Mapping the Knowledge Domains of Research Data Management: A Co-occurrence Analysis

Mohd Ikhwan Ismail[1,2], A. Abrizah[2] and **Samsul Farid Samsuddin[2]**
[1]Hamzah Sendut Library, Universiti Sains Malaysia, Penang, MALAYSIA
[2]Department of Library & Information Science,
Faculty of Arts and Social Sciences,
Universiti Malaya, Kuala Lumpur, MALAYSIA
e-mail: ikhwanismail@usm.my; abrizah@um.edu.my; samsulfarid@um.edu.my

ABSTRACT
This paper provides a knowledge presentation and mapping of research data management (RDM) based on a bibliometric analysis of the subject. The study investigates the keywords co-occurrences network and develops clusters to find the main themes from RDM publications in Scopus between 1977 and 2021, to uncover RDM development and identify the potential field of research on RDM. VOSviewer is used for constructing and visualising the data, while Microsoft Excel is used for analysing and presenting the descriptive statistics. The study focuses on searching the author and index keywords of "research data management" in the Scopus database to identify the RDM publications. The data gathered covers publication from 1977 to 2021, in which from 2012 onwards the publications produced were over ten documents per year. The study identified 442 documents with different languages and reference types. The study found four research clusters developed from 442 documents that present multiple themes derived from a specific paper related to the RDM. The main four themes identified are: RDM services, data sharing, information systems, and data management. The study also provides suggestions for potential research directions according to each RDM cluster. Findings from this paper highlights how RDM related research grows over time. This paper also contributes to the understanding of the underlying structure of RDM in addition to highlighting the important input towards the development of the RDM in scientific research.

Keywords: Mapping knowledge domains, Research data management, Information services; Data sharing; RDM services; Bibliometrics

INTRODUCTION

Research data management (RDM), which is part of the research process, has become widely discussed in many empirical works within this decade. RDM, which concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results, is important to its stakeholders which include (i) government & funders, (ii) university leadership, (iii) research management/support units, and (iv) researchers (Flores et al. 2015). There are different drivers and influencing factors on how RDM services could be created. The perspectives of RDM could be derived from different components related to strategies, policies, guidelines, processes,

technologies and services (Pinfield, Cox, and Smith 2014), and the development of RDM would be interrelated between the stakeholders, drivers and influencing factors.

There has been rapidly growing discussions on RDM in various scholarly platforms, and it has been a popular research topic in the library and information science (LIS) literature. Several studies systematically review and bibliometrically analyse RDM literature, but they are limited to certain areas. Perrier et al. (2017) conducted the scoping review on RDM in academic institutions, covering 301 articles published from 1995 to 2016. The study found that 85% of articles were published from 2010 onwards. The articles were grouped into five areas: stakeholder, data, library, tool/device, publication, and data quality. Ashiq et al. (2020) covered RDM literature published between 2016 and 2020, and the study focused on the challenges, services, skills, and factors on RDM practices by researchers and services by academic libraries. Zhang & Eichmann-Kalwara (2019) studied the RDM literature in the Scopus database, applying bibliometric analysis and data visualisation using CiteSpace. Using various keywords as search strategies, such as "research data management" and areas related to RDM such as digital curation, data steward, data curation, data management plan, the retrieved documents were published between 1945 and 2018. The study found six RDM themes: scientific collaboration, research support service, data literacy, knowledge manager, organisational environment, information literacy and particular matter. Zhang & Eichmann-Kalwara (2019), conducted a bibliometric study on RDM, and found major research clusters within this interdisciplinary field which include “scientific collaboration,” “research support service,” and “data literacy”. Additionally, there was a sharp increase in several LIS specific topics, such as “digital library”, “big data”, “data sharing” and “data curation”. However, there is still limited profiling of RDM literature in terms of knowledge structure, to gauge the trends, and future research focus.

This study investigates the keywords co-occurrences network analysis and applying knowledge domains mapping by developing clusters from RDM publications. The keywords co-occurrence network is a valuable tool for identifying research areas (Liu and Mei 2016). The mapping knowledge domains aimed to describe a newly evolving interdisciplinary area of knowledge while looking at the process of mining, analysing, sorting, enabling navigation of, and displaying knowledge (Shiffrin and Borner 2004). Mapping knowledge domains could also be called knowledge graph or knowledge visualisation as part of social network analysis (Zhu et al. 2015). This analysis could reveal that the hidden connections between the publications and the mapping knowledge domains cannot be easily interpreted if doing manually (Li et al., 2017). Therefore, this study could present the mapping knowledge domains as an approach of scientific literature on RDM by focusing on the following research question; “How can the relationships among research data management literature be described and analysed in a representative, dynamic, and scalable way?”.

MATERIALS AND METHOD

The data for this study was generated from Elsevier’s Scopus database as of April 2021. Scopus has been chosen because it one of the largest citation databases of peer-reviewed literature, with strength in inter-disciplinary feature, the nature of RDM as a research field (Zhang & Eichmann-Kalwara 2019). Scopus covered many types of research publications

with ensuring the quality of data indexed through the meticulous data selection and re-evaluation by the Scopus Content Selection and Advisory Board (Baas et al. 2020). According to Baas et al. (2020) and Martín-Martín et al. (2021), the Scopus database has more than 27 million publication records (1966-2004), and it has grown up to over 76 million publication records which covering publications from 1788 until 2019. The Scopus database contained different publications covering the global coverage of journals, conference proceedings, books, and others (Baas et al. 2020). Therefore, the research community has recognised the Scopus database as the main relevant database, providing a comprehensive overview of the research outputs (Wahid, Ahmi, and Alam 2020).

Social network analysis, a technique in identifying the clusters of related nodes within the network (Benckendorff and Zehrer 2013) was used as the research approach. This network was analysed using cluster analysis, centrality, betweenness, and any relevant metrics; when all analysis combined, they contributed the overview of the knowledge domains. This analysis could permit the analysis of the relationship and collaboration development in the research area.

The keyword "research data management" contained in the title, abstract, and keywords fields were used to search the relevant publications in the Scopus database. The exact keyword "research data management" in the keyword searching was used as an additional query to retrieve specific publications focusing on RDM. The query search keyword in the Scopus database covers all publication years, languages, source types, and document types. The search was conducted on 28th April 2021. Figure 1 presents the literature search flow used in this study. All the publications found were analysed using two application tools; (i) VOSviewer version 1.6.16 to generate and visualise the bibliometric networks and mapping analysis such as keywords and citation co-occurrence networks; (ii) Microsoft Excel to retrieve the frequencies and percentage of publications and to develop graphs/tables.

Data were retrieved straightly from the Scopus database using two types of export data; (i) RIS and (ii) CSV. The data was analysed from these two types of export data according to applications such as RIS for VOSviewer and CSV for Microsoft Excel. This study developed the clusters from the 442 RDM publications retrieved using the mapping knowledge domains analysis method.

