



## PUBLICATION OF RESEARCH DATA MANAGEMENT IN OPEN ACCESS JOURNAL ANALYSIS BASED ON SCOPUS DATA

Tupan<sup>1\*</sup>, Kamaludin<sup>2</sup>

<sup>1,2</sup>The Center for Scientific Data and Documentation, Indonesia Institute of Sciences

\*Correspondence: [tupan71210@gmail.com](mailto:tupan71210@gmail.com)

Submission: 02-06-2020; Review: 30-08-2020; Accepted: 07-09-2020; Revised: 30-10-2020

### ABSTRACT

The study aims to determine: (1) the number of open access resources for research data management publications indexed by Scopus, including the year of publication, source of publication, authors, institutions, countries, types of documents and funding agencies; (2) mapping research data management based on keywords. The results of the study showed that the number of open access resources for research data management publications has started since 1981 and the number has continued to increase starting in 2014 and the highest number occurred in 2019, namely 49 publications. The most publicized journals that open access to research data management was the Data Science Journal, which was 11 publications. The most productive author of conducting research data management publications was Cox, A.M. and Pinfield, S. The largest institutions contributing to the publication of open access research data management were the University of Toronto and New York University. The countries that contributed the most were the United States with 50 publications, then China with 38 publications. The most open access research data management in the form of articles as many as 107 and 37 conference paper publications. The institutions that provided the most funding sponsors were the Deutsche Forschungsgemeinschaft and the National Science Foundation. The results of keyword mapping with VOSViewer showed that big data, research data management, information management, data management, medical research topics, software, information processing, and metadata were the most researched topics.

### ABSTRAK

Kajian bertujuan untuk mengetahui: (1) jumlah sumber daya akses terbuka publikasi manajemen data penelitian yang terindeks Scopus yang mencakup tahun publikasi, sumber publikasi, pengarang, lembaga, negara, jenis dokumen dan lembaga pendanaan; (2) pemetaan manajemen data penelitian berdasarkan kata kunci. Hasil kajian menunjukkan bahwa jumlah sumber daya akses terbuka publikasi manajemen data penelitian sudah dimulai dari tahun 1981 dan jumlahnya terus meningkat mulai tahun 2014 dan jumlahnya terbanyak terjadi pada tahun 2019, yaitu sebanyak 49 publikasi. Jurnal terbanyak mempublikasikan akses terbuka manajemen data penelitian adalah Data Science Journal, yaitu sebanyak 11 publikasi. Pengarang paling produktif melakukan publikasi manajemen data penelitian adalah Cox, A.M. dan Pinfield, S. Lembaga terbanyak berkontribusi dalam publikasi akses terbuka manajemen data penelitian adalah University of Toronto dan New York University. Negara terbanyak berkontribusi adalah Amerika Serikat dengan jumlah publikasi sebanyak 50, kemudian disusul China sebanyak 38 publikasi. Publikasi akses terbuka manajemen data penelitian yang terbanyak dalam bentuk artikel sebanyak 107 dan Conference paper 37 publikasi. Lembaga yang paling banyak memberikan sponsor pendanaan adalah Deutsche Forschungsgemeinschaft dan National Science Foundation. Hasil pemetaan kata kunci dengan VOSViewer menunjukkan bahwa big data, research data management, information management, data management, medical research topic, software, information processing, metadata adalah topik yang paling banyak diteliti.

**Keywords:** Research data management; Open access; Journal; Scientific publication; Data science; Indexing journal; Scopus

## 1. INTRODUCTION

Data management is a comprehensive approach to effectively sharing data, managing, curating, preserving, and reusing the rapidly developing data generated from the research process. Research data is an integral part of the knowledge base that underpins scholarship, provides insights into our complex world and informs decisions about the present and future. Research data is very diverse and can include observational, computational, and experimental data. The types of data can be

spatial, temporal, resulting from instruments, models, and simulations to images, videos, etc. (Charbonneau, 2013). A similar statement was also conveyed by Angus & Tedds (2011), Research Data Management (RDM) is an activity related to how individuals manage data in research or scientific writing activities, this activity includes the collection, organization, management, storage, security, backup, maintenance and sharing of data, and the same is said by Borghi, et al. (2018) that RDM is a term that includes activities related to the storage, organization, documentation and dissemination of research data.

