RESEARCH DATA MANAGEMENT IN HIGHER EDUCATIONS: KNOWLEDGE MAPPING USING BIBLIOMETRIC ANALYSIS

Lis Setyowati*[1], Heriyanto[2]
[1]Faculty of Engineering, Diponegoro University
[2]Faculty of Humanities, Diponegoro University

*Correspondence: lis@ft.undip.ac.id

ABSTRACT
Research Data Management is gaining popularity nowadays, due to the widespread awareness of the importance of research data curation and sharing. Development is currently taking place. This paper aimed to map the knowledge domain in Research Data Management, particularly within the higher education setting. Using Vosviewer, the bibliometrics software, this research identified the most influential authors, publications, as well as the trending topics in the field. The result shows that Tenopir and her publications are the prominent ones in the field and that the trending topic in RDM-HEs research was academic library-related issues.

ABSTRAK
Manajemen Data Penelitian semakin populer saat ini, karena kesadaran luas akan pentingnya kurasi dan pembagian data penelitian. Pengembangan saat ini sedang berlangsung. Makalah ini bertujuan untuk memetakan domain pengetahuan dalam Manajemen Data Penelitian, khususnya dalam hal pengaturan pendidikan tinggi. Dengan menggunakan Vosviewer, penelitian ini mengidentifikasi penulis, publikasi, dan topik yang paling berpengaruh di lapangan. Hasilnya menunjukkan bahwa Tenopir dan publikasinya adalah yang menonjol di lapangan dan bahwa topik yang sedang hangat dalam penelitian RDM-HE adalah masalah yang berkaitan dengan perpustakaan akademik.

Keywords: *Bibliometrics; Research data management; Vosviewer; Visual network software; Visual analysis*

1. INTRODUCTION

Researchers are harnessing new methods and tools such as software, hardware, instruments, and equipments to conduct their research, as well as to use new data sources (Surkis et al. 2015). Thus, science has embraced a new paradigm of what is so-called "e-science" or "e-research". E-science/e-research is characterized as more collaborative, more computational, and more data intensive than the previous paradigms (Hey, Tansley, and Tolle 2009).

Consequently, data plays an important part in e-science, more importantly, "data is the new currency for research" (Markoff 2013). Heavily relying on computation technology, today's researches generate a larger volume of digital research data than ever before. The various formats of the digital data have added the complexity of the increasing number of the data generated. As e-science is data-driven, such complexity becomes a challenge to manage. Therefore, researchers need to apply effective and efficient data management (Surkis et al. 2015).

Research Data Management (RDM), is defined as "a method that enable the integration, curation and interoperability of data created during the scientific process, i.e. the production, access, verification, persistent storage and reuse of this data with the help of adequate and easy-to-use tools in virtual research infrastructures"(Schneider, n.d.). RDM enable researchers to carefully plan, archive, and preserve data, thus making it easier to have reproducible and transparent research data (Surkis et al. 2015). Through research data sharing, researchers can gain many benefits, i.e.: (1) data reuse by other researchers in other fields or for other contexts (Patel, Authors, and Patel 2016)(Tam, Fry, and Probets 2014); (2) getting citation, thus leading to reputation enhancement(Patel, Authors, and Patel 2016) (Tam, Fry, and Probets 2014); (3) data

shared will be reaffirmed by other researchers thus increasing trust on the datasets (MacMillan 2014)(Patel, Authors, and Patel 2016)(Tam, Fry, and Probets 2014); (4) data sharing imply the transparency in the research process; (5) using the available research data will save the researchers' time and efforts(Patel, Authors, and Patel 2016); (6) serving the obligation to funders to share data (Solutions 2013); (7) meeting publishers' requirements to publish the datasets(MacMillan 2014).

By looking at its importance, RDM is gaining more attention and wider acceptance from many stakeholders, either from research institutions, research funders, or publishers. RDM stakeholders has been variously categorized in the literature, they can be assembled into four main categories: government and funders, research support units, university leadership and researchers (Flores et al. 2015) The growing application of RDM is due to several factors, i.e. the demand for research accountability from the research funders (MacMillan 2014), the changing nature of science, which is more collaborative and interdisciplinary, (MacMillan 2014) data reuse, which means maximum return-on-investment for funders (Surkis et al. 2015)(MacMillan 2014), and data sharing which comply with Open Access agenda (Solutions 2013).

Higher Educations (HEs), which also serve as research institutions, are among the many bodies which encourage RDM. Many universities have already applied RDM, as evidenced by the research conducted by various researchers (Hayes, Harroun, and Temple, n.d.; Steeleworthy 2014; Dixon 2012; Feature 2011; Arias-coello, n.d.; Choudhury 2008; Tripathi, Shukla, and Sonker 2017; Johnson 2012; Chad and Enright 2014). These researches also revealed the growth of knowledge on RDM in HEs context. The knowledge growth can also be seen from a literature review.

A literature review can be used to identify the conceptual content of the research field. It also provides guidance for future research(Raghuram, Tuertscher, and Garud 2010). A literature review on RDM in HEs context was conducted by Perrier et.al (Perrier et al. 2017) who conducted a large scale literature review on RDM in HEs context. They gathered data from 40 databases and analysed them using "Data Lifecycle" which was proposed by the United Kingdom Data Archive as a framework. Although this was a comprehensive study, it did not consider individual articles' research impact.