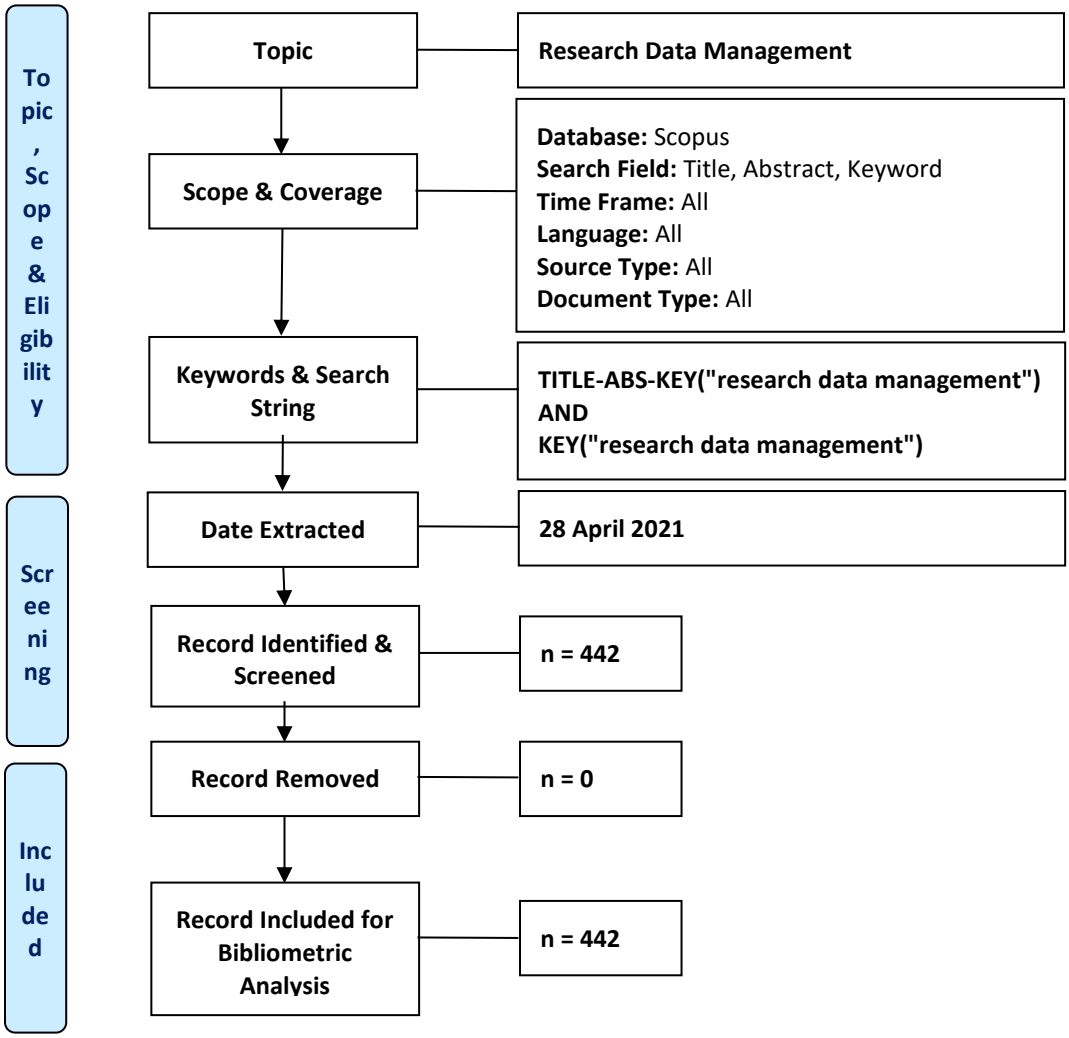

Figure 1: Literature Search Flow for this Study

RESULTS

RDM Topics

A total number of 442 publications were identified in the Scopus database based on the keyword "research data management". A total of 160 keywords were identified, which consist of both author keywords and index keywords. All these keywords have a total of 2552 times appearing in 442 publications. The study found issues for similar keywords but differ by the singular and plural terms such as "Research Data Management" and "Research Data Managements", where for proper noun the singular term is preferred, and plural term is used for the common noun. Some keywords also have the same meaning or connotation but in different terms or spelling, such as "Fair" and "Fair Principles". Therefore, this study combined some keywords, which become a unique keyword using the thesaurus method in VOSviewer to get accurate results. As a result, the study found 145 keywords after the cleaning process and used them as descriptors to identify the subject of the publication. Table 1 shows the top 20 keywords from the total keywords in 442 publications. Each publication may have one or more than one keyword

to it. The most appeared keyword is obviously the general term "Research Data Management", "Information Management" and "Research Data'. "Libraries" appears 4th in the list of top keywords used. Other keywords used are associated with concepts and areas in the management and delivery of digital content – "Metadata", "Data Repository", "Data Sharing", "Data Curation", "Digital Storage" and "Information Services". RDM is also associated with openness and transparency, reflected through the keywords used such as "Open Science", "Open Data", "Open Access" and "Reproducibility".

Table 1: Top 20 keywords from 145 keywords in 442 publications

	Keywords	Total number of appearing in publications
1	Research Data Management	576
2	Information Management	230
3	Research Data	85
4	Libraries	79
5	Metadata	59
6	Open Science	47
7	Data Repository	45
8	Data Sharing	41
9	Data Curation	39
10	Open Data	33
11	Digital Libraries	32
12	Digital Storage	27
13	Information Systems	27
14	Information Services	26
15	Open Access	26
16	Big Data	25
17	Semantics	22
18	Reproducibility	22
19	FAIR Principles	21
20	Research	21

** The total number of appearing in publications for each keyword could be more than the actual number after merging some keywords.*

RDM Research Areas

Research area or clusters based on the 145 keywords found from RDM publications were identified. The cluster had been developed with the 10 minimum number of occurrences of keywords and excluded the keywords with low occurrences. Based on this threshold, 57 keywords were selected from this process and visualized based on their respective clusters using VOSviewer. Figure 2 shows the network visualisation of the keywords co-occurrence network in which the circle size, font size, colour, and the thickness of the connecting lines indicate the link strength of the relationship between keywords. The network indicates that the related keywords by the same colour as commonly listed together. Four clusters emerged from the keywords co-occurrences network, namely

"RDM services" (Cluster 1, red, 16 keywords); "Data sharing" (Cluster 2, dark green, 15 keywords); "Information systems" (Cluster 3, blue, 14 keywords) and "Data management" (Cluster 4, light green, 12 keywords).