RDM is central to the open science agenda in Europe in the field of research and innovation. Open Science is the transformation, opening and democratization of science, research and innovation, to make science more efficient, transparent, and interdisciplinary, to change the interaction between science and society, and allow for wider community impact and innovation. Open research data is one of the key components of the new ecosystem that has emerged from standards and services, and priorities from the European Commission include raising awareness about data management, infrastructure and dataset interoperability, and data reuse. In France, the University President Conference puts the issue of conservation Research data and data sharing are their priorities during the annual conference held in 2015 by the Ministry of Higher Education and Research to support and promote related actions, and all public research organizations such as the CNRS (Center National de la Recherche Scientific) contribute to the development of data infrastructure and the main problem and objective are the same as in Germany or other countries, namely the preservation of long-term scientific outputs and open global data policies to increase transparency and stimulate research, innovation, and economic activity.

Consensus on the complexity and diversity of fields. Not only are research data difficult to define but their handling and processing are highly dependent on clear institutional, disciplinary and behavioural practices, the term research data must always be seen with certain scientific disciplines. All requirements for management and the long-term availability of research data must be distinguished from each other in terms of aspects and general and specific disciplinary solutions (and) so far, there is no general agreement on the definition of digital curation, not only in Germany but also internationally (Schöpfel & Prost, 2016).

Based on the aforementioned problems, a study of publication analysis on RDM was conducted in open access journals indexed by Scopus. The study aims to find out the number of publications by year; top ten journals that publish open access on RDM; the productivity of the authors of open access journal publications on RDM; the number of RDM open access publications by the institution; the number of RDM open access journal publications; the number of open access publications for RDM by document type; the number of open access publications based on funding sponsors; mapping and publishing trends in open access on RDM.

## 2. LITERATURE REVIEW

In the scientific community today there has been a significant shift towards open science, which prioritizes collaboration and open access in scientific research. Management and sharing of research data have become an important part because of the valuable role of researchers in improving science through work produced previously by reusing data produced by other researchers. As stated by Tenopir et al. (2011), research data is the infrastructure of science. Research data includes every piece of data obtained and generated during the research process and may consist, among others, text, spreadsheets, questionnaires, photos, films, test responses, slides, laboratory notes, statistics, observations, experimental results, measurements, samples, algorithms, scripts and workflows.

ÜNAL (2011) said that RDM is about organizing data, from entry into the research cycle to the dissemination and archiving of valuable results. Beyond the acquisition of their specific data,

related to the investigation, researchers have a special role to play in RDM, which requires a clear understanding of, and a set of applied skills related to, their role in the life cycle of research data. The purpose of this paper was to examine the level of awareness and understanding of various issues, challenges, and training needs related to RDM for researchers at Hacettepe University, Turkey. Preliminary findings indicate a very low level of data literacy, RDM awareness, knowledge, and skills among the participants. The majority of researchers are unsure about issues related to RDM. The findings also showed that, while participants did not have formal training in RDM-related subjects, they were willing to take part in the training, if provided. The majority of participants agreed that the university should be the body responsible for RDM.

The results of previous studies on data management research that have been carried out include Akhoun & Ganaie (2018) through the studied of data management studies in open access journals in developed countries comparisons between the United States and the United Kingdom. This study would assist researchers in exploring the number of open access titles in different subject areas. The results of the study showed that the contribution of developed countries to the issuance of open access was very significant. The United States and Britain took the lead in publishing open access journals indexed on Scopus.

Corrall, Kennan, & Afzal (2013) conducted a study of the role of academic libraries in the four countries surveyed in providing bibliometric support using quotation reports and impact calculations for academic units and at the institutional level as well as training and guidance for individual researchers and research group. The results showed that the level of library involvement at this time was still low in RDM. The bibliometric method among the libraries surveyed showed that there were plans for significant expansion of services with the support offered and the audience identified seems to have prioritized providing technology, infrastructure and tools, support for storing data in institutional repositories, and developing institutional policies for managing data.

Onyancha (2016) examined the need to share research data in the context of current world trends in open access scientific publishing. The study was conducted by exploring the status of research data publication in Sub-Saharan Africa (SSA) as a determinant of how research data was shared among researchers in the region and internationally. The results showed that articles were the most widely published type of publication in the form of open access. Lone & Rather (2008) conducted a study to evaluate India's contribution to making this intellectual output accessible to everyone by publishing it in open access resources such as open access journals and filing them in an open access archive or repository.

