

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

Winter 2-1-2021

Metadata Research Development: A Bibliometric Study on Science Direct

Seno Yudhanto

Universitas Indonesia, seno.yudhanto91@ui.ac.id

Taufik Asmiyanto

Universitas Indonesia, tasmiy@ui.ac.id

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>



Part of the [Library and Information Science Commons](#)

Yudhanto, Seno and Asmiyanto, Taufik, "Metadata Research Development: A Bibliometric Study on Science Direct" (2021). *Library Philosophy and Practice (e-journal)*. 5059.

<https://digitalcommons.unl.edu/libphilprac/5059>

Metadata Research Development: A Bibliometric Study on Science Direct

Seno Yudhanto^{1&2}, & Taufik Asmiyanto^{1*}

¹Departement of Library and Information Science, Faculty of Humanities, Universitas Indonesia, Depok, 16424, Indonesia

²Center for Scientific Data and Documentation, Indonesian Institute for Sciences, Jakarta, 12710, Indonesia

seno.yudhanto91@ui.ac.id
tasmiy@ui.ac.id *corresponding author

Abstract

This study aims to identify (1) the number of publications by year; (2) number of publications by type of document; (3) the top ten publications that publish metadata; (4) the distribution of metadata research topics based on subjects and keywords; and (5) mapping research trends in the discussion of metadata. The method uses a bibliometric study sourced from the Science Direct database (www.sciencedirect.com). Data is collected through published searches with the keyword "metadata" in the title, abstract or author-specified keywords area and limited to the last 10 years (2011-2020) and it is managed and analyzed using Ms. Excel. Meanwhile, metadata international publication developments trends were analyzed using the VosViewer application. The result shows that the most publications were in 2016 with 33 publications from 221 publications obtained. The most publications based on document type are journal articles with 163 documents, while the most publication with titel of metadata research are Procedia Computer Science with 22 publication titles. In the distribution analysis, research topics result in 10 research subjects and 40 keywords related to 8 clusters, with the most keywords which are "metadata", "linked data", and "data management".

Keywords: *metadata, bibliometrics, VOSViewer, science direct, information organization*

INTRODUCTION

The progress of information and communication technology (ICT) has an impact to the acceleration of the science development and at the same time it causes information growing exponentially. This phenomenon is often referred to as the information revolution. Information then becomes a reality that affects all human life. Information is understood not only to relate to ‘*about reality*’ (information *about* reality) but also interpreted as ‘*as reality*’ (information *as* reality) and ‘*for reality*’ (information *for* reality). (Asmiyanto, 2020) The existence of this information in turn changes the way and / or methods of collecting, selecting, managing, and interpreting and passing it back on. This information cycle creates the needs and demands of related fields (library science, information science, archival science and others) to develop an information management mechanism that is capable of dealing with the revolution that occurred.

The diversity of information that is attained, both in terms of format, type and medium, creates different perspectives in managing information resources (Joudrey *et al.*, 2018). One of the focuses in organizing information resources is metadata. Metadata can include bibliographic information such as those in traditional library catalogs, subject catalogs, such as descriptors, classification designations, abstracts, etc. (Nair & Jeeven, 2004). The greater the number of resources, the adequate systems and infrastructure are needed in their management.

There are quite a few studies on metadata, but not yet mapped. Kuźma and Mościcka (2020) examined the evaluation criteria for cartographic collection metadata through the systematic literature review method. The results of his research indicated that there are no universal criteria for evaluating metadata. Gregg *et al.* (2019) identified the challenges, opportunities, and knowledge gaps related to the use of metadata in scientific communication. They compile and interpret the literature in sections based on groups of professionals in scientific communication metadata: researchers, funders, publishers, librarians, service providers, and data curators. Meanwhile, Alemu *et al.* (2017) examined the need to re-conceptualize existing metadata principles and technical formats with a Linked Open Data framework based on papers from the 11th International Conference on Metadata and Semantics Research (MTSR-2017). The results of the study showed that the future of metadata, ontology, and semantics can be enriched, linked, exposed, and filtered. In addition, ontologies need to reflect the diversity of interpretations inherent in humans and the existence of multilingual, cross-cultural, and multidisciplinary content, so it must be designed, developed and maintained by considering diversity, scalability and interoperability. Thus, it is very important to examine how the development of studies on metadata so far. One way to evaluate this is through a bibliometric study (Tupan *et al.*, 2018).