As identifying the most relevant and influential articles is critical to a literature review's contribution(Wang et al. 2017), a different approach to literature review needs to be taken using a bibliometric analysis. The bibliometric analysis relies on the use of quantitative methods to examine documents, thus, helping to uncover knowledge structure and development of the research field (Pitchard 1969). Measuring publications that are heavily cited by others over time enable us to get to get an overview of the prominent research issues on RDM, particularly in higher education settings.

Bibliometrics analysis yields various maps of relations among authors or even relations between key research terms in the field. This paper aimed to map the knowledge domain in Research Data Management, particularly within the higher education setting. Using Vosviewer, a bibliometrics software, this research seeks to identify the most influential authors, publications, as well as the hot topics in the field.

2. METHOD

This study uses bibliometric analysis, which applies mathematical and statistical methods to books and other media of communication (Pitchard 1969). Therefore, the primary data collected are bibliographic data, for instance, title, abstract, descriptors, identifiers, and references. Data were collected on July 4, 2019, by performing a search on SCOPUS database to retrieve all RDM-

HEs related studies. The keywords used for the search were “research data”, “management”, “curation”, “preservation”, “university*”, “college*”, “higher education*” and “academic*”. The search was also limited to articles, conference papers, and book chapters. The search was also limited to a 20-year timespan, e.i. 2000 to 2019, resulting in 655 records. The dataset was then analyzed to identify any duplications and invalid data, leaving 652 records for further analysis.

The datasets were then analyzed using VosViewer, a software used to conduct a bibliometric analysis. Bibliometric analysis is a method that has been used to understand the knowledge base of a research field (Vogel and Güttel 2013), by measuring the relationship among documents through citation, either direct citation (the citing of earlier document by a new document) or bibliographic coupling (sharing one or more references by two documents)(Small 1973). The bibliographic coupling used herein is aimed to identify citation or co-citation which can be used to identify prominent authors and journals, the most influential papers (Chen, Song, and Zhu 2007), as well as research hotspot in the field(Eck and Waltman 2017).

3. RESULTS AND DISCUSSION

Based on the data that were collected from some academic databases, it was found that there a growing body of knowledge on RDM-HEs. This can be witnessed from the number of papers on the topic for the last 20 years (Fig. 1). Figure 1 shows various topics discussed. Major development can be witnessed from the last 10 years which shows promising exponential growth.

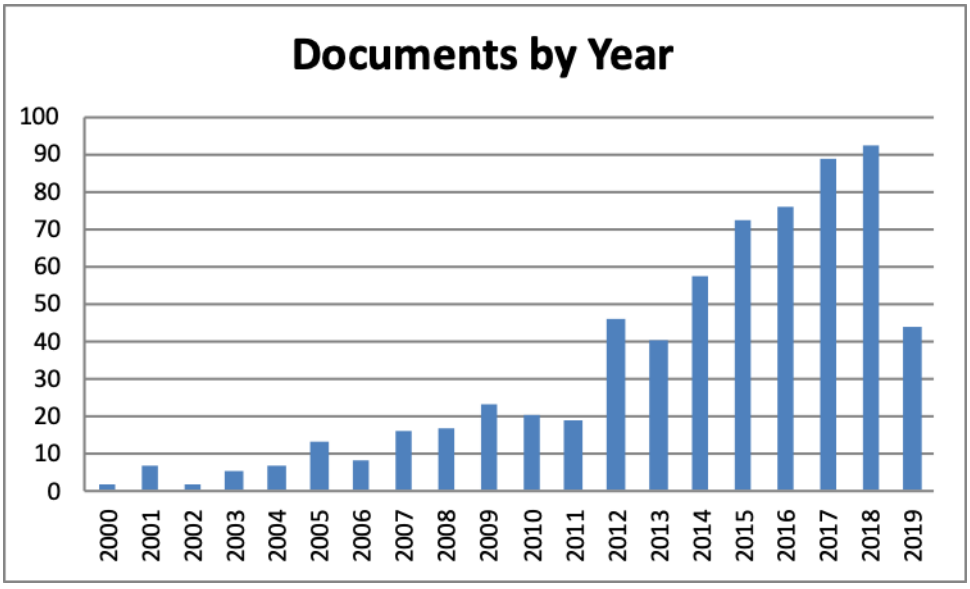

Fig. 1. Documents by year, per July 4, 2019
(Source: Scopus)

The further descriptive analysis provided data on the journal that publishes RDM-HEs-related papers. It showed that journals that published RDM-HEs-related papers were mostly from the library and information science and computer science, as is shown in Table 1.

Table 1. Top 10 Journals that publish RDM-HEs related papers

No.	Title	Documents	Citation
1	Voeb-Mitteilungen	14	9
2	Communications In Computer And Information Science	12	9

3	Ifla Journal	11	71
4	Lecture Notes In Computer Science (Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics)	11	17
5	Grey Journal	9	6
6	Program	7	51
7	Liber Quarterly	7	42
8	D-Lib Magazine	7	24
9	Procedia Computer Science	7	18
10	Ceur Workshop Proceedings	7	3

The identification was continued to find the most prominent journals which published RDM-HEs articles. These journals are presented in Table 2 and illustrated in Figure 2 along with their citations.