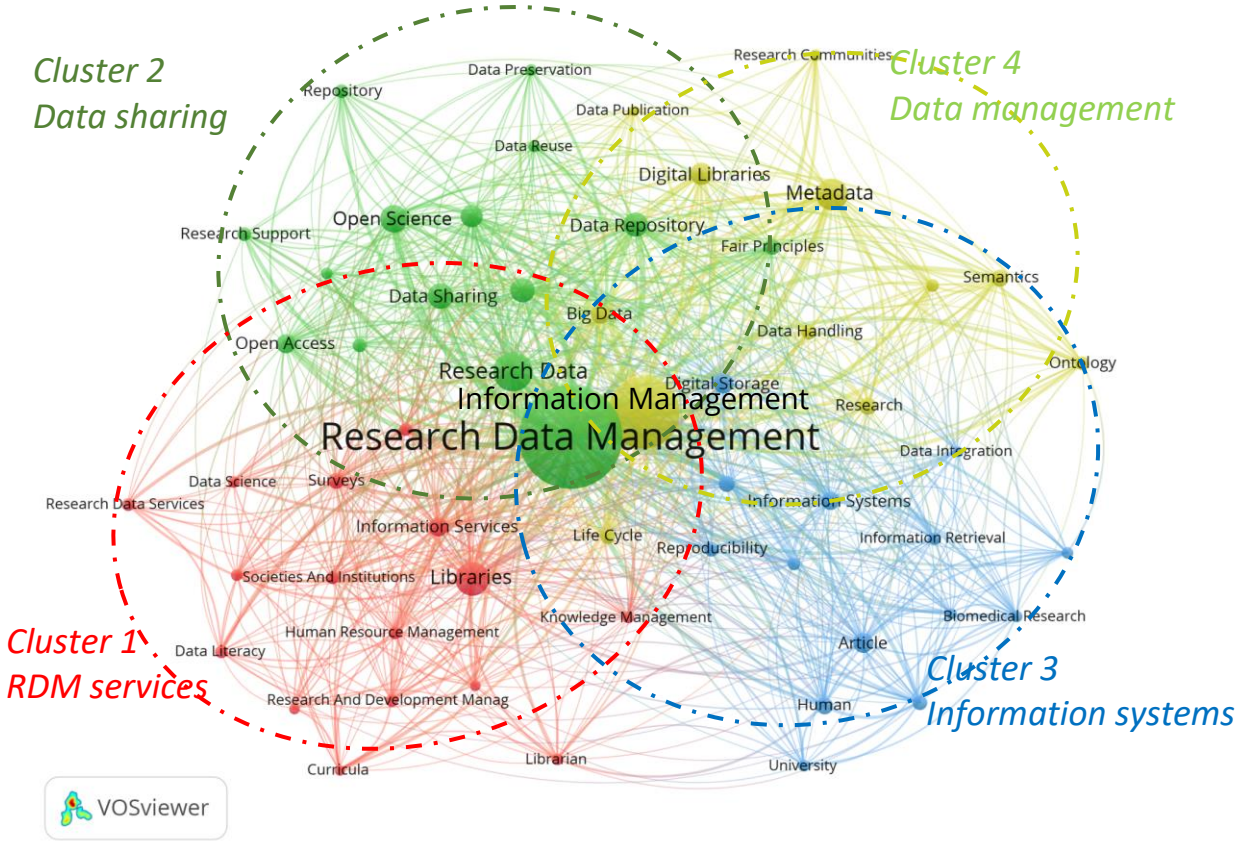

Figure 2: Keywords co-occurrence network

All keywords, showing the full counting of the occurrences, indicate the total number of occurrences of a term in all publications (Van Eck and Waltman 2020). Each theme has keywords that occur in its cluster, and they were sorted first by cluster and then by total link strength (Appendix). The total link strength is referred to the number of publications in which two or more keywords occur together (Van Eck and Waltman 2020). The keyword "Libraries" is the most common occurrence (4%) with the highest total link strength (3.35%) in RDM services (Cluster 1). Overall, "Libraries" is at fourth place for the total occurrences and fifth place for total link strength among the keywords. The primary keyword "Research Data Management", is placed in Data sharing (Cluster 2). It has the highest occurrence (421, 23%) and the highest total link strength (1345, 17.35%). The highest occurrence in Information systems (Cluster 3) is "Digital Storage" and "Article" respectively (27, 1%). "Information Management" is the highest occurrence (230, 13%) in Data management (Cluster 4) and is the second-highest occurrence overall. Findings show that only "Research Data Management" and "Information Management" have more than 200 occurrences with more than 900 total link strength. However, these two keywords are not positioned in only one cluster. All other keywords have below 100 occurrences. It

is worth mentioning that some keywords although were low in occurrences, but were high in total link strength around the keywords co-occurrences network. For example, some keywords were between 18 and 32 occurrences, but the total link strength was between 100 and 151, such as “Information Services” (Cluster 1), “Digital Libraries” (Cluster 4), “Digital Storage” (Cluster 3), “Article” (Cluster 3), “Surveys” (Cluster 1), “Human” (Cluster 3), “Big Data” (Cluster 4), “Open Access” (Cluster 2), “Information Systems” (Cluster 3), “Semantics” (Cluster 4) and “Information Processing” (Cluster 3) (see Appendix).

RDM publications associated with the keyword “Research Data Management” becomes the biggest node in the keywords co-occurrence network (Figure 2 and 3). Figure 3 shows the average publication year of the publications in which a keyword or a term occurs or the average publication year of the publications published by a source, an author, an organisation, or a country (Van Eck and Waltman 2020). Most publications have the publication year from 2017 onwards, specifically in RDM services (Cluster 1) and Data sharing (Cluster 2), indicating many keywords (shown in yellow, orange and red). Specifically, Cluster 2 falls under the theme of Data Sharing, indicating the most recent keywords occurred in recent publications from 2018 onwards (red). The keywords related to “Data Repository”, "Open Science", "Open Data", "Data Sharing", “Data Reuse” and “FAIR Principles” were indicated as the big nodes and red colour in Data sharing (Cluster 2) from this co-occurrence network. While the keyword "Libraries" was indicated as the biggest node and red colour in Cluster 1 under the theme of RDM services. Overall, there were many keywords with red colours in Data sharing (Cluster 2).

