator, workforce development, library collaboration, and so forth.

Compared to those in the "scientific collaboration" and "research support service" clusters, publications in the "data literacy" cluster are a little newer. The average publication year is 2012 and the median 2013. Key topics, such as data quality, sequence analysis, cluster analysis, personal information management, data librarian, team-based data management instruction, and analytic tools, stay at the top of this category. Research themes about the main clusters are summarized in Table 1.

Cluster #	Mean Year	Research Theme
0	2009	Scientific collaboration
1	2010	Research support service
2	2012	Data literacy
3	2009	Knowledge manager/Digital curation
4	2011	Institutional support
5	2010	Information literacy/Data life cycle management
6	2012	Data service

Table 1. Characteristics of major research clusters

Development of the Clusters over Time

To examine the major research clusters from a chronological perspective and to identify durations of these research areas, we performed a timeline visualization (Figure 3). The main

research clusters are listed on the right of the figure. The publication years are shown on the top. The colored arcs indicate co-citation links. The tree rings (circles) represent the citation histories of cited documents. The size of tree rings correlates to the frequency of citations. The red dots imply a significant citation burst, which means citations to that document increased rapidly in a given time period.

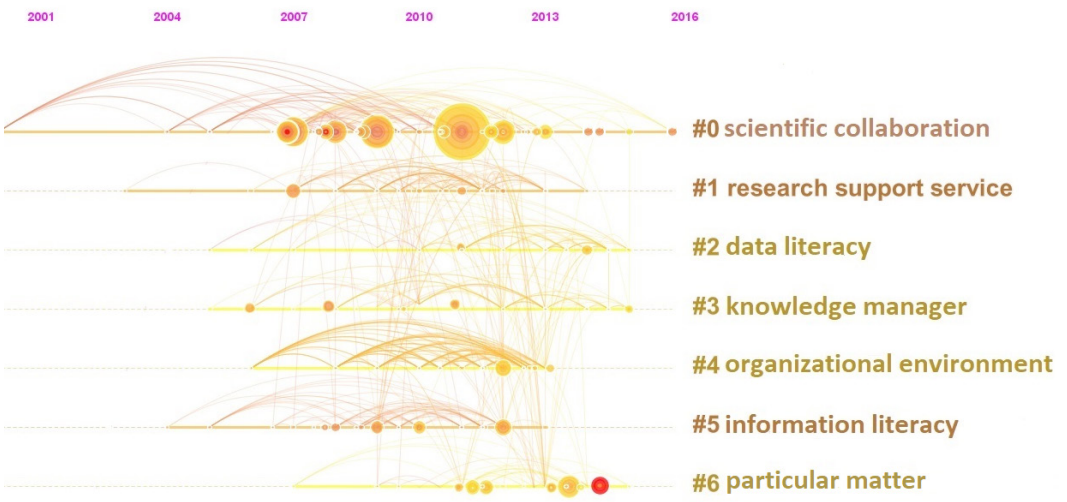


Figure 3. Timeline of major research clusters

Of the network clusters, the "scientific collaboration" cluster is the most active. Scholarly works in this cluster span a period of almost 20 years, from 2000 to 2016. From 2000 to 2006, few articles related to RDM generated a large number of citations or significant interest. Arzberger et al. 2004 was one of the earliest survey studies that explored the benefits and barriers of making publicly funded research data openly available. For the period of 2007 to 2016, Figure 3 displays many tree rings and red dots (i.e., citation bursts), indicating that a wealth of high-impact articles were published during this phase.

Similarly, in the "research support service" cluster, peak research activity occurred between 2007 and 2014. Although no citation bursts were seen in this branch, an earlier article by Gold (2007) that explored cyberinfrastructure and the roles of libraries and librarians attracted a noticeable amount of attention in the library community. Gold's article corresponded to several notable works in the "knowledge manager/digital curation" cluster: for example, a study on preserving data so that data can be discovered, shared, and reused in the long term (Witt, 2008) and an article articulating the opportunities and potentials for librarians to serve as data curation managers (Lyon, 2012).