Based on these problems, the study of metadata publication analysis was conducted In Science Direct indexed publications. This study aims to identify (1) the number of publications by year; (2) number of publications by type of document; (3) the top ten publications that publish metadata; (4) the distribution of metadata research topics based on subjects and keywords; and (5) mapping research trends in the discussion of metadata.

LITERATURE REVIEW

Metadata

The definition of metadata is data about data. Metadata can be divided into several types (Riley, 2017), namely descriptive, administrative, structural, and markup languages. Administrative metadata is information about the content of a resource that helps

locate or understand it. Administrative metadata refers to information that is required to manage a resource or is associated with its creation. Structural metadata describes the relationship of resource parts to one another. In entirety, metadata markup languages combine all metadata and content into one.

Metadata is descriptive and classification information about digital resources such as web pages, computer files, images, multimedia documents, etc (Ramesh *et al.*, 2015). Furthermore, he argued that metadata is important for conserving and sharing these resources. Currently the role of metadata is expanding due to the development of digital resources. The complexity, quantity and variety of information that is required to manage digital resources are often greater than conventional or printed resources (Joudrey *et al.*, 2018). Thereupon, the ultimate goal of metadata is to find, identify, select, obtain, and explore the information.

Bibliometric Studies

Bibliometric is a scientific study of discourse, which is the application of mathematics and statistical methods for books and other communication media through the information transfer process, with the aim of analysis and process control (Rousseau *et al.*, 2018). Bibliometric studies have links to other “metrics” studies, such as scientometrics, informetrics, or webometrics. (Bellis, 2009) stated that the purpose of this research area is to analyze and measure communication phenomenon to build an accurate formal representation of behavior to make it clearer, evaluative, and administrative.

The difference between these studies lies in the order of the factors and the limitations of the objects being measured. One of the areas of bibliometric study is co-word analysis (Chellappandi & Vijayakumar, 2018). In addition, (Tupan *et al.*, 2018) stated that co-word analysis can be mapped based on co-occurrences, important or unique terms contained in the article and can be seen from the title or abstract. Through this analysis, it can be determined the distribution of research or study topics that develop from time to time.

RESEARCH METHODS

This research is a bibliometric study using international publication data on metadata sourced from the Science Direct database (www.sciencedirect.com). Data is collected through searches published in Science Direct in January 2020 with the keyword "metadata" in the title, abstract or author-specified keywords. This study limits the search time in the period 2011 - 2020 (the last 10 years). Data is conducted in the form of the number of publications per year containing articles on metadata. Then the bibliography was extracted in csv format. Data is

managed and analyzed using Ms. Excel, while for the development of international publication trends, the field of metadata is mapped using the VosViewer application.

DISCUSSION

Publications per year

The number of publications on metadata, indexed in Science Direct from 2011 to 2020, are 221 titles. Table 1 and Figure 1 show the development of research studies on metadata by year. The highest number of publications regarding metadata occurred in 2016 with 33 publications (14.93%). In second place, the number of research publications on metadata shows that in 2013 there are 32 publications (14.48%), followed by 2015 with 30 publications (13.57%), in 2020 there are 26 publications (11.76%), 2018 and 2019 with 20 publications (9.05%), in 2012 and 2017 about 16 publications (7.24%), and in 2014 about 15 publications (6.79%). The least number of publications occurred in 2011 with 13 publications (5.88%).