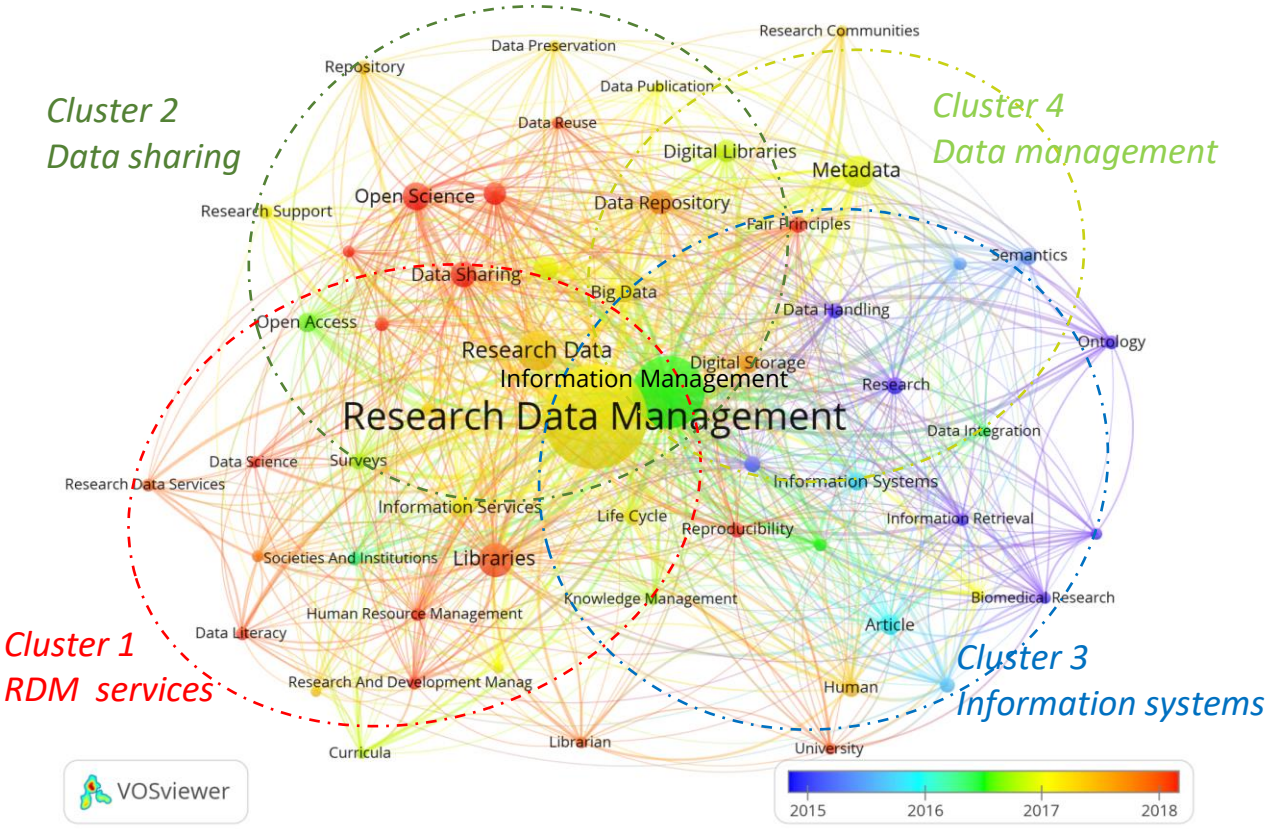

Figure 3: Keywords co-occurrence network by year of publications

Figure 4 presents the number of RDM publications from 1977 to 2021 (April 2021) based on the research areas/clusters. Each cluster has a different number of publications - RDM services (Cluster 1, 149), Data sharing (Cluster 2, 424), Information systems (Cluster 3, 118) and Data management (Cluster 4, 264). The study indicates that most publications were published in Data Sharing (Cluster 2) among the other clusters. There were not more than three publications for each cluster from 1977 to 2011. However, it is noticeable that there was a spectacular increase in most publications from 2012 onwards. Data Sharing (Cluster 2) shows an increase from 12 publications in 2012 to more than 70 publications in 2019. Data management (Cluster 4) was second of the most increasing RDM publications, from 8 publications in 2012 to 41 publications in 2018 and 2019. Overall, each cluster shows between 13% to 18% of total publications between 2017 and 2020. Most publications were gradually published between these four years, and it is still continually increasing. However, the number of publications in 2021 was not completed yet because this study was conducted in April 2021. All these numbers will be reflecting the growing interest in RDM, especially in relation to Data sharing (Cluster 2) and Data management (Cluster 4).

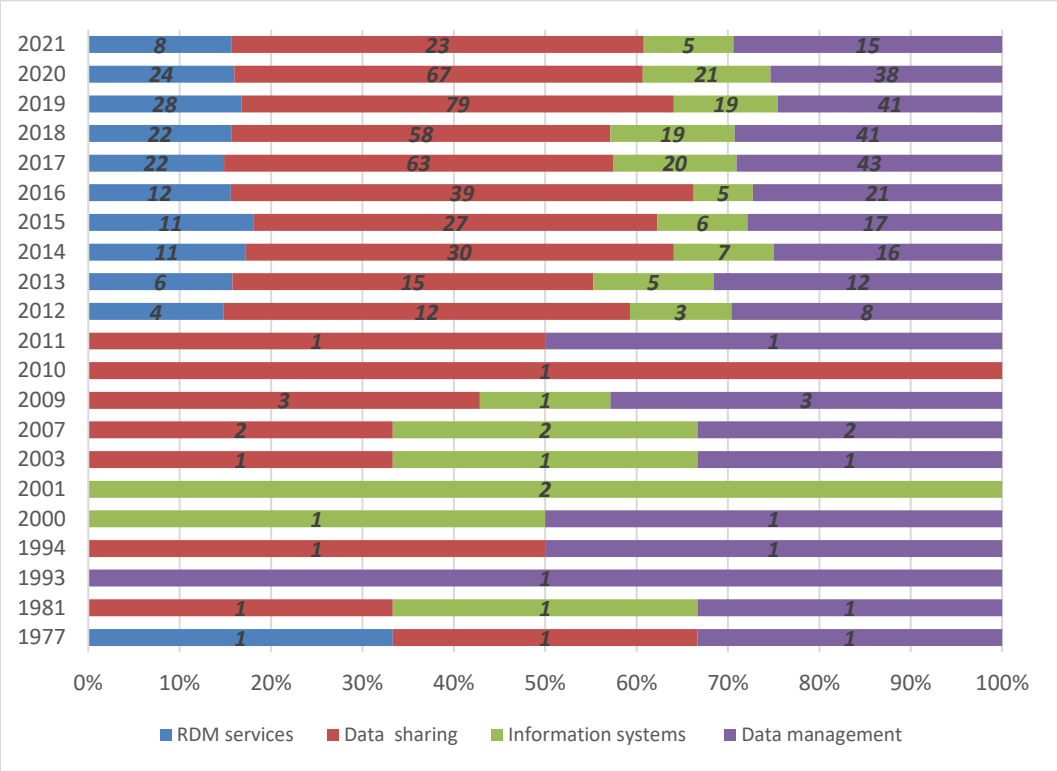

Figure 4: Number of RDM publications by years and clusters

RDM Subject Areas

This study categorised the subject areas according to the publications in each cluster, as presented in Table 2. The subject areas are referring to the Scopus database, and there are 26 subject areas with additional general subject areas containing multidisciplinary journals; (i) Life Sciences; (ii) Physical Sciences; (iii) Social Sciences and (iv) Health Sciences (García, Rodriguez-Sánchez, and Fdez-Valdivia 2011). Table 2 shows that RDM as a topic shares with a variety of subject areas. Because some publications would have more than one subject area, data in Table 2 presents more than the actual total number of publications. The cluster that has the highest number of publications covering the subject areas is Data sharing (Cluster 2, 660), followed by Data management (Cluster 4, 433), RDM services (Cluster 1, 220) and Information systems (Cluster 3, 209). Computer Science and Social Sciences are the two highly represented subject areas in RDM publications. The subject area Computer Science has the most number of publications in each cluster (more than 50% of total publications) - RDM services (Cluster 1, 53.69%), Data sharing (Cluster 2, 58.96%), Information systems (Cluster 3, 63.56%) and Data management (Cluster 4, 79.92%). Social Sciences represents more than 40 percent of total publications in RDM services (Cluster 1, 54.36%) and RDM sharing (Cluster 2, 44.58%). It was clear that the subject area of Computer Science and Social Sciences has been dominant in RDM services (Cluster 1) and Data sharing (Cluster 2).