Table 2. Journals that publish the most cited papers

No.	Title	Documents	Citation
1	Journal of The American Medical Informatics Association	6	104
2	IFLA Journal	11	71
3	Program	7	51
4	Liber Quarterly	7	42
5	Journal of Academic Librarianship	6	40
6	New Review of Academic Librarianship	6	33
7	D-Lib Magazine	7	24
8	Electronic Library	5	23
9	Journal of The Medical Library Association	5	19
10	Procedia Computer Science	7	18

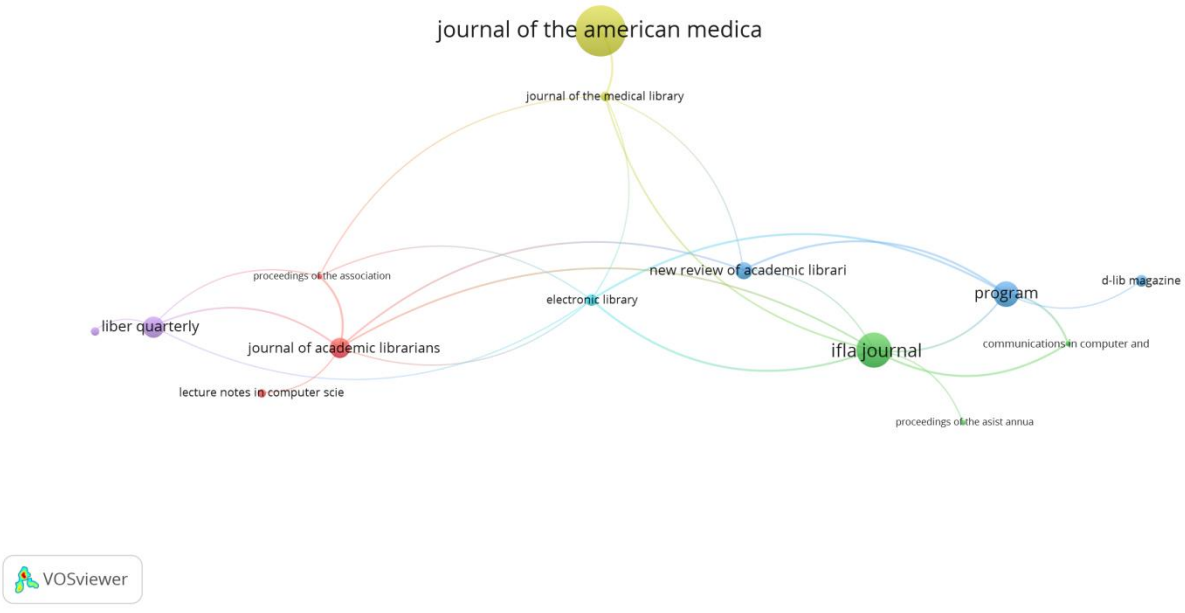

Fig. 2. Prominent Journals on RDM-HEs
(Source: Scopus, 2019)

Journal of The American Medical Informatics Association is a journal that continuously gets high impact factor every year. Their impact factor for the year of 2017 was 4.270, whereas they reached up to 4.292 in 2018. This fact can be one of the reasons that this journal becomes the most cited journal identified in this research.

Further, this journal and the citation list can be useful for academics and researchers in the field as guidance when they are looking for the most cited journals in terms of looking for a place to publish their research.

Another analytical test was conducted to identify the most productive authors along with the numbers they have and the citations they received. The data showed Andrew M. Cox, Riberio and P. Budroni as the most productive authors. The complete list is shown in Table 3.

Table 3. Most Productive authors

No.	Author	Documents	Citations
1	Cox Andrew. M	7	148
2	Ribeiro C.	7	29
3	Budroni P.	7	5
4	Tenopir C.	6	123
5	Allard D.	6	121
6	Da Silva J.R.	6	25
7	Ganguly R.	6	5
8	Solis B.S.	6	5
9	Smith P.L.	6	1
10	Castro J.A.	5	22

(Source: Scopus, 2019)