Table 1. Publications about metadata per year

Year	Number of Publications	Percentage
2011	13	5,88%
2012	16	7,24%
2013	32	14,48%
2014	15	6,79%
2015	30	13,57%
2016	33	14,93%
2017	16	7,24%
2018	20	9,05%
2019	20	9,05%
2020	26	11,76%
Total	221	100%

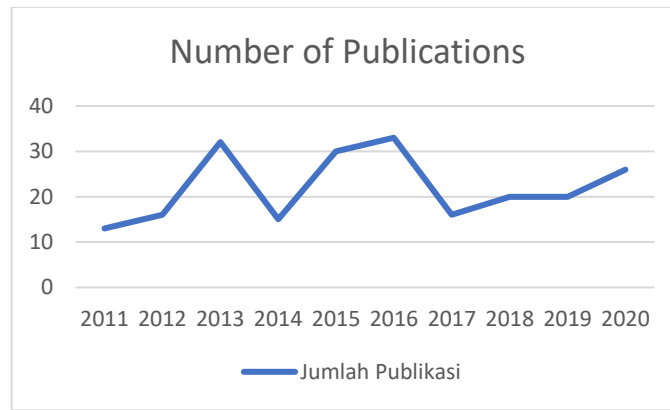


Figure 1. Publication Trends on Metadata per year

In addition, based on search results through Science Direct during the 2011-2020 period, Table 2 represents the most metadata research trends are in the form of journal articles about 163 documents (73.76%), then followed by Book Chapters with 44 documents (19,91%), conference abstracts about 7 documents (3.17%), 4 documents for book review (1.81%) and review articles with 3 documents (1.36%).

Table 2. Number of publication types about metadata

Type of Document	Number of Publications	Percentage
journal article	163	73,76%
book chapter	44	19,91%
conference abstract	7	3,17%
book review	4	1,81%
review article	3	1,36%
Total	221	100%

Top 10 Publication Titles

Based on search results in Science Direct, there are 221 publications with the keyword "metadata". From this number, it is known that the largest number of international publications are published in Procedia Computer Science (22 publications). The top ten publications that publish developments regarding metadata can be seen in Table 3.

Table 3. Publication Titles about metadata

Publication Title	Number of Publications
Procedia Computer Science	22
An Emergent Theory of Digital Library Metadata	8
Managing eBook Metadata in Academic Libraries	8
Fusion Engineering and Design	7
Digital Investigation	7
Future Generation Computer Systems	6
The Journal of Academic Librarianship	6
Journal of Biomedical Informatics	5
Environmental Modelling & Software	4
Ecological Informatics	4

Number of research topics

Based on search results on the Science Direct database, it shows that the subject of computer science is the highest subject. Then, it was followed by social sciences, engineering, medicine and dentistry, business, management, and accounting, decision sciences, environmental sciences, agricultural and biological sciences, energy, and arts and humanities.

Table 4. The Subject of research publication metadata

Subject areas	Klas DDC	Number of Publications
Computer Science	004	115
Social Sciences	300	50
Engineering	620	37
Medicine and Dentistry	611	21
Business, Management and Accounting	658	17
Decision Sciences	658	16
Environmental Science	363	15
Agricultural and Biological Sciences	630	10
Energy	621	10
Arts and Humanities	700	7

The specified term to be analysed in VOSviewer is minimum of 3 keywords, then it is found 40 topics with a total of 864 keywords. The results are shown by extracting keywords from the publication bibliography. Based on the search results using keywords shows that the trend of metadata research in the period 2011-2020 which is based on the Science Direct

database, 10 main keywords are the keyword "metadata" which is the largest number with 66 times appearing in publications, followed by the keyword "linked data" is 9 times , " data management " is 8 times , " digital forensics ", " interoperability ", " ontology " and " rdf " are all 5 times, " digital libraries ", "Dublin core" and "metadata management" are 4 times. All metadata research trends can be seen in table 5 below.