Table 2: RDM Subject areas

Subject Area	**Total Publications (TP)**			
	Cluster 1 RDM SERVICES	**Cluster 2 DATA SHARING**	**Cluster 3 INFORMATION SYSTEMS**	**Cluster 4 DATA MANAGEMENT**
Agricultural and Biological Sciences	1 (0.67 %)	4 (0.94 %)	3 (2.54 %)	2 (0.76 %)
Arts and Humanities	3 (2.01 %)	14 (3.03 %)	2 (1.69 %)	6 (2.27 %)
Biochemistry, Genetics and Molecular Biology	1 (0.67 %)	11 (2.59 %)	8 (6.78 %)	8 (3.03 %)
Business, Management and Accounting	2 (1.34 %)	8 (1.89 %)	-	3 (1.14 %)
Chemical Engineering	1 (0.67 %)	3 (0.71 %)	1 (0.85 %)	2 (0.76 %)
Chemistry	4 (2.68 %)	7 (1.65 %)	3 (2.54 %)	6 (2.27 %)
Computer Science	**80 (53.69 %)**	**250 (58.96 %)**	**75 (63.56 %)**	**211 (79.92 %)**
Decision Sciences	5 (3.36 %)	22 (5.19 %)	7 (5.93 %)	18 (6.82 %)
Dentistry	-	1 (0.24 %)	-	-
Earth and Planetary Sciences	1 (0.67 %)	8 (1.89 %)	1 (0.85 %)	5 (1.89 %)
Economics, Econometrics and Finance	1 (0.67 %)	3 (0.71 %)	-	-
Energy	-	2 (0.47 %)	1 (0.85 %)	1 (0.38 %)
Engineering	8 (5.37 %)	24 (5.66 %)	11 (9.32 %)	19 (7.20 %)
Environmental Science	1 (0.67 %)	6 (1.42 %)	5 (4.24 %)	6 (2.27 %)
Health Professions	5 (3.36 %)	8 (1.89 %)	9 (7.63 %)	8 (3.03 %)
Mathematics	12 (8.05 %)	62 (14.62 %)	19 (16.10 %)	61 (23.11 %)
Medicine	9 (6.04 %)	29 (6.84 %)	27 (22.88 %)	19 (7.20 %)
Multidisciplinary	1 (0.67 %)	1 (0.24 %)	1 (0.85 %)	1 (0.38 %)
Neuroscience	1 (0.67 %)	4 (0.94 %)	4 (3.39 %)	1 (0.38 %)
Nursing	-	-	1 (0.85 %)	-
Pharmacology, Toxicology and Pharmaceutics	1 (0.67 %)	2 (0.47 %)	1 (0.85 %)	1 (0.38 %)
Physics and Astronomy	1 (0.67 %)	1 (0.24 %)	-	1 (0.38 %)
Psychology	1 (0.67 %)	1 (0.24 %)	1 (0.85 %)	1 (0.38 %)
Social Sciences	**81 (54.36 %)**	**189 (44.58 %)**	**29 (24.58 %)**	**53 (20.08 %)**
Total	**220**	**660**	**209**	**433**

DISCUSSIONS

This study aims to identify the main themes from RDM publications and the potential field of research on RDM, examining the relationship among RDM publications through mapping of knowledge domains analysis. This study has developed the themes from the publications related to the RDM using the primary keyword “Research Data Management”. It has applied the bibliometric approach to evaluate research productivity (Moed, Luwel, and Nederhof 2001). Bibliometric studies have been growing to reveal this study's statistics and literature growth (Ahmi and Mohd Nasir 2019). Indeed, there were limited bibliometric studies related to the topic of RDM that have been published.

This section discusses the findings from the bibliometric analysis above by revisiting the research question in this study. The selection of publications from Scopus databases was based on the primary keyword “Research Data Management”. VOSviewer was used for data analysis to extract the main themes related to the RDM publication by emerging the

four clusters. This study summarised the four clusters represented by the four main themes: RDM Services, Data Sharing, Information Systems, and Data Management.

Cluster 1: RDM Services

The first cluster was represented by keywords oriented in the topics around RDM services. The current library services have been changed rapidly because of the various materials and formats available online and offline. Indeed, the research data could be in various types of materials and formats involved by the library in determining the new library services, especially related to the RDM. Therefore, the co-occurrences analysis shows that most publications discussed the library, information services, data literacy, and institutional repository, indicating higher frequencies from this analysis. The library has to play the roles of RDM because of the reliable capability and expertise of RDM in the range of skills and knowledge (Avuglah and Underwood 2019). The library has been accustomed to coordinate the RDM services, especially in advocacy and giving training on RDM to the researchers. In addition, the librarian can handle the RDM services, who are well experienced in library services such as information services, knowledge management, institutional repository and reference services. Most library services could be related to the RDM services, such as creating the metadata of research data in institutional repository and information services in advocating the researchers through training and consultation services and others. Indeed, the library could increase the awareness of RDM by advocacy program and conduct training and workshops (Marlina and Purwandari 2019; Wiorogórska, Leśniewski, and Rozkosz 2018; Wiljes and Cimiano 2019; Y. Li, Dressel, and Hersey 2019). Many publications also mentioned that the library had been played the primary roles in developing and delivering the RDM services to the researchers (Bunkar and Bhatt 2020; Nitecki and Davis 2019; Harrison 2018; Mushi et al. 2020; Henderson and Knott 2015; Tammaro et al. 2019; Chawinga and Zinn 2020; Hickson et al. 2016; Koltay 2016a; 2016b; Cox and Pinfield 2014; Pinfield, Cox, and Smith 2014)

Cluster 2: Data Sharing

The second cluster was closely related to the keywords co-occurrences related to data sharing issues. The primary keyword “Research Data Management” has been placed in this cluster. According to Zhu et al. (2015), the highest frequency keywords showed high concentration and focused on a particular topic. This keyword demonstrates that it will help develop application research by emphasising the execution of the research findings. It occurs in this cluster when the primary keyword of this study has been placed with other important keywords, which are oriented around data sharing such as “Research Data”, “Open Science”, “Data Sharing”, “Data Curation”, “Open Data” And “FAIR Principles”. The importance of RDM for researchers and supporting research units is growing in the context of the open science movement (Vilar and Zabukovec 2019). Data sharing is one of the main aspects of open science that promotes excellent managing of the research data as it is a prerequisite of open science and RDM policies (Timmermann 2019). Open data could be considered particularly important for achieving the open science agenda, with open data is frequently indicated to data sharing and data reuse (Mosconi et al. 2019). Therefore, comprehensive research support is needed for integrating the RDM and open science strategies (Rice 2019). The FAIR principles are also very important when the researchers are encouraged to have a greater engagement with RDM and openness.

Therefore, the FAIR principles have been introduced when data should be made as findable, accessible, interoperable, and reusable (Schöpfel et al. 2018). Most keywords in this cluster were closely related to each other and easy to understand the relationship among them.

Cluster 3: Information Systems

This cluster was related to the information systems issues in developing and implementing RDM. Generally, information systems are an integral element for collecting, storing, and processing data. The data is used to provide information and contribute to knowledge and digital products. Therefore, clinical and biomedical research was widely used as the information systems for managing the research data, which most research data has been digitised as the primary source of research in these fields (Tang et al. 2018). Digital storage was an essential part of the information systems which many institutions, especially medical institutions, providing the storage for research data (Suhr et al. 2020; Tang et al. 2018). The information systems in RDM involve data processing, integration and retrieval (Pinfield, Cox, and Smith 2014). Instead of the needed skills and knowledge in developing the RDM services, the technology infrastructure may also be another resource allocation for RDM, such as software and hardware to support the researchers regarding RDM activities (Aydinoglu, Dogan, and Taskin 2017). Lack of information and technology infrastructure becomes a challenge in providing the facilities to the researchers for storage, preservation, and open data to fulfil the institution's requirement, funders and publishers (Kruse and Thestrup 2014). The future aspects of RDM could be explored in research related to information systems and technology infrastructure issues that will be effective to development and activities on RDM.