The results showed that India continues to contribute to the Open Access literature because several key institutions, especially in the fields of science and technology, provide Open Access to their research publications. India's position in terms of the number of journals in the Directory of Open Access Journals (DOAJ) was 7th, far ahead of countries such as China, Australia, and Japan and shares the 10th position with Sweden and Spain in the Directory of Open Access Repositories (Open DOAR) in terms of the number of repositories in the world. In terms of the number of journals, India ranks number 7 in the Directory of Open Access Journals (DOAJ), far ahead of countries such as China, Australia and Japan. The first position that contributes to open access was the United States. Xu & Yu (2019) conducted bibliometric analysis and visual analysis methods to analyze the publication of big data indexed in the Science Citation Index (SCI) and the Social Science Citation Index (SSCI). One of the analyzes conducted was to analyze the publication of big data based on countries, journals, and research institutions that contribute the most. On the other hand, the emergence of keywords, authors of publications, and current research hotspots and future development trends were also examined.

### 3. METHOD

The study was conducted by searching through the Scopus database with keywords research; data; and management with a period of 1981-2020. The search was conducted in April 2020. The search results obtained were 1070 documents. Of the total search results, there were around 170 documents that could be accessed openly which were then used as research data. Research data that could be accessed openly were grouped based on parameters that were following the research objectives. The results of grouping the data were then analyzed using Ms. Excel. The results of the analysis were then outlined in the form of results and discussion. Whereas for keyword mapping VOSViewer software was used.

### 4. RESULTS AND DISCUSSIONS

#### 4.1 The Number of Open Access Resources on RDM Publications Indexed By Scopus

##### 4.1.1 Number of publications by year

The search results through the Scopus database showed that the publication of 1070 research data management documents that could be accessed openly was only about 170 documents (15.89%). RDM publications that could be accessed have started since 1981 and the number continues to increase starting in 2014 and the most number occurred in 2019 with 49 publications. The number of open access publications management of research data that could be accessed openly can be seen in Table 1 and Figure 1 below.

Table 1. Number of Open Access on RDM Publications

Year	Number of Publications
2020	8
2019	49
2018	25
2017	21
2016	18
2015	10
2014	15
2013	5
2012	2
2011	3
2010	2
2009	3
2008	1
2007	2
2002	1
1998	1
1995	1
1990	1
1982	1
1981	1
Total	170

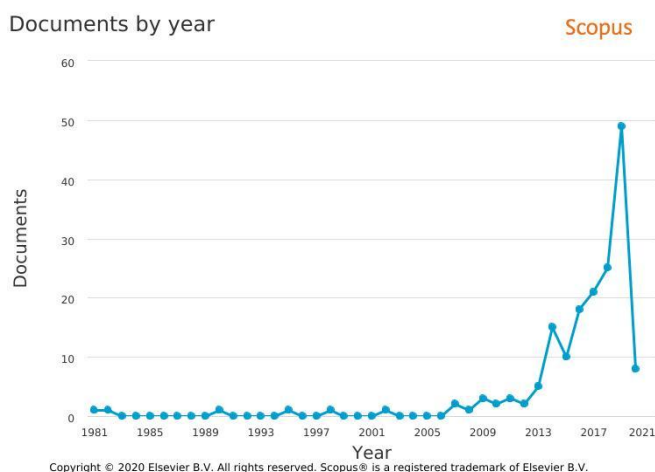


Figure 1. Number of open access on RDM publications

##### 4.1.2 The top ten journals that publish open access on RDM

The top ten journals that had the most open access on RDM were the Data Science Journal, with 11 publications. The next sequence was Journal of Physics Conference Series, SPRS International Journal of Geo-Information, UInsights The Uksg Journal, Liber Quarterly, Procedia Computer Science dan Bioinformatics. Data Science Journal. Data Science Journal was a journal with the title Q3 with H Index 17. The eleven publications of the Data Science Journal could be

accessed openly. Details of the publication published in the Data Science Journal from the manual analysis were identifying and implementing relevant RDM services for the library at the University of Dodoma, Tanzania.

The impact of targeted data management training for field research projects - A case study, Teaching RDM for students, Teaching RDM for students, Expanding the RDM services portfolio at Bielefeld University according to the three-pillar principle towards data FAIRness, Supporting the interdisciplinary, long-term research project ‘patterns in soil-vegetation-atmosphere-systems’ by data management services, Enhancing the RDM of computer-based educational assessments in Switzerland, Science metadata management, interoperability and data citations of the National Institute of Polar Research, Japan, RDM in research institutions in Zimbabwe, Metadata management at the polar data centre of the National Institute of Polar Research, Japan, Data and structure characteristics analysis for a scientific research e-management platform in the humanities and social sciences. A complete list of the top ten journals that publish open access journals on research data management can be seen in Table 2 and Figure 2.