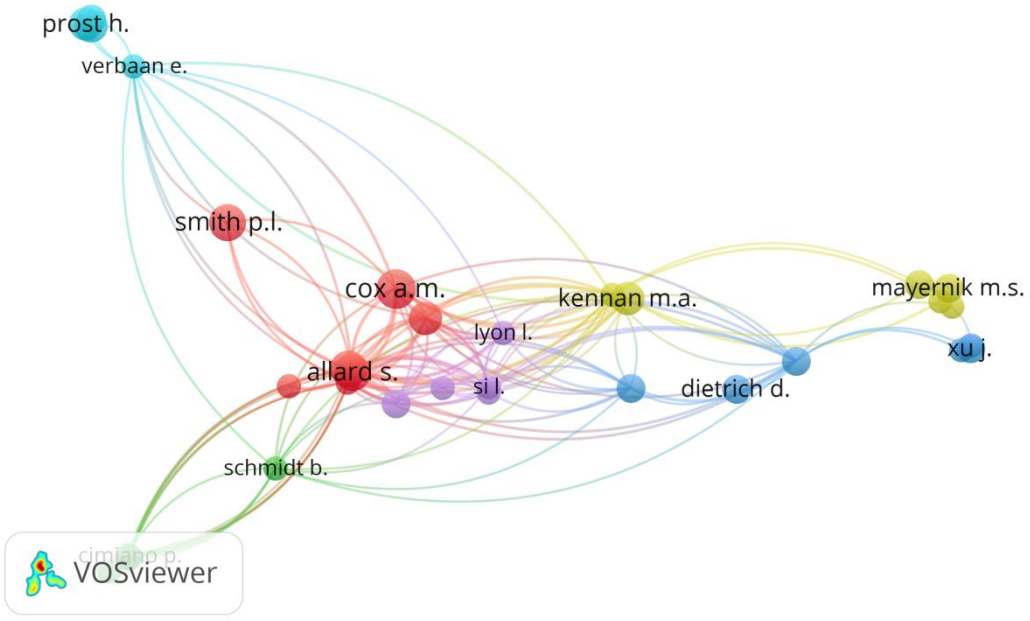

Fig. 3. Authors with the highest number of publications
(Source: Scopus, 2019)

Further analysis was conducted by using co-citation method. Co-citation analysis is a technique to identify papers' clusters that are frequently cited. Such pattern depicts how scientists collectively attribute their work for published work (Chen, Song, and Zhu 2007). Thus, the

visualization mapping provided in Figure 4 highlighting the prominent authors that inspired the field. The authors' influence can be seen from the relative size of each author's depiction (diameter of cycle symbol). Clustering in the figure symbolizes the tight in-between linkage. Based on this visual data, it can be clearly seen that these authors have shaped the body of literature as illustrated in Figure 4.

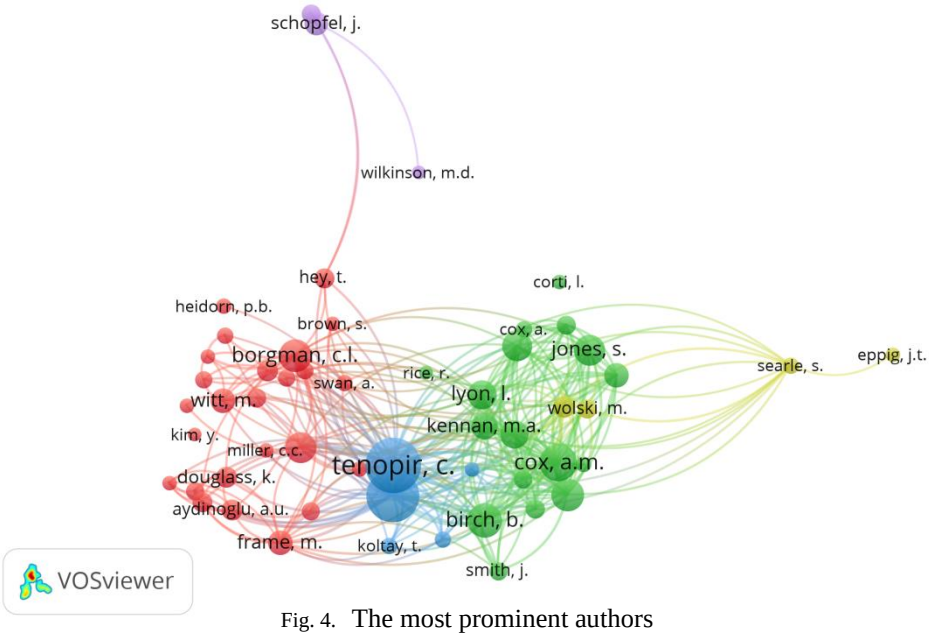

Fig. 4. The most prominent authors
(Source: Scopus, 2019)

Co-citation analysis was also conducted to observe the most influential reference in the field. It was apparent from the data that Tenopir has 3 influential papers (Tenopir, Birch, and Allard 2012), (Tenopir et al. 2014), (Tenopir et al. 2011). Tenopir is considered as the most influential researcher in the field which then makes her the most prominent author in the field. The second author that is considered as the influential author was Corral (Corrall, Kennan, and Afzal 2013), and Cox (Cox and Pinfield 2014). These authors' contributions have led their papers to be key papers in field. The complete list of these influential authors is presented in the Table 4.

