8th International Conference on Libraries, Information & Society
18th – 19th August 2021
Kuala Lumpur Malaysia

UNIVERSITI MALAYA

Mapping the knowledge domains of research data management: A co-occurrence analysis

Mohd Ikhwan Ismail
Hamzah Sendut Library, Universiti Sains Malaysia, Penang, MALAYSIA
Faculty of Computer Science & Information Technology,
Universiti Malaya, Kuala Lumpur, MALAYSIA
ikhwanismail@usm.my

Abrizah Abdullah
Department of Library & Information Science, Faculty of Arts and Social Sciences, Universiti Malaya, Kuala Lumpur, MALAYSIA
abrizah@um.edu.my

Samsul Farid Samsuddin
Department of Library & Information Science, Faculty of Arts and Social Sciences, Universiti Malaya, Kuala Lumpur, MALAYSIA
samsulfarid@um.edu.my

8th International Conference on Libraries, Information & Society
18th – 19th August 2021
Kuala Lumpur Malaysia

UNIVERSITI MALAYA

Introduction

Research data management (RDM), which is part of the research process, which concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results

A popular research topic in the library and information science (LIS) literature.

It has become widely discussed and rapidly growing discussions in many empirical works within this decade.

Purpose of study

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, to uncover RDM development and identify the potential field of research on RDM.

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)

The keywords co-occurrence network is a valuable tool for identifying research areas (Liu and Mei 2016).

UNIVERSITI MALAYA

Background of study

Several studies systematically review and bibliometrically analyse RDM literature, but they are limited to certain areas.

However, there is still limited profiling of RDM literature in terms of knowledge structure, to gauge the trends and future research focus.

1. 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.

2. Ashiq et al. (2020) **reviewed 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.

3. Zhang & Eichmann-Kalwara (2019) studied the RDM literature in the Scopus database, applying **bibliometric analysis** and **data visualisation** using **CiteSpace**.

8th International Conference on Libraries, Information & Society
18th – 19th August 2021
Kuala Lumpur Malaysia

UNIVERSITI MALAYA

Materials and method

1

The data for this study was generated from Elsevier's Scopus database as of April 2021

Scopus®

2

The keyword "research data management" contained in the title, abstract and keywords fields were used to search the relevant publications in the Scopus database and used the same keyword in additional query to retrieve specific publications focusing on RDM.

3

All the publications found were analysed using two application tools;