Active publications in the "data literacy" cluster appeared between 2005 and 2015. Compared to other clusters, topics covered in this cluster were rather diversified, ranging from data quality to personal information management to sequence analysis to team-based data management instruction, but all associated with building knowledge about analyzing and managing data. This cluster is related to Cluster #5, "information literacy."

For the clusters focusing on "institutional support/organizational environment" and "data service/particular matter," while a larger number of publications related to institutional support was seen during 2005 and 2013, those related to data service were primarily published after 2010.

Hot Topics and Emerging Trends

In contrast to the cited references approach discussed in previous sections, the study then used keywords provided by authors to detect the hot topics associated with RDM. Inside CiteSpace, the top 50 most frequently appearing keywords from each year between 2000 and 2017 were selected for analysis. In the visualized network (Figure 4), the largest node (keyword) is "data curation," indicating it has the highest frequency of appearance in the publications on RDM. "Information processing" and "information management" are the second and third largest nodes. In addition, nodes (i.e., keywords) such as "digital library," "big data," "data sharing," "metadata," "data acquisitions," and "data preservation" are displayed with red dots, revealing that they are hot spots with high-burst values (A burst indicates an abrupt rise in the volume of occurrence). The bursts indicate that those keywords were the fast growing topics in the articles during the studied time period. Among the above hot-spot nodes, "digital library" has the longest burst history, lasting from 2006 until 2013. The apparent drop-off since 2013 points to the evolution of "digital library" serving as an umbrella term for digital content (including research data), processes, and repositories, to specifically digital collections, redefining it in the context of growing issues in data curation (see Figure 4).

Journal Network on RDM

Examining what journals have been frequently cited may inform what disciplines that are actively involved in RDM. Figure 5 displays the top 20 most frequently co-cited journals based on the present study's bibliographic data. The text size correlates to the frequency of citations (see Figure 5).

Figure 4. Keyword frequencies

Figure 5. Journal Network on RDM

Table 2 lists the 20 journals by number of citation count; their subject categories are also provided.

Rank	Frequency	Journal Title	LC Subject Classification
1	164	*PLoS One*	Science
2	151	*Nucleic Acids Research*	Physiology
3	149	*Nature*	Science
4	130	*Bioinformatics*	Biology
5	129	*Science*	Science
6	99	*International Journal of Digital Curation*	Library Science
7	92	*Library Trends*	Library Science
8	90	*D-Lib Magazine*	Information Resources
9	78	*BMC Bioinformatics*	Biology
10	70	*Nature Genetics*	Biology
11	64	*Journal of the American Society for Information Science and Technology*	Library Science
12	55	*Communications of the ACM*	Mathematics and Computer Science
13	44	*Nature Biotechnology*	Chemical Technology, Microbiology
14	42	*Genome Research*	Physiology
15	41	*Journal of Library Administration*	Library Science
16	39	*Genome Biology*	Biology
17	36	*Library Hi Tech*	Library Science
18	35	*Data Science Journal*	Science
19	32	*International Journal of Digital Libraries*	Library Science
20	31	*Journal of the American Medical Informatics Association*	Medicine

Table 2. Most frequently cited journals

DISCUSSION

Research on RDM is relatively new. Of the articles that this study examined, 90% were published after 2007. Since then, an exponential growth of publications has been seen, which reflects the increased importance of study in this area.

Among various research topics in the RDM literature network, the top three major research areas (or clusters) are "scientific collaboration," "research support service," and "data lit-

eracy." By checking the metrics (such as citation burst, betweenness centrality, and citation frequency) that CiteSpace uses for citation analysis, important and remarkable publications are identified, which can help researchers track the development or paths of transformative changes in the RDM knowledge domain. (A citation burst indicates a sudden rise in the volume of citations to an article/author, which indicates that the article is of particular importance. Betweenness centrality identifies the ability of a publication to connect to other publications, which is another indicator showing the importance of a node/publication in a network.)

