

# ADVANCED TECHNOLOGY FOR DIGITAL ARCHIVE RESEARCH

**Dr. Wirapong Chansanam**

Department of Information Science, Faculty of Humanities and Social Sciences,  
KhonKaen University, KhonKaen, Thailand

## INTRODUCTION

Digital archival (DA) descriptions have to reflect the peculiarities of the archive, retain all the informative power of a record, and keep track of the provenance and original order. This is emphasized by the central concept of fonds, which reflects an organic process in which a record creator produces or accumulates a series of records (Pearce-Moses, 2005; Cook, 1993; Duranti, 1998; Gilliland-Swetland, 2000). Duranti (1998) argues that maintaining provenance leads archivists to evaluate records based on the importance of their creator's mandate and functions and fosters the use of a hierarchical method. The hierarchical structure of an archive expresses the relationships and dependency links between the records of the DA. This article employs the bibliometrics technique to investigate the current state and perception of digital archives through systematic analysis and then uses the findings to suggest directions for its future development. This article examines the development and historical movement (past and current) and contemporary (present). Furthermore, it explores current trends and developments advance technology in digital archives study (the future). The research answers two critical questions:

RQ1: The complete information and publishing trend in digital archives is becoming more popular, includes top-cited articles, top-ranked authors, publishing institutes and countries, and keywords.

RQ2: The conceptual structure of extent and influence of the digital archives.

The quantitative features of citation information and Scopus publishing structures will be investigated to answer the research issues raised above. There are many characteristics, including a keyword co-occurrence network, a temporal overlay keyword visualization, and co-citation and bibliography information. Furthermore, this study shows how authors, organizations, and nations worked together to structure the knowledge domains of journals in the context of knowledge domain structuring. By identifying existing and emerging research topics that will aid in developing intelligent future studies, this study assesses its contribution to the digital archives field. The following section provides a breakdown of the paper's organizational structure. The next part covers the research methodologies, followed by the findings and comments; the last section discusses the study's conclusion and its contribution and flaws.

## METHOD

This article quantitative analyzes the digital archives using worldwide publishing statistics from the Scopus database. This research was carried out on September 17th, 2022, and included the terms "digital archives" in the article titles and abstracts, as shown below.:

```
TITLE-ABS-KEY ( "digital archives" ) OR ( "digital archive" ) AND ( LIMIT-TO ( SRCTYPE , "j" ) )
AND ( LIMIT-TO ( DOCTYPE , "ar" ) ) AND ( EXCLUDE ( PUBYEAR , 2022 ) ) AND ( LIMIT-TO
(LANGUAGE , "English" ) )
```

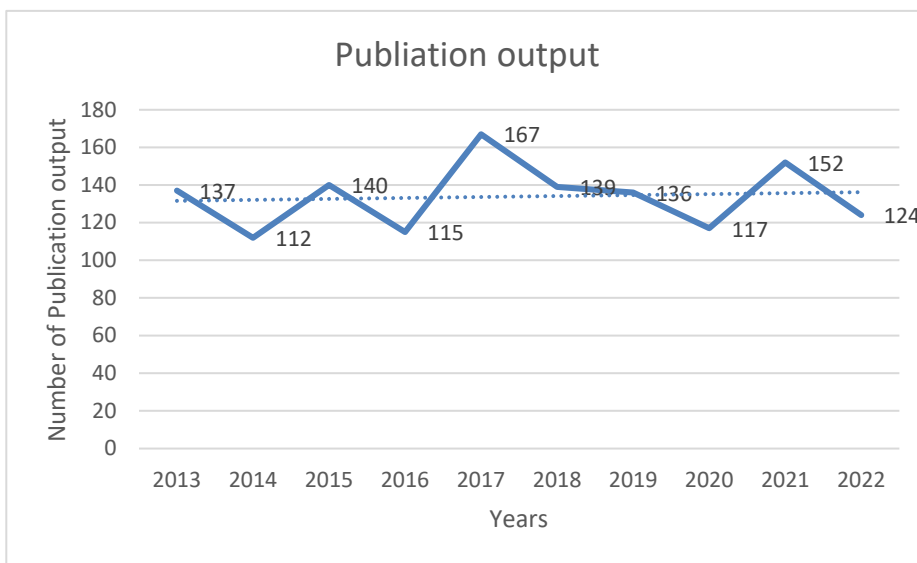
This study examines a wide range of research trends, authors, universities, keywords, publications, citation structures, and the development of research trends through time. The bibliometric approach was used to examine the visual representation of the growth of digital archives research. Additionally, data from bibliometric analyses were examined with the use of the Bibliometrix R programme (Aria & Cuccurullo, 2017).