Table 5. Metadata research trends based on Keywords

Keyword	Occurrences	Total Link Strength
metadata	66	55
linked data	9	25
data management	8	7
digital forensics	5	5
interoperability	5	7
ontology	5	6
rdf	5	15
digital libraries	4	4
dublin core	4	14
metadata management	4	2
preservation	4	6
provenance	4	4
semantic web	4	10
semantics	4	9
standards	4	7
xml	4	7
bibframe	3	10
big data	3	7
cerif	3	2
cloud storage	3	4
controlled vocabularies	3	6
crowdsourcing	3	5
data analysis	3	3
data integration	3	3
data mining	3	5
data model	3	4
deep learning	3	4
dicom metadata	3	1
document metadata	3	3
ead	3	9
folksonomy	3	3

information management	3	4
information retrieval	3	1
machine learning	3	6
marc	3	8
metadata granularity	3	4
metadata mapping	3	6
mods	3	13
open data	3	2
rda	3	10

The analysis of mapping and publication trends about metadata using network and density visualization is utilizing VOSViewer version 1.6.14. It aims to find out the network map that exists between publications from the metadata that has been downloaded. The network map based on the network visualization is divided into 8 clusters as in Figure 2. The clusters are divided as follows:

Cluster 1 (red): bibframe, controlled vocabularies, Dublin core, ead, marc, metadata mapping, mods, rda, dan rdf

Cluster 2 (green): big data, data mining, dicom metadata, digital libraries, document metadata, dan standards

Cluster 3 (blue): crowdsourcing, folksonomy, metadata, metadata management, semantics, dan xml

Cluster 4 (yellow): certif, data integration, interoperability, linked data, ontology, open data

Cluster 5 (purple): data analysis, data management, information management, provenance, semantic web

Cluster 6 (light blue): cloud storage, data model, digital forensics, preservation