Cluster 4: Data management

According to the results of this study, most RDM publications were under the subject area of Computer Sciences and Social Sciences, specifically in relation to the library and information science area. In other words, the RDM also could be under the topic related to information management. Basically, the components of RDM are intimately related to information management. The keywords in this cluster were oriented to the data management activities, such as metadata management. The metadata of research data may be different from research publications such as journal article. The creation of metadata for research data may need to researcher's contribution in describing its metadata while regularly the research publication’s metadata could be described accurately by the librarians. Certainly, metadata management is vital in the flexibility and efficiency of data management because the metadata could be accessible on data platforms or databases for the long-term preservation of research data (Finkel et al. 2020). However, many researchers have not used the standardised ontology or metadata schema, giving the library challenges in providing and creating semantically linked sources for research data (Schirrwagen et al. 2019). Research on data management should be focused on by the researchers involved in data handling, from creating the metadata until data publication throughout the research lifecycle. The researchers are critical to explore more aspects of RDM to fulfil personal needs, the requirements of institutions, funders and publishers.

Literature growth

The publications from the Scopus database were extracted for all years of publications in the Scopus database. The study found that there were publications with the keyword "research data management" from 1977 to the date of this study in April 2021. The analysis done was prominent when many publications on RDM have been started from 2012 and increasing until today. This growing of publications shows that the area of RDM is increasingly fascinating attention of the researchers to study more about the related issues of RDM. According to the analysis has been done by Zhang & Eichmann-Kalwara (2019), the study revealed that the top three clusters from the literature related to RDM were scientific collaboration, research support service and data literacy. This present study shows that the research support services and data literacy were part of RDM services (Cluster 1), which is in one cluster. The average publication years of data literacy as a cluster in Zhang & Eichmann-Kalwara (2019) study was 2010. However, the growth of the literature around RDM is increasing, and the present study indicates that data literacy is average from 2017 onwards. It is the same that occurred with other clusters. This study plainly demonstrates that the publications about RDM are growing, especially in Data sharing (Cluster 2), when the issues related to open science and data sharing. These issues are currently discussed in many works of literature as research on open data has proliferated since 2009 with the development of various initiatives (Zhang et al., 2018).

The study illustrates the results using keywords co-occurrence network analysis that could be utilised for classification. The classification shows the clusters that represent the co-occurrences that developed with easy to be interpreted. The various knowledge domains have been demonstrated using the VOSviewer as a primary tool for this study. The approach of this study was representative when the publications have chosen using the primary keyword "Research Data Management" with multidisciplinary and more expansive representations of views and opinions. This study also uses the analysis that allows the dynamic relationship among the RDM publications using the author and index keywords available in each publication. As a result, four clusters developed from this study show the dynamic relationship among numerous keywords even within different clusters. It can be demonstrated by analysing the link among the keywords and the calculation of total link strength. This study also used the scalable approach when the study examined the RDM publications for all years (1977-April 2021) with not limited to a specific period. This approach can analyse the relationship and trends of RDM publications in the Scopus database and obtain wide-ranging findings.

CONCLUSION

This study specifically focuses on developing the main themes of RDM publications from the Scopus database using bibliometric analysis. The publications were extracted for all years from 1977 to April 2021. The main themes found in this study indicated the keywords co-occurrences network and literature growth of RDM publications. The findings of this study covered the publications data obtained from the Scopus database only. Undeniable that the Scopus database is one of the largest databases, and there are unindexed journals related to RDM topics that might have been missed (Sweileh et al.

2017). The publications were obtained by searching the keyword "research data management" under the title and abstract, and this keyword was mandatory for keyword searching. Further study could be expanded by using specific keywords to focus on specific themes or fields related to the RDM.

This paper stands out from any other bibliometric study of RDM literature that has previously been published as it started with developing the clusters to identify the main themes based on the keywords co-occurrence network. Then, the bibliometric approach was used to answer the research question. Most RDM publications have focused mainly on subject areas of computer science and social science that were interrelated. However, most of them were explicitly focused on the area of library and information science. The issues on open science, data sharing and open data (Data sharing, Cluster 2) and libraries (RDM services, Cluster 1) were discussed in recent years and have been most published.

This study's findings could help researchers in the RDM field understand the current state of RDM publications and their issues discussed in the literature. This study also could help them to propose further research on related topics of RDM. The issues of RDM could be discussing more in the future due to the adoption of the RDM with current research consciousness, such as open science. Bibliometric approaches are used in this analysis to expand and complement previous studies on RDM literature. This study suggests more studies on RDM to be explored and discussed the related issues since the RDM still in the development and implementation process in many countries. Future research could obtain the attention of researchers and practitioners to contribute the knowledge of RDM and highlight the importance of RDM.

REFERENCES

Ahmi, Aidi, and Mohd Herry Mohd Nasir. 2019. "Examining the Trend of the Research on Extensible Business Reporting Language (Xbrl): A Bibliometric Review." *International Journal of Innovation, Creativity and Change* 5 (2): 1145–67. https://www.researchgate.net/publication/336055342.

Ashiq, Murtaza, Muhammad Haroon Usmani, and Muhammad Naeem. 2020. "A Systematic Literature Review on Research Data Management Practices and Services." *Global Knowledge, Memory and Communication*. https://doi.org/10.1108/GKMC-07-2020-0103.

Avuglah, Bright Kwaku, and Peter G. Underwood. 2019. "Research Data Management (RDM) Capabilities at the University of Ghana, Legon." *Library Philosophy and Practice*. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85066066748&partnerID=40&md5=52b10dfefaa0f4cef9146b7930947bea.

Aydinoglu, Arsev Umur, Guleda Dogan, and Zehra Taskin. 2017. "Research Data Management in Turkey: Perceptions and Practices." *Library Hi Tech* 35 (2): 271–89. https://doi.org/10.1108/LHT-11-2016-0134.

Baas, Jeroen, Michiel Schotten, Andrew Plume, Grégoire Côté, and Reza Karimi. 2020. "Scopus as a Curated, High-Quality Bibliometric Data Source for Academic Research in Quantitative Science Studies." *Quantitative Science Studies* 1 (1): 377–86. https://doi.org/10.1162/qss_a_00019.

Benckendorff, Pierre, and Anita Zehrer. 2013. "A Network Analysis of Tourism Research." *Annals of Tourism Research* 43: 121–49. https://doi.org/10.1016/j.annals.2013.04.005.

Bunkar, Anjana R., and Dhaval D. Bhatt. 2020. "Perception of Researchers & Academicians of Parul University towards Research Data Management System & Role of Library: A Study." *DESIDOC Journal of Library and Information Technology* 40 (3): 139–46. https://doi.org/10.14429/djlit.40.3.15302.

Chawinga, Winner Dominic, and Sandy Zinn. 2020. "Research Data Management at a Public University in Malawi: The Role of 'Three Hands.'" *Library Management* 41 (6–7): 467–85. https://doi.org/10.1108/LM-03-2020-0042.