vosViewer

Excel

RIS

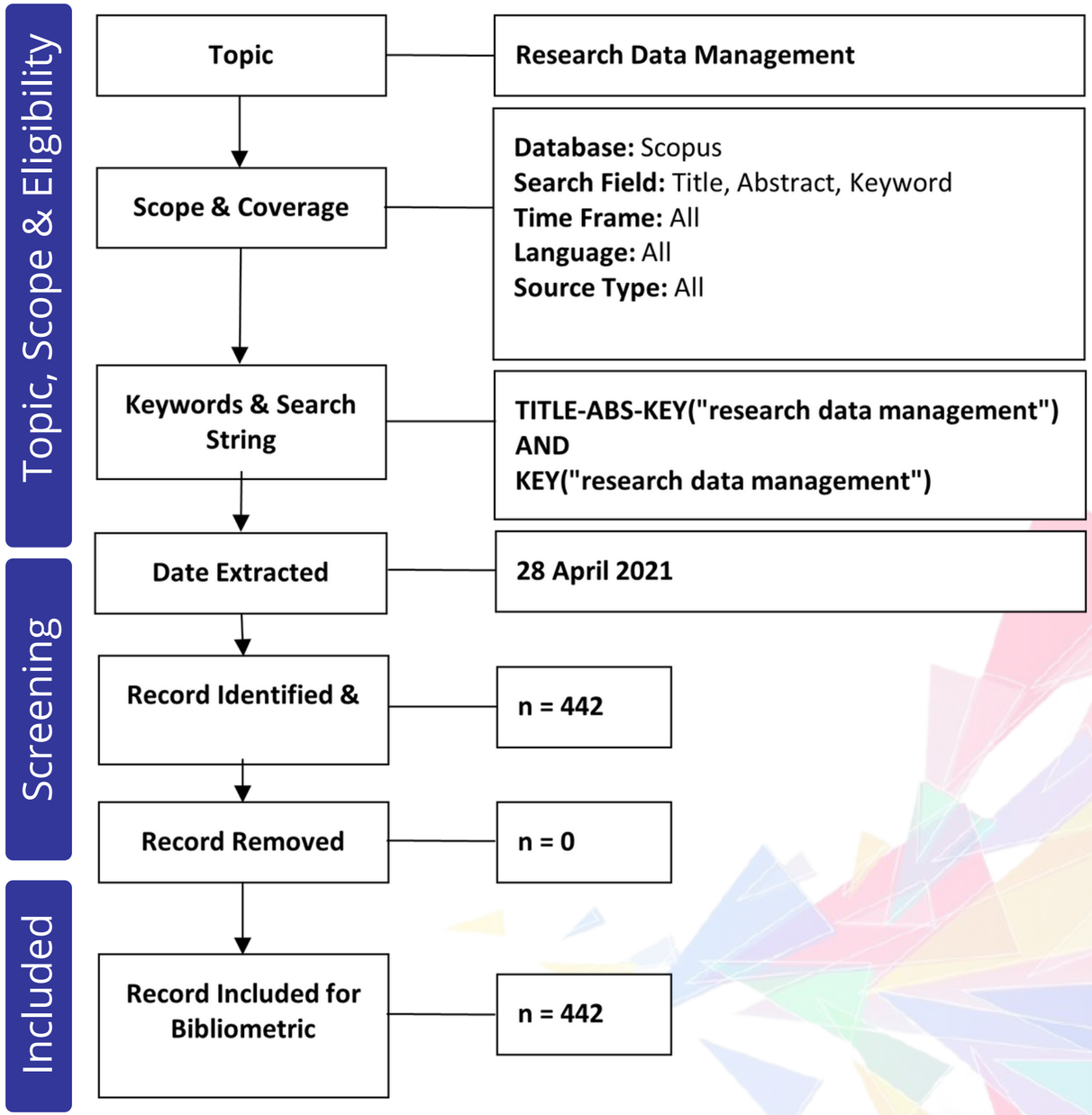
CSV

4

Social network analysis, a technique in identifying the clusters of related nodes within the network was used as the research approach (Benckendorff and Zehrer 2013).

5

This study developed the clusters from the **442 RDM publications** retrieved from this study.



Results

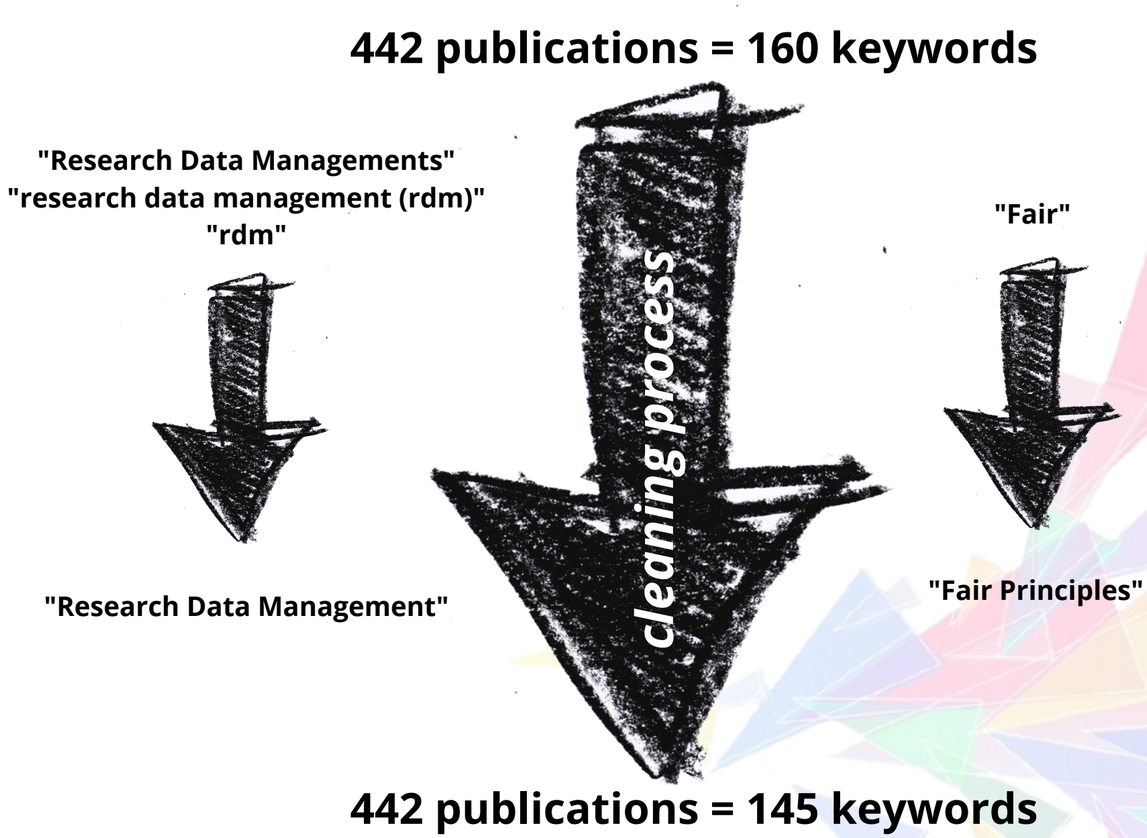
RDM Topics

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.*

This study combined some keywords, which become a unique keyword using the thesaurus method in VOSviewer to get accurate results.