A study on requirements for data in digital libraries by Borgman, Wallis, Mayernik, and Pepe (2007a) showed a strong citation burst and high betweenness centrality. The paper not only stimulated numerous citing articles in its own cluster (scientific collaboration), but also connected to many publications in the "institutional support/organizational environment" cluster. This 2007 contribution thus marks a change or progression in the research about RDM, leading to further scholarly discussions about building sustainable information infrastructures. Some additional publications in the "scientific collaboration" cluster are worth mentioning, because they have focused on different aspects related to RDM and also attracted numerous citations. These works investigated collaborative efforts among scientists and engineers on data practices (Borgman, Wallis, & Enyedy, 2007b), the infrastructures for organizing large volumes of data (Lynch, 2008), a new paradigm (eScience) for scientific exploration (Hey, Tansley, & Tollie, 2009), rationales and challenges for sharing data (Borgman, 2012), trends in data sharing among scientists (Tenopir et al., 2011), the relationship between data publications and citation impact (Piwowar, Day, & Fridsma, 2007), role of metadata (Edwards, Mayernik, Batcheller, Geoffrey, & Borgman, 2011), and issues related to data curation (Heidorn, 2008). During 2014 and 2016, several publications presenting recent developments in the RDM area triggered a relatively high number of citations. For example, one is a discussion about research centers working together to provide digital services in a more comprehensive and cohesive way (Towns et al., 2014). By providing many case studies, a book explored various specific aspects of RDM (Ray, 2014). A more recent article evaluated Figshare, a repository service for sharing academic resources (Thelwall and Kousha, 2016).

Articles in the "data literacy" cluster are quite new, with the average publication year being around 2012. Although no citation bursts were detected in this cluster, two works showed high betweenness centrality, which means that they connected different research clusters and served as an intermediary in the communication over data literacy. One was a collection of works providing practical guides and advice to RDM (Pryor, 2012). The other was an article advocating for the development of a data information literacy curriculum in collaboration with disciplinary faculty (Carlson, Fosmire, Miller, & Sapp Nelson, 2011).

The lack of citation burst in the data literacy publications implies that limited articles are researching and referencing studies in this area. As a form of information literacy, data literacy needs to gain increasing attention in the scholarly literature, especially it is considered as an essential competency in an age of big data and information deluge.

Likewise, this study also found that articles related to data service were primarily published after 2010. As a fairly new support system, data service was developed to meet research demands and help promote research activities of scholars who use and produce data. In the "data service" research cluster that this study identified, important contributions include a large-sample survey conducted in Australia, New Zealand, Ireland, and the United Kingdom pertaining to the roles that libraries and librarians can play to carry out research data management services (Corrall, Kennan, & Afzal, 2013); a report on US and Canadian libraries' practices in the implementation of research data services, noting a gradual expansion of traditional information retrieval service (for example, locating data or repositories) to more technology-focused approaches, such as creating metadata and archiving data (Tenopir, Sandusky, Allard, & Birch, 2014); and a survey among UK academic libraries about their involvement in RDM, which showed that only large research institutions were offering limited RDM services (Cox & Pinfield, 2014).

Apart from citations, tracking topics will aid in understanding emerging trends in the research of RDM. With regard to hot topics related to RDM, data curation and information processing/management have been heavily studied in the literature. Data curation has the highest frequency of appearance in the publications. This scholarly attention of data curation echoes what Flanders and Muñoz's (n.d.) finding that there was increased effort in libraries to actively and continually capture and preserve research data. On the other hand, the study's finding also suggests that probably more research should be conducted to address those less-studied areas (marked as smaller nodes in Figure 4). In addition, examining what journals have been frequently cited provides additional insights into relevant or potential disciplines that are actively involved in RDM. This may suggest areas in which libraries can develop strategies for collaborative partnerships.