Table 2. The Top Ten Journals that Publish Open Access on RDM

No.	Journal Name	Number of Publications
1	Data Science Journal	11
2	Journal of Physics Conference Series	9
3	ISPRS International Journal of Geo-Information	5
4	UInsights the Uksg Journal	5
5	Liber Quarterly	5
6	Procedia Computer Science	5
7	Bioinformatics	4
8	College and Research Libraries	4
9	Desidoc Journal of Library and Information Technology	4
10	Information Services and Use	4

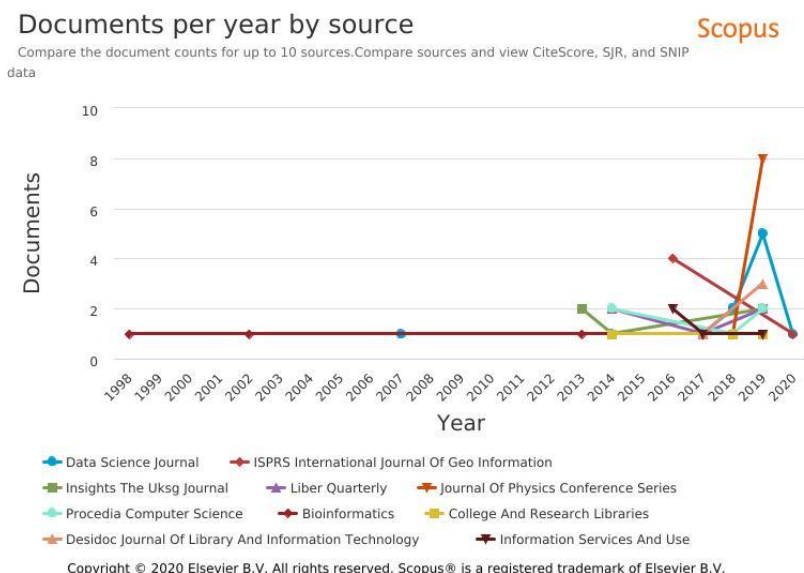


Figure 2. The top ten journals that publish open access on RDM

#### 4.1.3 Author productivity open access journal publications on RDM

The ten most productive authors conduct to RDM publications with ten recommendations. The ten most productive authors were Cox, A.M. and Pinfield, S. each with 3 publications. The ten most

productive authors can be seen in Table 3 and Figure 3 below. The three publications were written by Cox, A. M in collaboration with Pinfield, S. was development in RDM in academic libraries; Towards an understanding of research data service maturity, Moving a brick building: UK libraries coping with RDM as a ‘wicked’ problem, and RDM and libraries: relationships, activities, drivers and influences.

Table 3. The Productivity of the Author of Open Access on RDM

No	Author	Number of Publications
1	Cox, A.M.	3
2	Pinfield, S.	3
3	Bahls, T.	2
4	Cimiano, P.	2
5	Clements, A.	2
6	Hoffmann, W.	2
7	Kadokura, A.	2
8	Kanao, M.	2
9	Laroche, C.	2
10	Okada, M.	2

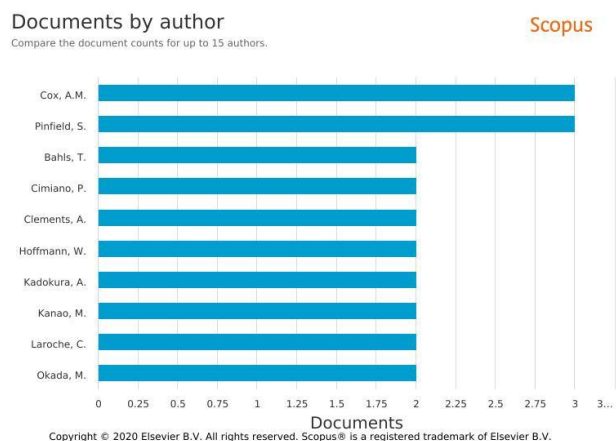


Figure 3. The productivity of the author of open access on RDM

#### 4.1.4 Number of open access publications for RDM by the institution

Table 4 and Figure 4 showed the ten institutions that most open research data management open access journals. At the first level was the University of Toronto and New York University each with 4 publications, then followed by University of Sheffield, Purdue University, Chinese Academy of Sciences, Universität Göttingen, Medizinische Fakultät der Universität Greifswald, Karlsruhe Institute of Technology, Deutsches Zentrum für Herz-Kreislauf-Forschung e. V. each of 3 publications. In the next sequence was Jibe Electric Power Company Limited Metering Centre with 2 publications. This was supported by Xu, Zeshui and & Yu, Dejian (2019) who conducted a study on the analysis of big data publications by country, journal, and the most contributing research institutions. The results showed that the University of California System and Chinese Academy of Sciences were the two institutions that contribute the most to big data publications.