Results

RDM Research Areas

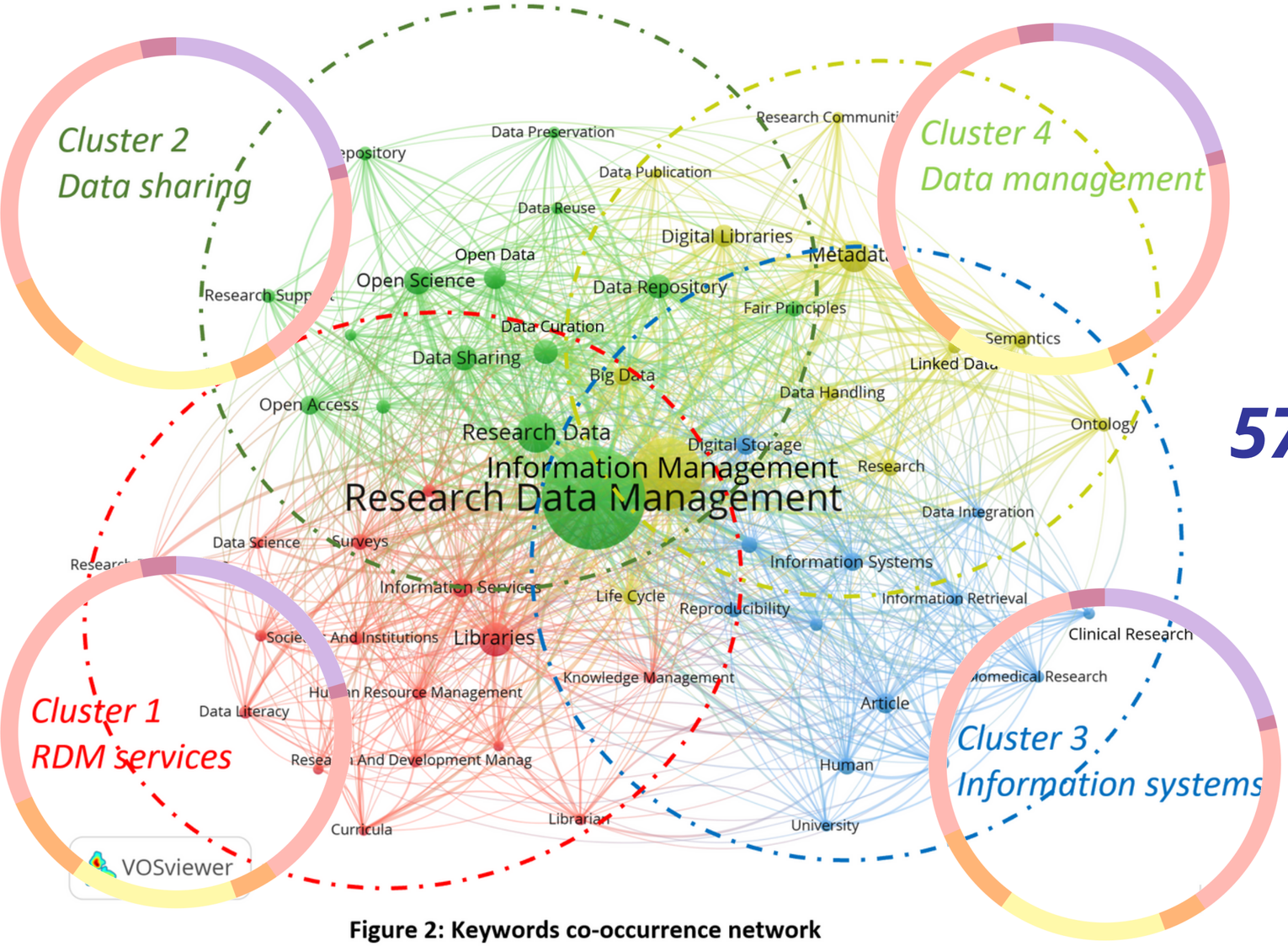


Figure 2: Keywords co-occurrence network

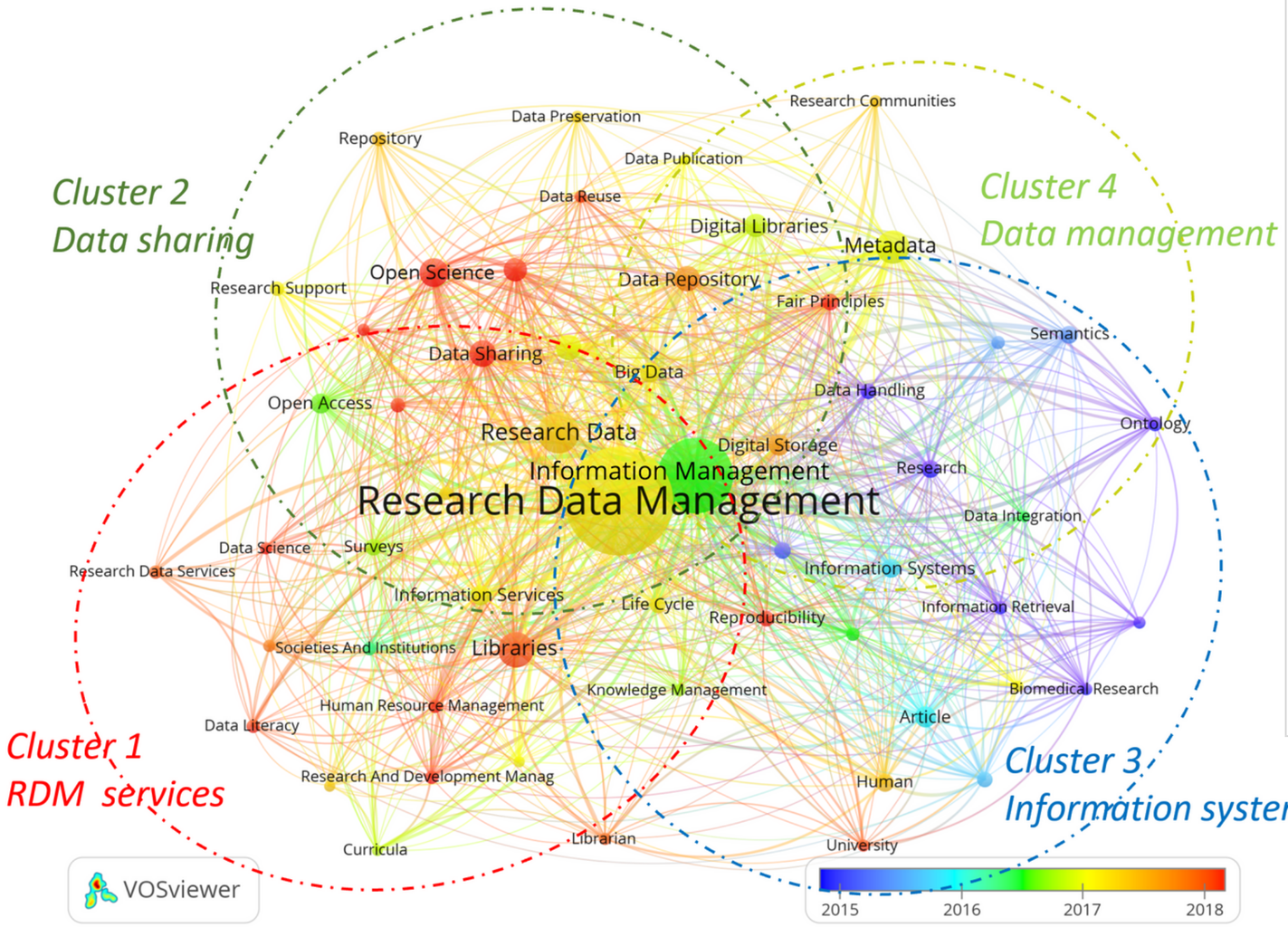
The cluster had been developed with the 10 minimum number of occurrences of keywords and excluded the keywords with low occurrences.