Some top multidisciplinary scientific journals, such as *Nature* and *Science*, serve as good forums for discussions concerning RDM. *PLoS ONE*, another interdisciplinary journal, has the highest citation counts (n=164). As is seen in Figure 5, researchers in biological sciences, library science, computer science, and health sciences are heavily engaged in this particular field, which is expected given recent Office of Science and Technology policies, and National Institutes of Health (NIH) and National Science Foundation (NSF) mandates for research data management plans. Since this study searched only one database (Scopus), we cannot say that research on RDM practices emphasizes STEM fields. But Akers and Doty

(2013) observed that articles on management of humanities data remain underpublished despite an increased focus in libraries and digital humanities. This study's finding suggests areas in which both subject-specialist and functional librarians can continue to develop strategies for collaborative partnerships with researchers in these areas. Additionally, librarians could collaborate with any research centers that sit outside of academic structures that are reflected in these journals' authorship. However, as Latham (2017) notes, "librarians must understand that RDM services, while increasingly important, should be but one of a suite of services offered to researchers in order to meet the needs of most, and to perpetuate research—in all its forms—throughout its lifecycle" (p. 265).

LIMITATIONS OF THE STUDY

There are several limitations of this study that are worth noting. First, the sample articles were obtained only from Scopus, one of the three English-language data sources that CiteSpace implements (The other two are Clarivate Web of Science and NCBI PubMed). Due to access restriction, Web of Science was not searched. PubMed was not searched either, because according to Elsevier, Scopus has significant duplicated coverage of the literature offered by PubMed. However, there may be additional sources in PubMed on RDM that were not included in this study. Second, the search terms that this study used are rather controlled and may not have gleaned the full relevant literature, although the selection of these words was intended to perform more focused searches for analysis of an enormous number of citation records. Future study should further refine the search strategies. Third, the data input format CiteSpace accepts is the bibliographic record style of Web of Science. Records from other databases need to undergo a conversion within the tool. While this study acquired a high conversion rate (91%), a small number of data (9%) was lost in the conversion. Fourth, one challenge to the software users is to set proper visualization parameters to optimize visualization (this study used the tool's default parameters). As the tool continues to improve, data visualization components will become more effective and helpful. This paper conducts an exploratory investigation and serves as the basis for further rounds of discussions and analyses.

CONCLUSION

The current work has conducted bibliographic and network analysis of studies on RDM to map the intellectual structure and research development related to this knowledge domain. The results provide an overview of the scholarly literature in the field. Research outputs on the topic have steadily increased since 2000, with a rapid rise during 2007 and 2017. Major research areas within this interdisciplinary field include "scientific collaboration," "research support service," and "data literacy," with 'scientific collaboration" being the most active

research cluster, containing many high-impact articles. Among them, Christine Borgman et al.'s (2007a) paper "Drowning in Data: Digital Library Architecture to Support Scientific Use of Embedded Sensor Networks," exhibits great visibility and importance, whereas Carol Tenopir et al.'s (2011) article "Data Sharing by Scientists: Practices and Perceptions," is the most cited. An analysis of the most-cited keywords reveals that "digital curation" and "information processing" are the primary general topics associated with research data management, and there is a sharp increase in the appearance of several specific topics, such as "digital library," "big data," "data sharing," "metadata," "data acquisitions," and "data preservation." The top three journals that have explored research data management and received a high number of citations are *PLoS ONE*, *Nucleic Acids Research*, and *Nature*. Disciplines such as biological sciences, library science, computer science, and health sciences are heavily engaged in the field of data management.