Table 4. Most influential papers

No.	Papers	Citations	Total link strength
1	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), pp. 636-674	11	20
2	Cox, A.M., Pinfield, S., (2014) Research Data Management and Libraries: Current Activities and Future Priorities. *Journal of Librarianship and Information Science*, 46 (4), pp. 299-316	10	17

3	Borgman, C.L. (2012) The conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology*, 63 (6), pp. 1059-1078	8	10
4	Akers, K.G., Doty, J. (2013) Disciplinary Differences in Faculty Research Data Management Practices and Perspectives. *International Journal of Digital Curation*, 8 (2), pp. 5-26	6	14
5	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. PLOSone, 6 (6)	5	6
6	Tenopir, C., Sandusky, R.J., Allard, S., Birch, B., (2014) Research Data Management Services In Academic Research Libraries And Perceptions Of Librarians. *Library & Information Science Research*, 36 (2), pp. 84-90	5	6
7	Whitmire, A.L., Boock, M., Sutton, S.C., (2015) Variability In Academic Research Data Management Practices: Implications For Data Services Development From A Faculty Survey. *Program: electronic library and information systems*, 49 (4), pp. 382-407	4	10
8	Witt, M., (2008) Institutional Repositories And Research Data Curation In A Distributed Environment. *Library Trends*, 57 (2), pp. 191-201	4	10
9	Carlson, J., Fosmire, M., Miller, C.C., Nelson, M.S., (2011) Determining Data Information Literacy Needs: A Study Of Students And Research Faculty. *Portal: Libraries And The Academy*, 11 (2), pp. 629-657	4	9
10	Hey, T., Hey, J., (2006) E-science and Its Implications for the Library Community. *Library Hi Tech*, 24 (4), pp. 515-528	4	6

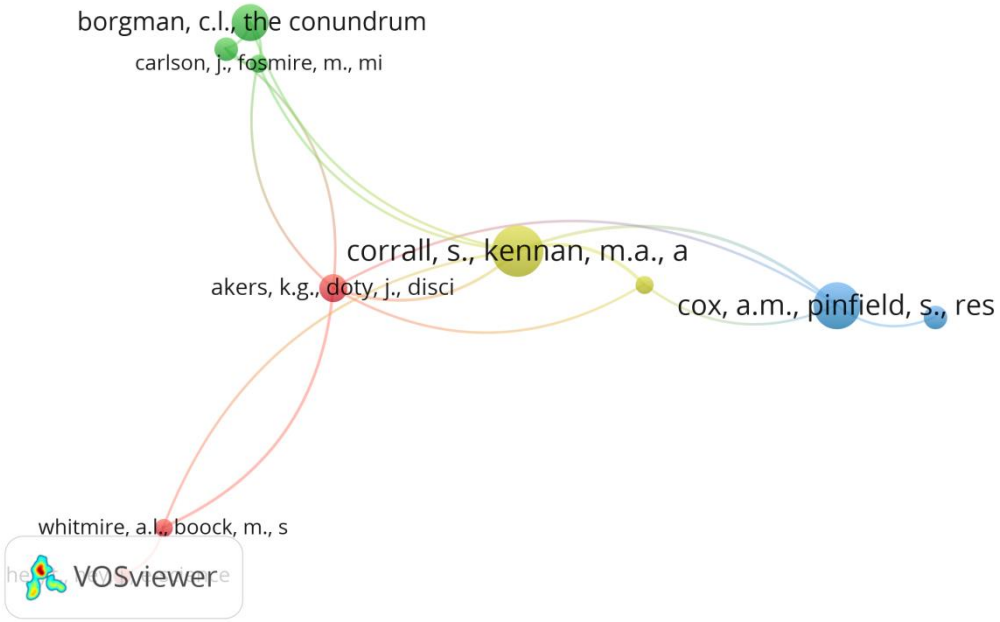

Fig. 5. Most influential papers
(Source: Scopus, 2019)

Following the process, this study identified the research trend by conducting Co-occurance analysis. By using Vosviewer, this process intended to count the frequencies of the terms that occured together for some articles. The number of co-occurrences of two terms was used as a measure of the relatedness of the terms. Therefore a map of key terms was constructed (Eck and Waltman 2017). However, there were a few keywords neglected during computation, i.e. "research data management", "university", "management", "research data", "data management", "universities" and "data", as these are the general terms that can be found in the data sets.

The visual keyword mapping showed the most popular keywords in the field. The result of this computation process was illustrated in the Figure 6. The close relationship that happened between two terms reflects the co-occurrence frequency, thus, the closest terms are located next to each other in the map. This map provides a visualization of trending terms for that year, which also show such topics that are the attention of researchers.

In addition, this visual map serves as a starting point for analyzing research that has been done to date. This also showed the concepts that have gained researchers' attention over time. The map also revealed that the most intense research, which is depicted in the red area, was related to academic libraries, data curation, knowledge management, education, technology, performance.

It is interesting to notice that academic libraries hold an important part in RDM in higher educations environment. This can be seen from the size of the node provided in the Figure 6.