57 keywords

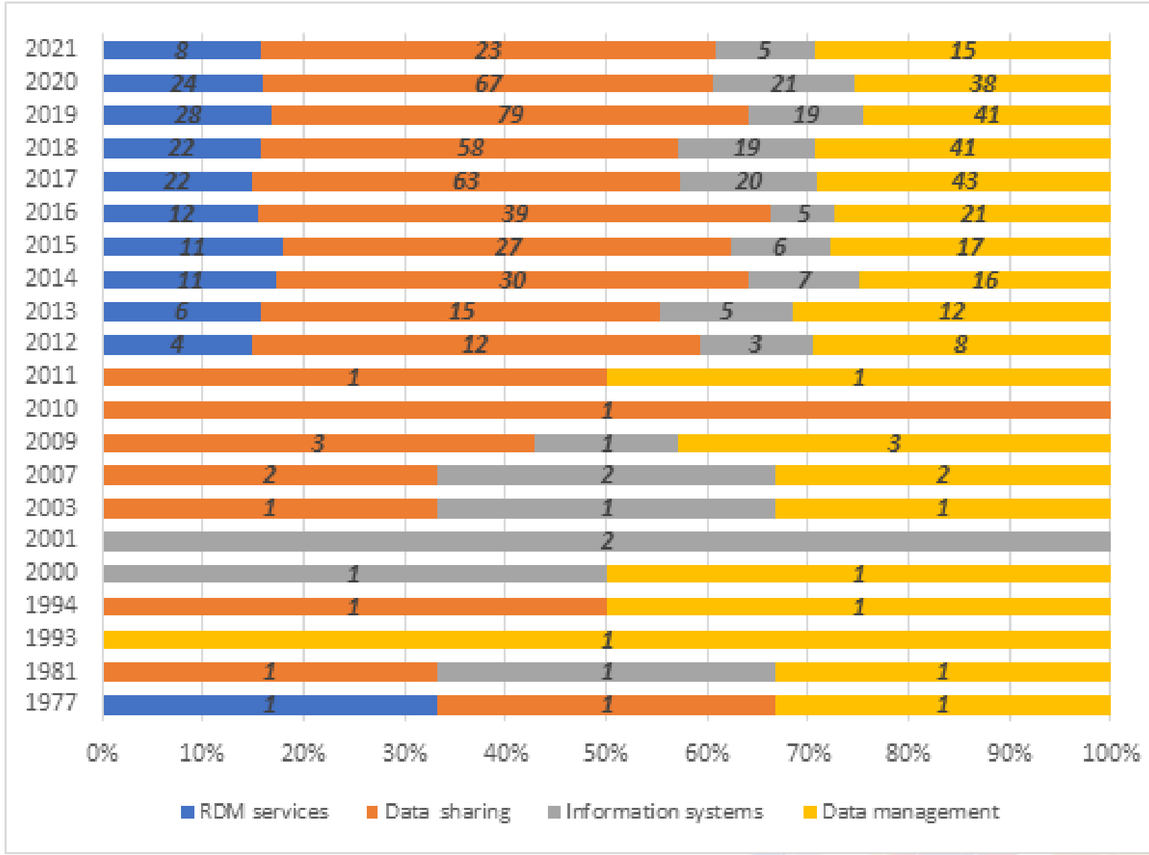
Figure 2 shows the network visualisation of the keywords co-occurrence network in which the circle size, font size, colour, and the thickness of the connecting lines indicate the link strength of the relationship between keywords.

Results

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).



Keywords co-occurrence network by year of publications



Number of RDM publications by years and clusters

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.

UNIVERSITI MALAYA

Results

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**

*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).

***Computer Science** and **Social Sciences** are the two highly represented subject areas in RDM publications.*

*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)**.*

***Some publications would have more than one subject area**

Discussion

8th International Conference on Libraries, Information & Society
18th – 19th August 2021
Kuala Lumpur Malaysia

UNIVERSITI MALAYA

Cluster 1: RDM services

The library has been accustomed to coordinate the **RDM services**, especially in advocacy and giving training on RDM to the researchers.
(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 3: Information Systems

The information systems in RDM involve data processing, integration and retrieval (Pinfield, Cox, and Smith 2014)

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).

Cluster 2: Data sharing

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).

The FAIR principles are also very important when the researchers are encouraged to have a greater engagement with RDM and openness.

Cluster 4: Data management

The keywords in this cluster were oriented to the data management activities, such as **metadata management.**

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)

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).

Literature growth

The analysis done was prominent when many publications on RDM have been started **from 2012 and increasing until today.**

This study plainly demonstrates that the publications about RDM are growing, especially in **Data sharing (Cluster 2).**

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)

This study also used **the scalable approach 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.**

UNIVERSITI MALAYA

Limitations & Further Study

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).

Further study could be **expanded to other databases.**

Further study could be expanded by **using specific keywords to focus on specific themes or fields** related to the RDM.

Conclusion

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.

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.

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.

UNIVERSITI MALAYA

Thank you for listening and watching our presentation.