Research data management is a great challenge for many disciplines. Exploring relevant literature and keeping pace with the dynamic area of study will help researchers better understand this fast evolving landscape, help identify research trends and gaps, and assist in enhancing capabilities across various fields. Complementing traditional practices of literature analysis, network and data visualization approaches allow us to not only quickly find patterns in large data sets that span long periods of time, but also display the patterns in a way that can be easily understood at a glance. In transforming data into effective visual forms, librarians will be able to develop evidence-based knowledge, strengthen decision-making when performing collection development, and conduct meaningful conversations when engaged in outreach activities. The methods described in this paper could be applied to analyzing other subjects of interest.

REFERENCES

Akers, K. G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, *8*(2), 5–26. https://doi.org/10.2218/ijdc.v8i2.263

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K, Laaksonen, L., . . . Wouters, P. (2004). Promoting access to public research data for scientific, economic, and social development. *Data Science Journal, 3*, 135–153. https://doi.org/10.2481/dsj.3.135

Borgman, C. L., Wallis, J. C., Mayernik, M. S., & Pepe, A. (2007a). Drowning in data: Digital library architecture to support scientific use of embedded sensor networks. *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital libraries*, 269–277. https://doi.org/10.1145/1255175.1255228

Borgman, C. L., Wallis, J. C. & Enyedy, N. (2007b). Little science confronts the data deluge: Habitat ecology, embedded sensor networks, and digital libraries. *International Journal of Digital Libraries*, *7*(1/2), 17–30. https://doi.org/10.1007/s00799-007-0022-9

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, *63*(6), 1059–1078. http://dx.doi.org/10.1002/asi.22634

Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science & Technology*, *37*(1), 179–255. https://doi.org/10.1002/aris.1440370106

Carlson, J., Fosmire, M., Miller, C., & Sapp Nelson, M. (2011). Determining data information literacy needs: A study of students and research faculty. *Portal: Libraries and the Academy*, *11*(2), 629–657. https://doi.org/10.1353/pla.2011.0022

Chen, C. (2004). Searching for intellectual turning points*:* Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(suppl. 1), 5303–5310. https://doi.org/10.1073/pnas.0307513100

Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *Journal of the American Society for Information Science and Technology*, *61*, 1386–1409. https://doi.org/10.1002/asi.21309

Corrall, S, Kennan, M. A., & Afzal, W. (2013). Bibliometrics and research data management services: Emerging trends in library support for research. *Library Trends*, *61*(3), 636–674. https://doi.org/10.1353/lib.2013.0005

Cox, A. M., Kennan, M. A., Lyon, L., & Pinfield, S. (2017). Developments in research data management in academic libraries: Towards an understanding of research data service maturity. *Journal of the Association for Information Science and Technology*, *68*(9), 2182–2200. https://doi.org/10.1002/asi.23781

Cox, A. M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, *46*(4), 299–316. https://doi.org/10.1177/0961000613492542

Dotsika, F., & Watkins, A. (2017). Identifying potentially disruptive trends by means of keyword network analysis. *Technological Forecasting and Social Change*, *119*, 114–127. https://doi.org/10.1016/j.techfore.2017.03.020

Durieux, V. & Gevenois, P. A. (2010). Bibliometric indicators: Quality measurements of scientific publications. *Radiology*, *255*(2), 342–351. https://doi.org/10.1148/radiol.09090626

Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Geoffrey, C. B., & Borgman, C. L. (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science*, *41*(5), 667–690. https://doi.org/10.1177/0306312711413314

Flanders, J. & Muñoz, T. (Eds.) (n.d). An introduction to humanities data curation. *DH curation guide: A community resource guide to data curation in the digital humanities.* https://doi.org/10.5703/1288284315650

Gherghina, S., & Katsanidou, A. (2013). Data availability in political science journals. *European Political Science*, *12*(3), 333–349. https://doi.org/10.1057/eps.2013.8

Gold, A. (2007). Cyberinfrastructure, data, and libraries, Part 2: Libraries and the data challenge: Roles and actions for libraries. *D-Lib Magazine*, *13*(9/10). https://doi.org/10.1045/september2007-gold-pt2

Heidorn, P. B. (2008). Shedding light on the dark data in the long tail of science. *Library Trends*, *57*(2), 280–289. https://doi.org/10.1353/lib.0.0036

Hey, T., Tansley, S. & Tolle, K. (Eds.). (2009). *The fourth paradigm: Data-intensive scientific discovery*. Redmond, WA: Microsoft Research.