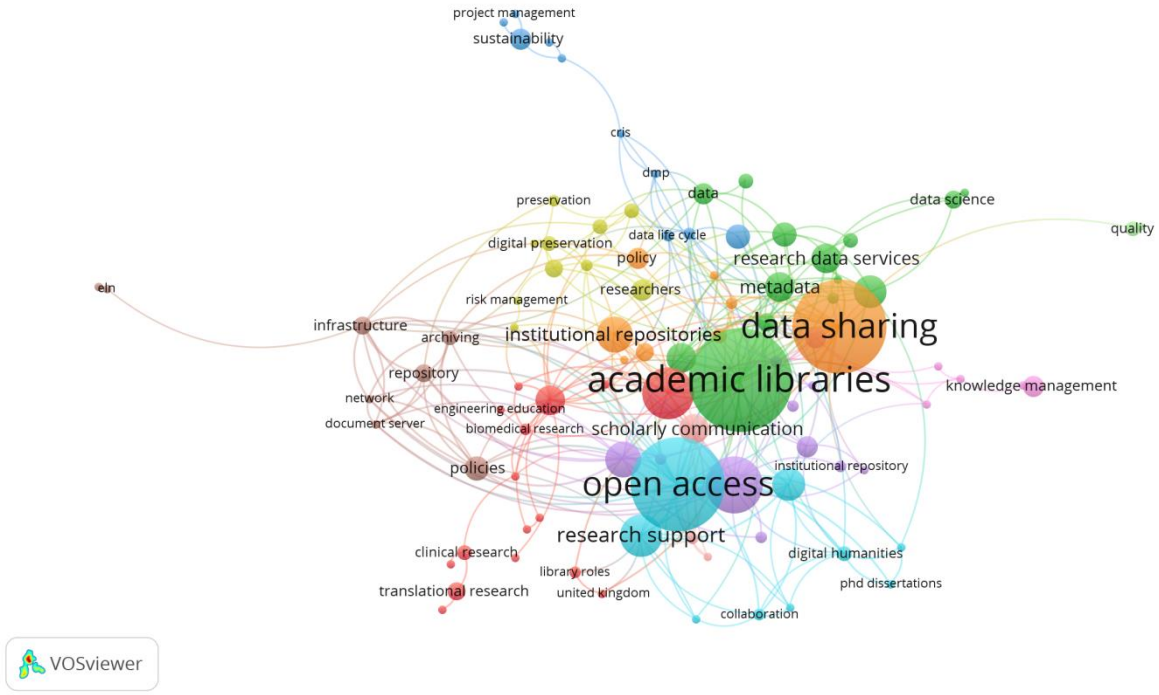

Fig. 6. Keywords Co-occurance
(Source: Scopus, 2019)

The analysis found 16 clusters of topics. The cluster related to academic library (Figure 7) consists of several topics including open access, data sharing, institutional repository, data collection, data services, data management plan, digital repositories, training, and data services.

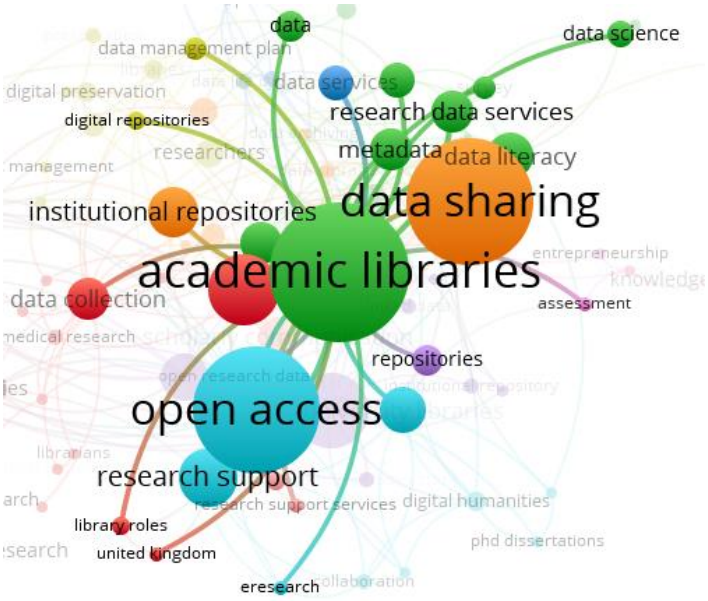

Fig. 7. Prominent topic
(Source: Scopus, 2019)

The cluster also shows some concepts that are related to each other, for example, data curation, e-science, faculty, data management services, and research data services. It reveals such

research trends that reflect unique topics investigated by those studies. From the result shown, it can be seen the research topic was about library services provided by academic libraries. The topics include support services, data management services, data management training and data curation infrastructure provision.