Table 4. Number of Open Access Publications for RDM by the Institution

No	Institution name	Number of Publications
1	University of Toronto	4
2	New York University	4
3	University of Sheffield	3
4	Purdue University	3
5	Chinese Academy of Sciences	3
6	Universität Göttingen	3
7	Medizinische Fakultät der Universität Greifswald	3
8	Karlsruhe Institute of Technology	3
9	Deutsches Zentrum für Herz-Kreislauf-Forschung e. V.	3
10	Jibe Electric Power Company Limited Metering Centre	2



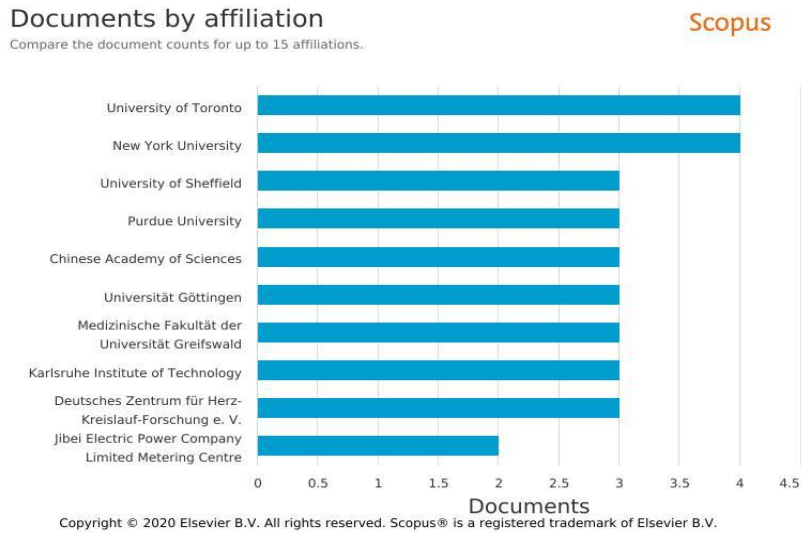


Figure 4. Number of open access publications for RDM by the institution

#### 4.1.5 Number of open access journal on RDM by country

The results of the analysis based on Table 5 and Figure 5 showed that there were ten countries that most contribute to the publication of open access research data management. The first level of countries that contributed the most was the United States with 50 publications, followed by China with 38 publications, Germany 22 publications, 21 English publications, 10 Australian publications, 9 Canadian publications, 7 Indian publications, 6 Dutch publications, 5 French publications and Indonesia 4 publications. This was consistent with the research of Lone, Fayaz and Rather, Rafiq and Shah, Gh Jeelani (2008) who conducted a study of Indian contributions in publishing open access journals in DOAJ. The results showed that the United States contributed the most in open access journal publishing. This was supported by Xu, Zeshui and Yu, Dejian (2019) in the study of the analysis of big data publications by country, journal, and research institutions that contribute the most. The results showed that the United States and China were the two countries that have contributed the most to the publication of big data. Indonesia was in 10th place with 4 publications that could be accessed openly namely designing engineering data management system in research and development company, management of potential data on websites for communicating research in education issue, strategy for RDM services in Indonesia dan The Role of Science in the Management of Biodiversity: A Case of Stingrays (*Dasyatidae*) Research to Provide Basic Data for Aquatic Fauna Conservation in South Sumatra.

Table 5. Number of Open Access Journal RDM by country

No	Country name	Number of Publications
1	United States	50
2	China	38
3	Germany	22
4	United Kingdom	21
5	Australia	10
6	Canada	9
7	India	7
8	Netherlands	6
9	France	5
10	Indonesia	4