Kraus, S., Filser, M., Eggers, F., Hills, G. E., & Hultman, C. M. (2012). The entrepreneurial marketing domain: A citation and co-citation analysis. *Journal of Research in Marketing and Entrepreneurship*, *14*(1), 6–26.

Kirk, A. (2012). *Data visualization: A successful design process.* Birmingham, UK: Packt.

Latham, B. (2017). Research data management: Defining roles, prioritizing services, and enumerating challenges. *Journal of Academic Librarianship*, *43*(3), 263–265. https://doi.org/10.1016/j.acalib.2017.04.004

Lynch, C. (2008). Big data: How do your data grow? *Nature*, *455*, 28–29. https://doi.org/10.1038/455028a

Lyon, L. (2012). The informatics transform: Re-engineering libraries for the data decade. *International Journal of Digital Curation*, *7*(1), 126–138. https://doi.org/10.2218/ijdc.v7i1.220

Molloy, L. 2014. Digital curation skills in the performing arts: An investigation of practitioner awareness and knowledge of digital object management and preservation. *International Journal of Performance Arts and Digital Media*, *10*(1), 7–20. https://doi.org/10.1080/14794713.2014.912496

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007), Sharing detailed research data is associated with increased citation rate. *PLoS ONE 2:* e308. https://doi.org/10.1371/journal.pone.0000308

Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation*, *25*, 348–349.

Pryor, G. (Ed.). (2012). *Managing research data*. London: Facet.

Ray, J. M. (2014). *Research data management: Practical strategies for information professionals*. West Lafayette, IN: Purdue University Press.

Rodrigues, S. P., van Eck, N. J., Waltman, L., & Jansen, F. W. (2014). Mapping patient safety: A large-scale literature review using bibliometric visualisation techniques. *BMJ Open*, *4*(3),1–8. https://doi.org/10.1136/bmjopen-2013-004468

Small, H, Boyack, K.W., & Klavans, R. (2014). Identifying emerging topics in science and technology. *Research Policy*, *43*(8), 1450–1467. https://doi.org/10.1016/j.respol.2014.02.005

Thelwall, M., & Kousha, K. (2016). Figshare: A universal repository for academic resource sharing? *Online Information Review*, *40*(3), 333–346. https://doi.org/10.1108/oir-06-2015-0190

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, *6*(6). https://doi.org/10.1371/journal.pone.0021101

Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library and Information Science Research*, *36*(2), 84–90. https://doi.org/10.1016/j.lisr.2013.11.003

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A. . . . Wilkins-Diehr, N. (2014). XSEDE: Accelerating scientific discovery. *Computing in Science & Engineering*, *16*(5), 62–74. https://doi.org/10.1109/mcse.2014.80

Verbeek, A., Debackere K., Luwel M., & Zimmermann E. (2002). Measuring progress and evolution in science and technology I: The multiple uses of bibliometric indicators. *International Journal of Management Reviews*, *4*(2), 179–211. https://doi.org/10.1111/1468-2370.00083

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science and Technology*, *49*(4), 327–355. https://doi.org/10.1002/(sici)1097-4571(19980401)49:4<327::aid-asi4>3.0.co;2-w

Williams, M., Bagwell, J., & Nahm Zozus, M. (2017). Data management plans: The missing perspective. *Journal of Biomedical Informatics*, *71*, 130–142. https://doi.org/10.1016/j.jbi.2017.05.004

Witt, M. (2008). Institutional repositories and research data curation in a distributed environment. *Library Trends*, *57*(2), 191–201. https://doi.org/10.1353/lib.0.0029