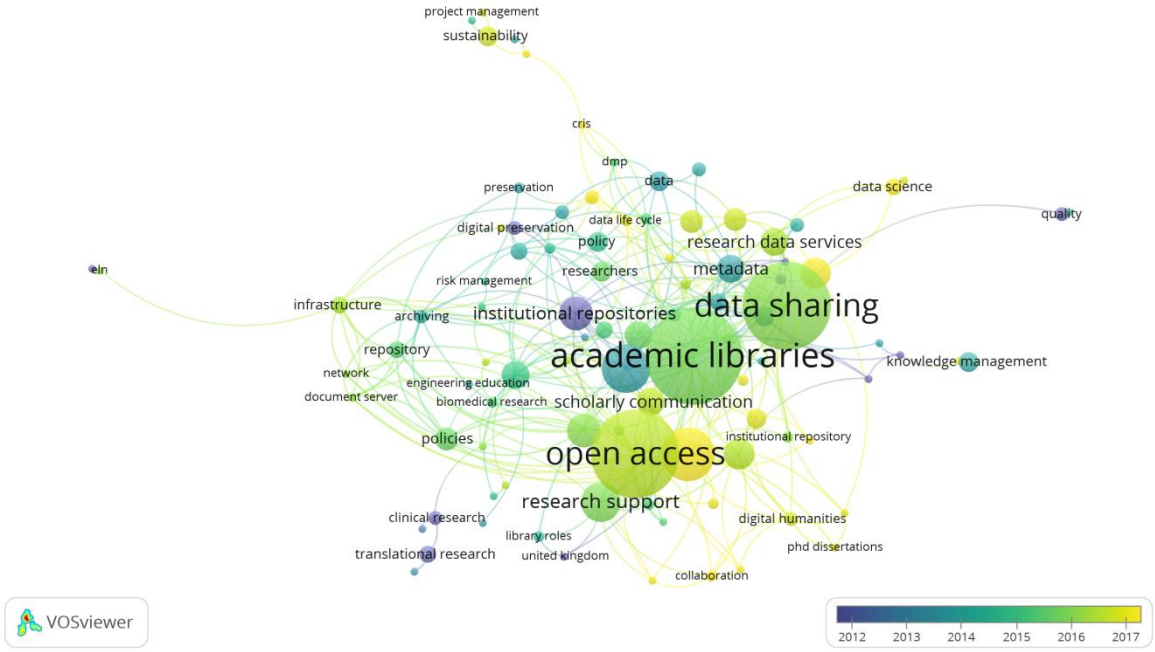

Fig. 8. Keywords Co-occurance by year
(Source: Scopus, 2019)

Analysis using a timeframe view (Figure 8) revealed that there are many new topics such as data science, digital humanities, and open access.

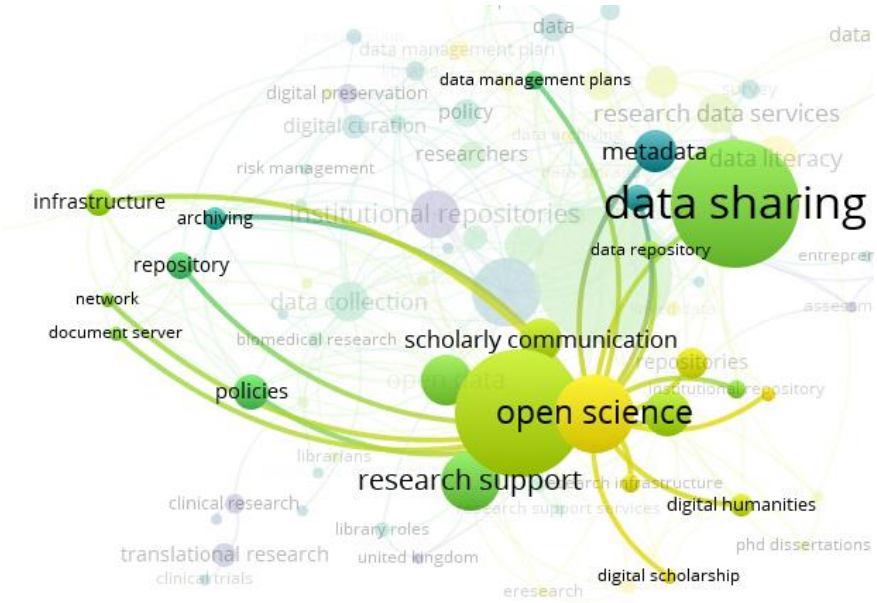

Fig. 9. Recent topics
(Source: Scopus, 2019)

It is interesting to know the recent studies (Figure 9) were related to open science, data science, data literacy, data management plan, digital scholarship. It shows that there are growing discussions among researchers about open data and data literacy worldwide. As this provides information about the development of these issues, it may encourage other researchers to join the discussions and collaborating with other researchers in the field.

4. CONCLUSION

This study found a number of most cited journals from different field of study, as well as the most influential authors that have contributed to the development of knowledge for some research subject. In addition, some topics that were mostly studied and published were also known. They are open access, data sharing, institutional repository, data collection, data services, data management plan, and digital repositories. However, it is important to note that this study identified the trend in RDM-HEs related research that the datasets are limited to papers indexed by Web of Science, and thus this provides limited opportunities to understand RDM-HEs field. On the other hand, this trend-spotting identifies any topic that needs to be explored further.

BIBLIOGRAPHY

Arias-coello, Alicia. n.d. "Management Literacy at Three Spanish Universities."

Chad, Ken, and Suzanne Enright. 2014. "The Research Cycle and Research Data Management (RDM): Innovating Approaches at the University of Westminster." *Insights: The UKSG Journal* 27 (2). doi:10.1629/2048-7754.152.

Chen, Chaomei, Il-yeol Song, and Weizhong Zhu. 2007. "Trends in Conceptual Modeling : Citation Analysis of the ER Conference Papers." *11th International Conference on the International Society for Scientometrics and Informatrics*. CSIC, Madrid, Spain. June 25-27, 189–200.

Choudhury, G. Sayeed. 2008. "Case Study in Data Curation at Johns Hopkins University." *Library Trends* 57: 211–20.

Corrall, Sheila, Mary Anne Kennan, and Waseem Afzal. 2013. "Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research." *Library Trends* 61 (3): 636–74. doi:10.1353/lib.2013.0005.