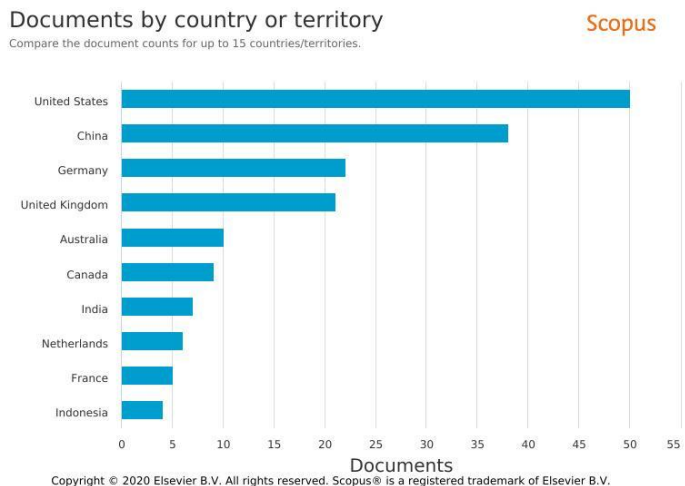


Figure 5. Number of open access journal RDM by country

Akhoon & Ganaie (2018) in a study of data management in open access journals in developed countries comparing the United States and the United Kingdom. The results of the study showed that the contribution of developed countries to the issuance of open access was very significant. The United States and Britain take the lead in publishing open access journals indexed on Scopus.

#### 4.1.6 Number of publications based on open access RDM document type

Analysis based on document types such as in Table 6 and Figure 6 showed that the highest number of open access research data management was in the form of articles, followed by Conference paper 37 publications, Review 13 publications, Editorial 5 publications, and Note 4 publications. This was following research from Bosire Onyancha (2016) who examines the need to share research data in the context of current world trends in open access scientific publishing. The study was conducted by exploring the status of research data publication in Sub-Saharan Africa (SSA) as a determinant of how research data was shared among researchers in the region and internationally. The results showed that articles were the most widely published type of publication in the form of open access.

Table 6. Number of Publications Based on Open Access RDM Document Type

No	Document Type	Number of Documents
1	Article	107
2	Conference Paper	37
3	Review	13
4	Editorial	5
5	Note	4
6	Erratum	2
7	Short Survey	2
8	<b>Total</b>	<b>170</b>

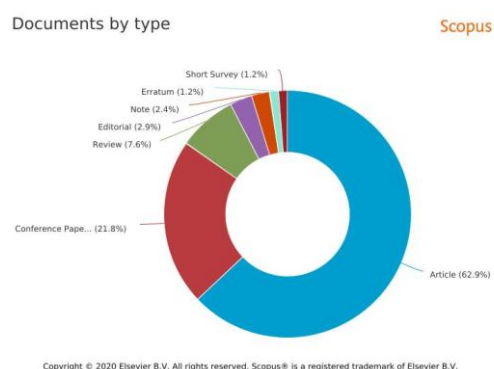


Figure 6. Number of publications based on open access RDM document type

#### 4.1.7. Number of open access publications based on funding sponsors

Based on the analysis of Table 7 and Figure 7 showed that the institutions that most provide funding sponsors were the Deutsche Forschungsgemeinschaft and the National Science Foundation each with 7 publications, then followed by the National Institutes of Health 6 publications, Science and Technology Foundation of the State Grid Corporation of China 3 Publications. The Indonesian institution that sponsors open access research data management publications was the University of Indonesia.

Table 7. Number of Open Access Publications Based on Funding Sponsors

No	Sponsoring Institution	Number of Publications
1	Deutsche Forschungsgemeinschaft	7
2	National Science Foundation	7
3	National Institutes of Health	6
4	Science and Technology Foundation of State Grid Corporation of China	3
5	Australian National Data Service	2
6	Central University of Finance and Economics	2
7	European Commission	2
8	Horizon 2020 Framework Programme	2
9	National Natural Science Foundation of China	2
10	U.S. Department of Defense	2



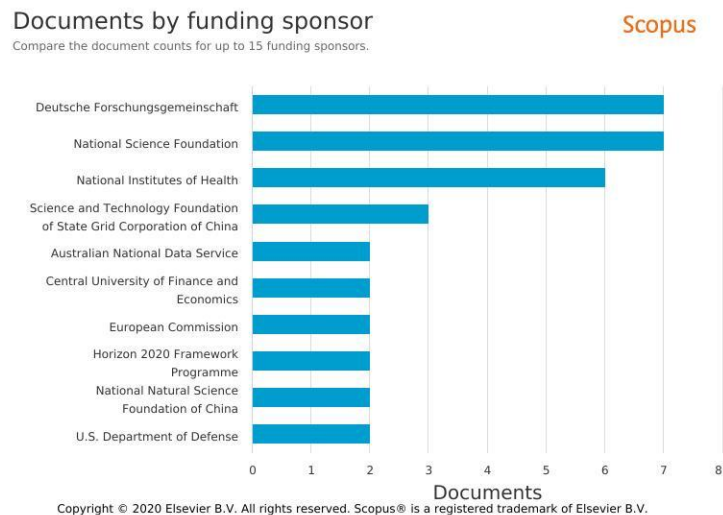


Figure 7. Number of open access publications based on funding sponsors

## 4.2 Mapping and Publishing Trends in Open Access on RDM

Mapping analysis and publication trends of open access to research data management using network visualization, overlays, and density by using VOSViewer version 1.6.14 to find out the network map that exists between articles from the metadata that had been downloaded. Network map based on network visualization showed divided into three clusters as in Figure 8 below. Cluster 1 was red consisting of 18 keywords that often appear and form interrelated circles. The eighteen keywords were big data, data analysis, data management, data mining, data sharing, data technologist, decision making, human resources management, information services, information management, information system, information use, libraries metadata, research and development management, research data, and RDM.