Cox, Andrew M., and Stephen Pinfield. 2014. "Research Data Management and Libraries: Current Activities and Future Priorities." J*ournal of Librarianship and Information Science* 46 (4): 299–316. https://doi.org/10.1177/0961000613492542.

Eck, Nees Jan van, and Ludo Waltman. 2015. "VOSviewer Manual." Leiden, Netherlands: Leiden University (Universiteit Leiden). Accessed 2020. http://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.1.pdf.

Finkel, M, A Baur, T K D Weber, K Osenbrück, H Rügner, C Leven, M Schwientek, et al. 2020. "Managing Collaborative Research Data for Integrated, Interdisciplinary Environmental Research." *Earth Science Informatics* 13 (3): 641–54. https://doi.org/10.1007/s12145-020-00441-0.

Flores, Jodi Reeves, Jason J. Brodeur, Morgan G. Daniels, Natsuko Nicholls, and Ece Turnator. 2015. "Libraries and the Research Data Management Landscape." In *The Process of Discovery: The CLIR Postdoctoral Fellowship Program and the Future of the Academy*, 82–102. http://www.clir.org/pubs/reports/pub167/.

García, J. A., Rosa Rodriguez-Sánchez, and J. Fdez-Valdivia. 2011. "Ranking of the Subject Areas of Scopus." *Journal of the American Society for Information Science and Technology* 62 (10): 2013–23. https://doi.org/10.1002/asi.21589.

Harrison, R. 2018. "The Academic Library and the Research Office: Providing Scholarly Communications Support at Imperial College London-A Case Study." In *Collaboration and the Academic Library: Internal and External, Local and Regional, National and International*, 143–50. Imperial College London, London, United Kingdom: Elsevier. https://doi.org/10.1016/B978-0-08-102084-5.00013-4.

Henderson, Margaret E., and Teresa L. Knott. 2015. "Starting a Research Data Management Program Based in a University Library." *Medical Reference Services Quarterly* 34 (1): 47–59. https://doi.org/10.1080/02763869.2015.986783.

Hickson, Susan, Kylie Ann Poulton, Maria Connor, Joanna Richardson, and Malcolm Wolski. 2016. "Modifying Researchers' Data Management Practices." *IFLA Journal* 42 (4): 253–65. https://doi.org/10.1177/0340035216673856.

Koltay, Tibor. 2016a. "Are You Ready? Tasks and Roles for Academic Libraries in Supporting Research 2.0." *New Library World* 117 (1–2): 94–104. https://doi.org/10.1108/NLW-09-2015-0062.

———. 2016b. "Facing the Challenge of Data-Intensive Research: Research Data Services and Data Literacy in Academic Libraries." In *Advances in Library Administration and Organization*, edited by W Baker, D and Evans, 35: 45–61. Advances in Library Administration and Organization. Howard House, Wagon Lane, Bingley, W Yorkshire BD16 1WA, England: Emerald Group Publishing Ltd. https://doi.org/10.1108/S0732-067120160000035008.

Kruse, Filip, and Jesper Boserup J.B. Thestrup. 2014. "Research Libraries' New Role in Research Data Management, Current Trends and Visions in Denmark." *LIBER Quarterly* 23 (4): 310–35. https://doi.org/10.18352/lq.9173.

Li, Xiao, Peng Wu, Geoffrey Qiping Shen, Xiangyu Wang, and Yue Teng. 2017. "Mapping the Knowledge Domains of Building Information Modeling (BIM): A Bibliometric Approach." *Automation in Construction* 84: 195–206. https://doi.org/10.1016/j.autcon.2017.09.011.

Li, Yuan, Willow Dressel, and Denise Hersey. 2019. "Research Data Management: What Can Librarians Really Help?" In *20th International Conference on Grey Literature: Research Data Fuels and Sustains Grey Literature, GL 2018*, 67–74. Loyola University New Orleans, United States: TextRelease. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85062960063&partnerID=40&md5=843ffb6d2b81174ab2cc46b010199f5a.

Liu, Linqing, and Shiye Mei. 2016. "Visualizing the GVC Research: A Co-Occurrence Network Based Bibliometric Analysis." *Scientometrics* 109 (2): 953–77. https://doi.org/10.1007/s11192-016-2100-5.

Marlina, Ekawati, and Betty Purwandari. 2019. "Strategy for Research Data Management Services in Indonesia." *Procedia Computer Science* 161: 788–96. https://doi.org/10.1016/j.procs.2019.11.184.

Martín-Martín, Alberto, Mike Thelwall, Enrique Orduna-Malea, and Emilio Delgado López-Cózar. 2021. "Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A Multidisciplinary Comparison of Coverage via Citations." *Scientometrics* 126: 871–906. https://doi.org/10.1007/s11192-020-03690-4.

Moed, Henk F., Marc Luwel, and A. J. Nederhof. 2001. "Towards Research Performance in the Humanities." *Library Trends* 50 (3): 498–520.

Mosconi, Gaia, Qinyu Li, Dave Randall, Helena Karasti, Peter Tolmie, Jana Barutzky, Matthias Korn, and Volkmar Pipek. 2019. "Three Gaps in Opening Science." *Computer Supported Cooperative Work: CSCW: An International Journal* 28 (3–4): 749–89. https://doi.org/10.1007/s10606-019-09354-z.

Mushi, G.E. Gilbert Exaud, Heila Pienaar, Martie van Deventer, and Martie van Deventer. 2020. "Identifying and Implementing Relevant Research Data Management Services for the Library at the University of Dodoma, Tanzania." *Data Science Journal* 19 (1). https://doi.org/10.5334/dsj-2020-001.

Nitecki, Danuta A., and Mary Ellen K. Davis. 2019. "Expanding Academic Librarians' Roles in the Research Life Cycle." *Libri* 69 (2): 117–25. https://doi.org/10.1515/libri-2018-0066.

Perrier, Laure, Erik Blondal, A. Patricia Ayala, Dylanne Dearborn, Tim Kenny, David Lightfoot, Roger Reka, Mindy Thuna, Leanne Trimble, and Heather MacDonald. 2017. "Research Data Management in Academic Institutions: A Scoping Review." PLoS ONE 12 (5). https://doi.org/10.1371/journal.pone.0178261.

Pinfield, S., Andrew M. Cox, and Jen Smith. 2014. "Research Data Management and Libraries: Relationships, Activities, Drivers and Influences." Edited by Pascal Launois. *PLoS ONE* 9 (12): 1-28. https://doi.org/10.1371/journal.pone.0114734.

Rice, Robin. 2019. "Supporting Research Data Management and Open Science in Academic Libraries: A Data Librarian's View." *Mitteilungen Der Vereinigung Österreichischer Bibliothekarinnen Und Bibliothekare* 72 (2): 263–73. https://doi.org/10.31263/voebm.v72i2.3303.

Schirrwagen, Jochen, Philipp Cimiano, Vidya Ayer, Christian Pietsch, Cord Wiljes, Johanna Vompras, and Dirk Pieper. 2019. "Expanding the Research Data Management Service Portfolio at Bielefeld University According to the Three-Pillar Principle towards Data FAIRness." *Data Science Journal* 18 (1). https://doi.org/10.5334/dsj-2019-006.