Cox, Andrew M., and Stephen Pinfield. 2014. "Research Data Management and Libraries: Current Activities and Future Priorities." *Journal of Librarianship and Information Science* 46 (4): 1–18. doi:10.1177/0961000613492542.

Dixon, Supervisor Pat. 2012. "Ewelina Melnarowicz a Case at the Loughborough University Library."

Eck, Nees Jan Van, and Ludo Waltman. 2017. "VOSviewer Manual," no. October.

Feature, Theme. 2011. "Research Data Management at Concordia University : A Survey of Current Practices," 15–18.

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/RDM.pdf.

Hayes, Barrie, James Harroun, and Brenda Temple. n.d. “Data Management and Curation of Research Data in Academic Scientific Research Environments Authors.”

Hey, y Tony, Stewart Tansley, and Kristin Tolle. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Washington: Microsoft Research. file:///C:/Users/Edson.MARLEY/Desktop/4th_paradigm_book_complete_lr.pdf.

Johnson, Vanessa E. 2012. “The Role of Information Professionals in Geoscience Data Management: A Western Australian Perspective.” *LIBRES* 22 (2): 1–23.

MacMillan, Don. 2014. “Data Sharing and Discovery: What Librarians Need to Know.” *Journal of Academic Librarianship* 40 (5). Elsevier Inc.: 541–49. doi:10.1016/j.acalib.2014.06.011.

Markoff, John. 2013. “How to Share Scientific Data.” *New York Times*. http://www.nytimes.com/2013/08/13/science/how-to-share-scientific-data.html?smid=fb-share&_r=1&.

Patel, Dimple, For Authors, and Dimple Patel. 2016. “Research Data Management : A Conceptual Framework.” *Library Review* 65 (4/5): 226–41. doi:10.1108/LR-01-2016-0001.

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): 1–14. doi:10.5281/zenodo.557043.Funding.

Pitchard, Alan. 1969. “Statistical Bibliomegraphy or Bibliometrics.” *Journal of Documentation* 25 (4): 348–49.

Raghuram, Sumita, Philipp Tuertscher, and Raghu Garud. 2010. “Research Note-Mapping the Field of Virtual Work : A Cocitation Analysis.” *Information Systems Research* 21 (4): 983–99. doi:10.1287/isre.1080.0227.

Schneider, Rene. n.d. “Information Literacy in the Workplace.” In *%th European Conference, ECIL 2017*, edited by Serap Kurbanoglu, Joumana Boustany, Sonja Spiranec, Esther Grassian, Diane Mizrachi, and Loriene Roy, 139–47. Switzerland: Springer.

Small, Henry. 1973. “Co-Citation in the Scientific Literature: A New Measurement of the Relationship between Two Documents.” *Journal of the American Societry for Information Science* July-Augus: 265–69.

Solutions, Mercury Project. 2013. “Research Data Management in Practice.” http://www.ands.org.au/__data/assets/pdf_file/0009/394056/research-data-management-in-practice.pdf.

Steeleworthy, Michael. 2014. “Research Data Management and the Canadian Academic Library : An Organizational Consideration of Data Management and Data Stewardship.” *Partnership: The Canadian Journal of Library and Information Practice and Research* 9 (1): 1–12.

Surkis, Alisa, Kevin Read, Carly Strasser, Alisa Surkis, and Kevin Read. 2015. “Research Data Management.” *Journal of Medical Library Association* 103 (3). Baltimore: Medical Library Association: 154–56. doi:10.3163/1536-5050.103.3.011.

Tam, Winnie, Jenny Fry, and Steve Probets. 2014. “The Disciplinary Shaping of Research Data

Management Practices." In *iConference 2014 Proceedings*, 721–28. doi:10.9776/14338.

Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. 2011. "Data Sharing by Scientists: Practices and Perceptions." Edited by Cameron Neylon. *PLoS ONE* 6 (6): e21101. doi:10.1371/journal.pone.0021101.

Tenopir, Carol, Ben Birch, and Suzie Allard. 2012. "Academic Librarues and Research Data Services." doi:http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf.

Tenopir, Carol, Robert J Sandusky, Suzie Allard, and Ben Birch. 2014. "Research Data Management Services in Academic Research Libraries and Perceptions of Librarians." *Library and Information Science Research* 36: 84–90. doi:10.1016/j.lisr.2013.11.003.

Tripathi, Manorama, Archana Shukla, and Sharad Kumar Sonker. 2017. "Research Data Management Practices in University Libraries : A Study." *DESIDOC: Journal of Library and Information Technology* 37 (6): 417–24. doi:10.14429/djlit.37.6.11336.

Vogel, Rick, and Wolfgang H. Güttel. 2013. "The Dynamic Capability View in Strategic Management: A Bibliometric Review." *International Journal of Management Reviews* 15 (4): 426–46. doi:10.1111/ijmr.12000.

Wang, Jian-Jun, Haozhe Chen, Dale S. Rogers, Lisa M. Ellram, and Scott J. Grawe. 2017. "A Bibliometric Analysis of Reverse Logistics Research (1992-2015) and Opportunities for Future Research." *International Journal of Physical Distribution & Logistics Management* 47 (8): 666–87. doi:10.1108/IJPDLM-10-2016-0299.

.