The two green cluster consists of 17 keywords namely clinical research, computer program, database, database management, databases, factual, factual database, information processing, knowledge management, medical informatics, methodology, priority journal, quality control, reproducibility, research, review, software, dan statistical analysis. The blue cluster consists of 14 keywords, namely biomedical research, data collection, education, information retrieval, information storage and retrieval, library, major clinical study, medical research, middle-aged, organization management, procedures, and scientist.

This was consistent with the study of Corral, Kennan, & Afzal (2013) who conducted a study of the role of academic libraries in the four countries surveyed in providing bibliometric support using quotation reports and impact calculations for academic units and at the institutional level as well as training and guidance for individual researchers and research groups. The results of bibliometric studies among the libraries surveyed showed that there were plans for significant expansion of services with support offered and the audience identified seemed to have prioritized providing technological assistance, infrastructure and tools, support for storing data in institutional repositories, and developing institutional policies for manage data.

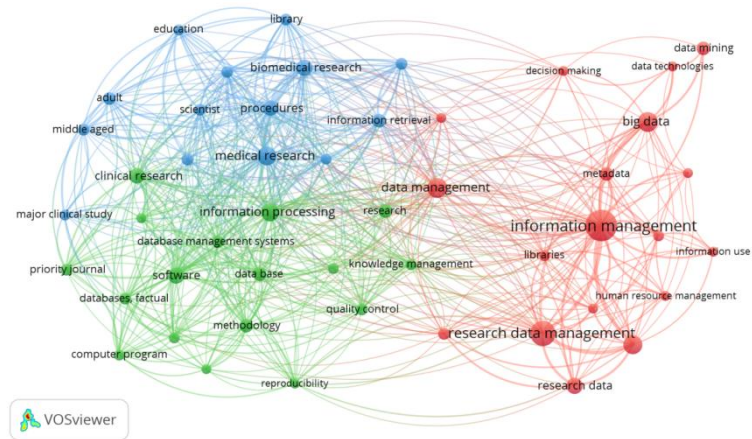


Figure 8. Map of the research data management network

Metadata analysis using VOSViewer also results in overlay visualization. In Figure 9 the overlay visualization showed that the colour of a node represents a keyword while the colour of a node indicates the year of publication in which the keyword was published. The darker colours on the node indicate that the topic had long been done. Figure 9 showed that information system, big data, data analysis, data management, data mining, data sharing, data technologist, decision making, human resources management, information services, information use, libraries metadata, research and development management, research data, RDM, and data technologist were topics discussed between 2013 and 2019. The topics were biomedical research, data collection, education, information retrieval, information storage and retrieval, library, major clinical study, medical research, middle-aged, organization management, procedures, and scientist were discussed from 2005 to 2016.

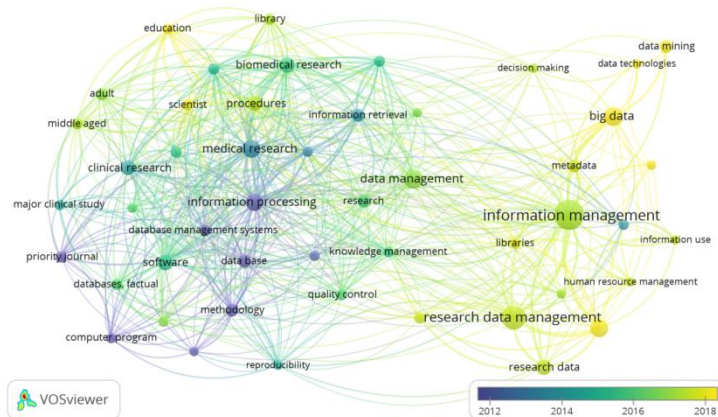


Figure 9. Overlay network map of RDM

Analysis based on visualization of density as shown in Figure 10 showed that big data, research data management, information management, data management, medical research, software, information processing, and metadata which are yellow areas were the most researched topics. The topics in the green areas such as computer program, database management, databases, factual, factual database, knowledge management, medical informatics, methodology, priority journal, quality control, reproducibility, research, review, and statistical analysis were topics that have not much researched.