Cluster 7 (orange): deep learning, information retrieval, machine learning

Cluster 8 (brown): metadata granularity

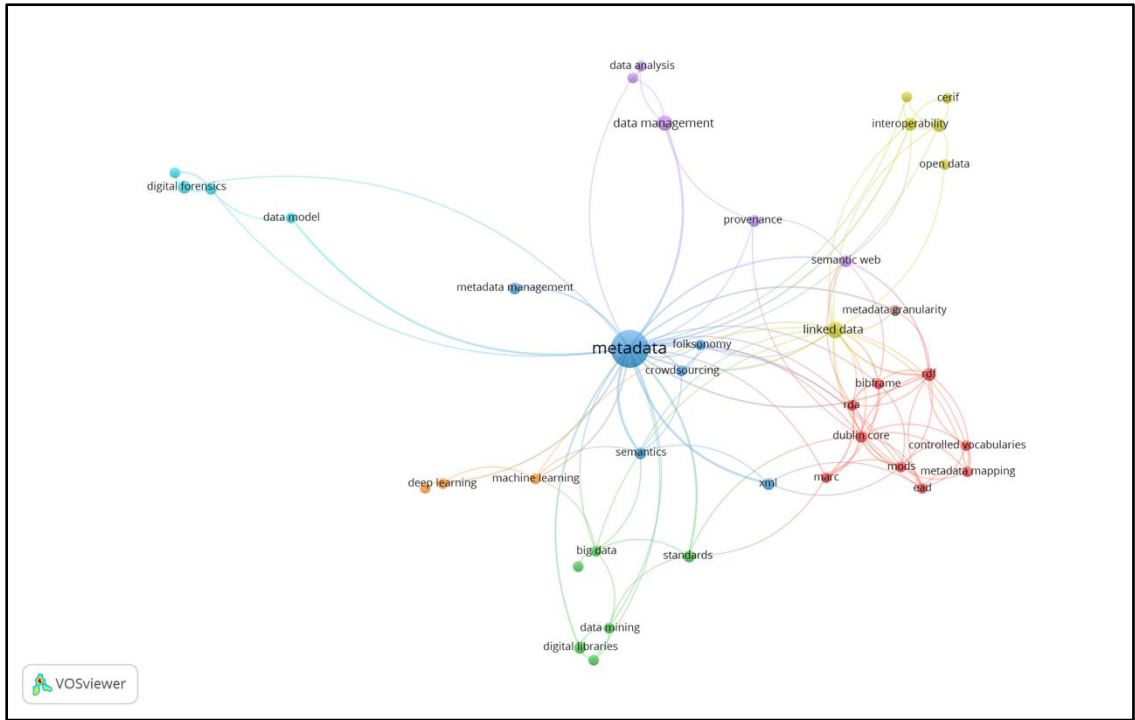


Figure 2. Map of metadata research development based on network visualization

Figure 3 below shows a map of the analysis results using all articles on metadata research developments. The visualization results show that the darker the color, the more research is carried out. Conversely, if the color is faded, the less research is done. For example, there are quite a lot of metadata studies that discuss folksonomy, crowdsourcing, data management, or linked data, yet research that discusses data models, provenance, or open data is still rare.