Schöpfel, Joachim, Coline Ferrant, Francis André, Renaud Fabre, Joachim Schopfel, Coline Ferrant, Francis Andre, et al. 2018. "Research Data Management in the French National Research Center (CNRS)." *Data Technologies and Applications* 52 (2): 248–65. https://doi.org/10.1108/DTA-01-2017-0005.

Shiffrin, Richard M., and K. Borner. 2004. "Mapping Knowledge Domains." *Proceedings of the National Academy of Sciences* 101 (Supplement 1): 5183–85. https://doi.org/10.1073/pnas.0307852100.

Suhr, M, C Lehmann, C R Bauer, T Bender, C Knopp, L Freckmann, B Öst Hansen, et al. 2020. "Menoci: Lightweight Extensible Web Portal Enhancing Data Management for Biomedical Research Projects." *BMC Bioinformatics* 21 (1). https://doi.org/10.1186/s12859-020-03928-1.

Sweileh, Waleed M., Samah W. Al-Jabi, Adham S. AbuTaha, Sa'ed H. Zyoud, Fathi M.A. Anayah, and Ansam F. Sawalha. 2017. "Bibliometric Analysis of Worldwide Scientific Literature in Mobile - Health: 2006-2016." *BMC Medical Informatics and Decision Making* 17 (1): 1–12. https://doi.org/10.1186/s12911-017-0476-7.

Tammaro, Anna Maria, Krystyna K. Matusiak, Frank Andreas Sposito, and Vittore Casarosa. 2019. "Data Curator's Roles and Responsibilities: An International Perspective." *Libri* 69 (2): 89–104. https://doi.org/10.1515/libri-2018-0090.

Tang, Haijing, Keyan Gao, Yangdong Zhou, Wenhao Zheng, Xu Yang, and Jinfeng Zhao. 2018. "Adopting Data Analysis and Visualization Technology to Construct Clinical Research Data Management and Analysis System." In *ACM International Conference Proceeding Series*, 49–53. Association for Computing Machinery. https://doi.org/10.1145/3301761.3301767.

Timmermann, Marie. 2019. "A Collective Challenge: Open Science from the Perspective of Science Europe." *Mitteilungen Der Vereinigung Österreichischer Bibliothekarinnen Und Bibliothekare* 72 (2): 424–30. https://doi.org/10.31263/voebm.v72i2.2831.

Vilar, Polona, and Vlasta Zabukovec. 2019. "Research Data Management and Research Data Literacy in Slovenian Science." *Journal of Documentation* 75 (1): 24–43. https://doi.org/10.1108/JD-03-2018-0042.

Wahid, Ratnaria, Aidi Ahmi, and A. S.A.Ferdous Alam. 2020. "Growth and Collaboration in Massive Open Online Courses: A Bibliometric Analysis." *International Review of Research in Open and Distance Learning* 21 (4): 292–322. https://doi.org/10.19173/IRRODL.V21I4.4693.

Wiljes, Cord, and Philipp Cimiano. 2019. "Teaching Research Data Management for Students." *Data Science Journal* 18 (1): 1–9. https://doi.org/10.5334/dsj-2019-038.

Wiorogórska, Zuzanna, Jędrzej Leśniewski, and Ewa Rozkosz. 2018. "Data Literacy and Research Data Management in Two Top Universities in Poland. Raising Awareness." In *5th European Conference on Information Literacy in the Workplace, ECIL 2017*, edited by Kurbanoğlu S., Boustany J., Špiranec S., Grassian E., Mizrachi D., and Roy L., 205–14. https://doi.org/10.1007/978-3-319-74334-9_22.

Zhang, Li, and Nickoal Eichmann-Kalwara. 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 (1). https://doi.org/10.7710/2162-3309.2266.

Zhang, Yun, Weina Hua, and Shunbo Yuan. 2018. "Mapping the Scientific Research on Open Data: A Bibliometric Review." *Learned Publishing* 31 (2): 95–106. https://doi.org/10.1002/leap.1110.

Zhu, Lin, Xiantao Liu, Sha He, Jun Shi, and Ming Pang. 2015. "Keywords Co-Occurrence Mapping Knowledge Domain Research Base on the Theory of Big Data in Oil and Gas Industry." *Scientometrics* 105 (1): 249–60. https://doi.org/10.1007/s11192-015-1658-7.

APPENDIX

Appendix: List of keywords co-occurrences with total link strength by cluster

Keyword	Occurrences	%	Total link strength	%	Cluster
Libraries	65	4%	260	3.35%	***Cluster 1 (RDM services)***
Information Services	26	1%	151	1.95%	
Surveys	20	1%	126	1.63%	
Data Literacy	14	1%	52	0.67%	
Institutional Repository	13	1%	76	0.98%	
Societies And Institutions	13	1%	75	0.97%	
Human Resource Management	13	1%	73	0.94%	
Research And Development Management	13	1%	73	0.94%	
Research Data Services	13	1%	41	0.53%	
Librarians	12	1%	54	0.70%	
Students	11	1%	65	0.84%	
Knowledge Management	11	1%	60	0.77%	
Data Science	11	1%	46	0.59%	
Curricula	10	1%	58	0.75%	
Education	10	1%	54	0.70%	
Training	10	1%	42	0.54%	
Research Data Management	421	23%	1345	17.35 %	***Cluster 2 (Data sharing)***
Research Data	85	5%	383	4.94%	
Open Science	47	3%	192	2.48%	
Data Sharing	41	2%	167	2.15%	
Data Curation	39	2%	158	2.04%	
Data Repository	38	2%	177	2.28%	
Open Data	33	2%	162	2.09%	
Open Access	26	1%	112	1.45%	
Fair Principles	20	1%	80	1.03%	
Repository	16	1%	57	0.74%	
Research Support	16	1%	56	0.72%	
Scholarly Communication	15	1%	59	0.76%	
Data Management Plan	12	1%	53	0.68%	
Data Preservation	12	1%	43	0.55%	
Data Reuse	11	1%	52	0.67%	
Digital Storage	27	1%	140	1.81%	***Cluster 3 (Information systems)***
Articles	27	1%	131	1.69%	
Human	22	1%	120	1.55%	

Information Systems	24	1%	120	1.42%	
Data Management	20	1%	88	1.14%	
Information Processing	18	1%	100	1.29%	
Reproducibility	17	1%	77	0.99%	
Information Use	14	1%	75	0.97%	
Information Retrieval	13	1%	71	0.92%	
Biomedical Research	12	1%	84	1.08%	
Clinical Research	11	1%	55	0.71%	
University	11	1%	50	0.65%	
Software	10	1%	61	0.79%	
Data Integration	10	1%	52	0.67%	
Information Management	230	13%	976	12.59 %	***Cluster 4 (Data management)***
Metadata	59	3%	262	3.38%	
Digital Libraries	32	2%	147	1.90%	
Big Data	25	1%	114	1.47%	
Semantics	22	1%	106	1.37%	
Research	21	1%	98	1.26%	
Life Cycle	17	1%	90	1.16%	
Data Handling	17	1%	80	1.03%	
Ontology	16	1%	85	1.10%	
Linked Data	14	1%	83	1.07%	
Research Communities	12	1%	44	0.57%	
Data Publication	10	1%	49	0.63%	