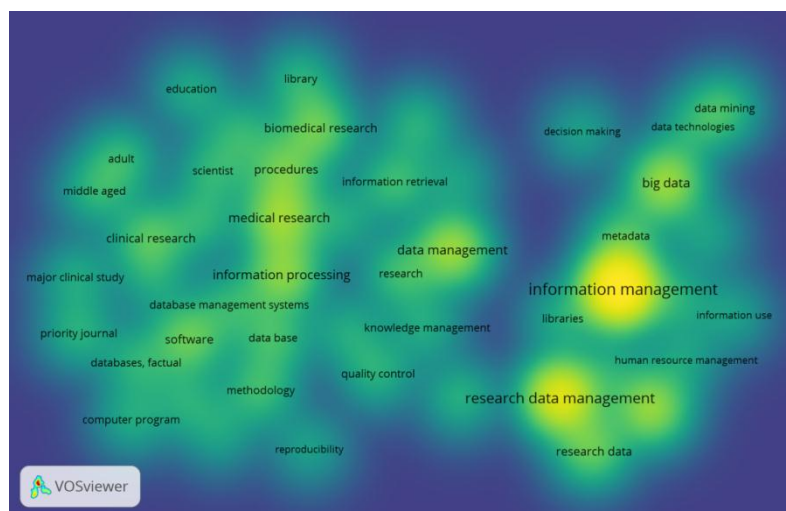


Figure 10. Density network map of RDM

## 5. CONCLUSION

Based on the results and discussion, it can be concluded that the publications of research data management that could be accessed openly have started from 1981 and the number continues to increase starting in 2014 and the most number occurred in 2019, namely 49 publications. The top journal that most publicizes open access to research data management was the Data Science Journal, which was 11 publications. The most productive author of conducting research data management publications was Cox, A.M. and Pinfield, S. The largest institutions contributing to the publication of open access research data management were the University of Toronto and New York University. The countries that contributed the most were the United States with 50 publications, then China with 38 publications. The most open access research data management research in the form of articles as many as 107 and 37 Conference paper publications. The institutions that provide the most funding sponsors were the Deutsche Forschungsgemeinschaft and the National Science Foundation. The results of density visualization with VOSViewer show that big data, research data management, information management, data management, medical research, software, information processing, and metadata were the most researched topics. The implication of this study is to assist researchers in exploring the number of open access titles in a variety of different subjects.

## REFERENCES

- Akhood, I.H., Ganaie, S.A. & Khazir, K. 2018. Research Data Management in Open Access Journals by Developed Countries. *5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ETTLIS)*, 116–20.
- Angus, W. & Tedds, J. 2011. Making the Case for Research Data Management Introduction – Doing More with Less Funders’ Data Policies. JISC, September.
- Borghi, J.A., Abrams, S., Lowenberg, D., Simms, S. & Chodacki, J. 2018. Support Your Data : A Research Data Management Guide for Researchers. *Research Ideas and Outcomes*, 4, 1–13. <https://doi.org/10.3897/rio.4.e26439>.
- Charbonneau, D.H. 2013. Strategies for Data Management Engagement. *Post-Print, Medical Reference Services Quarterly*, 32(3), 1–13.
- 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.
- Lone, F. & Rather, R. 2008. Indian Contribution to Open Access Literature: A Case Study of DOAJ

& OpenDOAR. *Chinese Librarianship: An International Electronic Journal*, 29, 1–10.

- Onyancha, O.B. 2016. Open Research Data in Sub-Saharan Africa: A Bibliometric Study Using the Data Citation Index. *Corenell for the Development of Social Science Research in Africa*, 1–18.
- Schöpfel, J. & Prost, H. 2016. Research Data Management in Social Sciences and Humanities : A Survey at the University of Lille (France). *LIBREAS: Library Ideas*, 29, 98–112.
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M. & Frame, M. 2011. *PLoS ONE*, 6(6), 1–21. <https://doi.org/10.1371/journal.pone.0021101>.
- Ünal, Y. & Kurbanoglu, S. 2011. Research Data Management Practices of Academic Researchers in Turkey. *Presentation Paper for the 2017 International Conference on Knowledge Management*.
- Xu, Z. & Yu, D. 2019. A Bibliometrics Analysis on Big Data Research (2009–2018). *Journal of Data, Information and Management*, 3–15. <https://doi.org/10.1007/s42488-019-00001-2>.