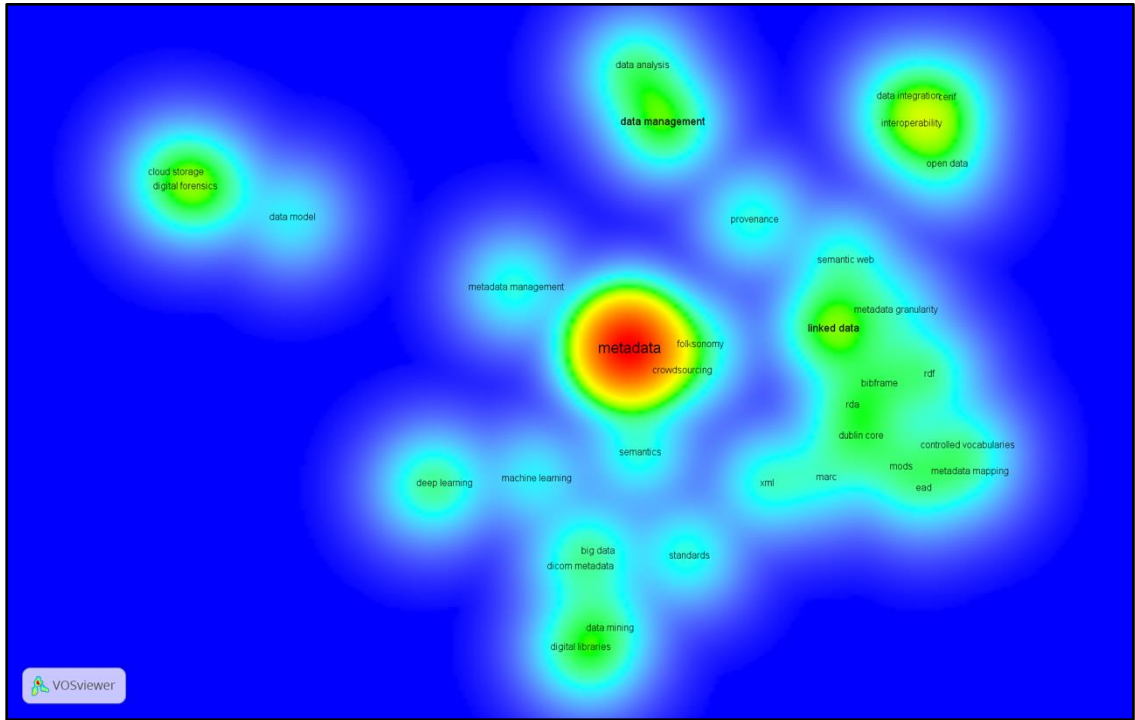


Figure 3. Map of metadata research development based on density visualization

CONCLUSION

This study can assist researchers in the study of information organization, especially metadata in determining the study boundaries. Based on the analysis of the bibliometric study, it can be concluded that the research development on metadata in the Science Direct database in the 2011-2016 period was the highest point in 2016 with 33 publications from 221 publications obtained. The most publications based on document type are journal articles with 163 documents, while the most publications that produce metadata research are Procedia Computer Science with 22 publication titles. In the analysis of the number of research topics resulted in 10 research subjects and 40 related keywords; The result of keyword mapping consist of 8 clusters with the most keywords are "metadata", "linked data", and "data management".

ACKNOWLEDGMENT

This research is supported by Indonesian Institute of Sciences (LIPI) and SAINTEK Scholarship, Ministry of Research and Technology/National Research and Innovation Agency of Republic Indonesia (KEMENRISTEK / BRIN)

REFERENCES

- Alemu, G., Garoufallou, E., Siatri, R., Koutsomiha, D., & Virkus, S. (2017). Metadata and semantics research: A case of an international conference paving toward a data-driven future. *International Journal of Metadata, Semantics and Ontologies*, 12(4), 167–181. <https://doi.org/10.1504/IJMSO.2017.093640>
- Asmiyanto, Taufik. (2020). Conceptual Foundations of Library and Information Sciences: Social Epistemology or Philosophy of Information. *International Review of Humanities Studies*, 5(1), 108-121.
- Bellis, N. De. (2009). *Bibliometrics and Citation Analysis: from the Science Citation Index to Cybermetrics*. Scarecrow Press, Inc.
- Chellappandi, P., & Vijayakumar, C. S. (2018). Bibliometrics, Scientometrics, Webometrics/ Cybermetrics, Informetrics and Altmetrics - An Emerging Field in Library and Information Science Research. *Shanlax International Journal of Education*, 7(1), 5–8. <https://doi.org/http/doi.org/10.5281/zenodo.2529398>

- Gregg, W., Erdmann, C., Paglione, L., Schneider, J., & Dean, C. (2019). A literature review of scholarly communications metadata. *Research Ideas and Outcomes*, 5. <https://doi.org/10.3897/rio.5.e38698>
- Joudrey, D. N., Taylor, A. G., & Wisser, K. M. (2018). *The Organization of Information* (Fourth Ed.). Libraries Unlimited.
- Kuźma, M., & Mościcka, A. (2020). Metadata evaluation criteria in respect to archival maps description: A systematic literature review. *Electronic Library*, 38(1), 1–27. <https://doi.org/10.1108/EL-07-2019-0161>
- Nair, S. S., & Jeeven, V. K. J. (2004). A Brief Overview of Metadata Formats. *DESIDOC Bulletin of Information Technology*, 24(4), 3–11. <https://doi.org/10.14429/dbit.24.4.3629>
- Ramesh, P., Vivekavardhan, J., & Bharathi, K. (2015). Metadata Diversity, Interoperability and Resource Discovery Issues and Challenges. *DESIDOC Journal of Library and Information Technology*, 35(3), 193–199. <https://doi.org/10.14429/djlit.35.3.8074>
- Riley, J. (NISO). (2017). Understanding Metadata - What is Metadata, and What is it for? In *NISO Primer Series*.
- Rousseau, R., Egghe, L., & Guns, R. (2018). *Becoming Metric-Wise: A Bibliometric Guide for Researchers*. Chandos Publishing.
- Tupan, T., Rahayu, R. N., Rachmawati, R., & Rahayu, E. S. R. (2018). Analisis Bibliometrik Perkembangan Penelitian Bidang Ilmu Instrumentasi. *BACA: Jurnal Dokumentasi Dan Informasi*, 39(2), 135–149. <https://doi.org/10.14203/j.baca.v39i2.413>