**RESULTS AND DISCUSSION**

Since its first article in 1981, Digital archives has been referenced in 2,707 publications, with an average of 0.185 citations for each item published. The findings of the digital archives Scopus bibliometric study are described in this part of the paper. Furthermore, the findings of this study are presented in the form of research questions.

**Publication output**

Figure 1 shows a substantial increase in the number of research papers published over the past ten years, showing the high level of interest that has grown in the academic community over this period. Compared to 2013, the pace of growth has risen from 137 documents per year to 124 documents per year now. (recorded in 2022), and the highest recorded in 2017 with 167 documents. Because just one paper was published in 1981, we can conclude that digital archives are a relatively new field based on the analysis of publications. The Scopus database has a total of 1,339 papers that have been collected for ten years. This trend is anticipated to become more pronounced in 2022 and the following years. (during this analyzing not the end of the year yet).



**Fig.1**Digital archives publication output.

**Bibliometric analysis**

**Digital archives articles developments in publishing and general description**

To answer the first question, you must first grasp the structure of digital archives publication descriptions, trends, and citations.

**Table 1**Average Citations per Year.

Year	N	MeanTCperArt	MeanTCperArt	CitableYears
2013	137	5.708029197	0.634225466	9
2014	112	8.910714286	1.113839286	8
2015	140	5.335714286	0.762244898	7
2016	115	6.086956522	1.014492754	6
2017	167	8.143712575	1.628742515	5

*Contd...*

Year	N	MeanTCperArt	MeanTCperArt	CitableYears
2018	139	5.366906475	1.341726619	4
2019	136	3.286764706	1.095588235	3
2020	117	1.837606838	0.918803419	2
2021	152	0.927631579	0.927631579	1
2022	124	0.637096774	-	0
<b>Total</b>	<b>1,339</b>	<b>4.62</b>	<b>1.05</b>	

Table 1 shows the average citations per year. It was reported in 2014 that the most significant total number of citations per article was 8.91, with an annual average of 1.11, indicating that the publications in 2014 were the most referenced in the field of digital archives. Articles with the second most significant number of citations received 8.14 in 2017, with an average of 1.62 each year.

### Three Fields Plot

The three-field plot depicts three components, including a list of journal names, authors, and subjects and a list of journal names, authors, and topics. The names of the journals are given first, followed by the names of the authors, and then each author is connected to the topic of the publication to which they contributed. Furthermore, grey connections are used to tie these three components together to demonstrate their interdependence. The number of articles on each element is represented by the rectangles on the lists, with journals being the first element on the left. Scopus has indexed ten journals, with the most articles appearing in the book series “Communications in Computer and Information Science.” It is represented as a purple rectangle that is connected to several other journals. The names of the authors are included inside the centre section. The main components are linked with the authors of papers that have been published in well-known publications. This graphic contains a list of the top ten authors, with the size of each author's rectangle reflecting the number of articles the author has authored. The Miller rectangle is the most significant shape in this graphic. This section contains the most frequently used terms connected to the subject. Each topic is linked with one or more authors who have published the most articles on the subject. There are six keyword topics, including digital archives, archives, digital archive, digital humanities, archive, and metadata, as shown in Figure 2.

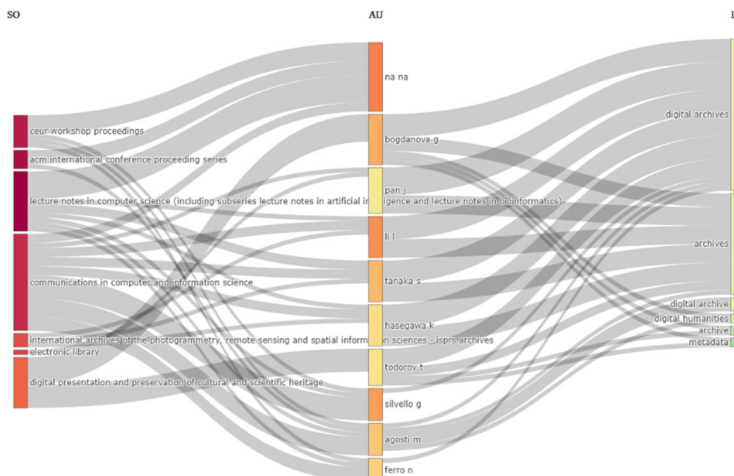
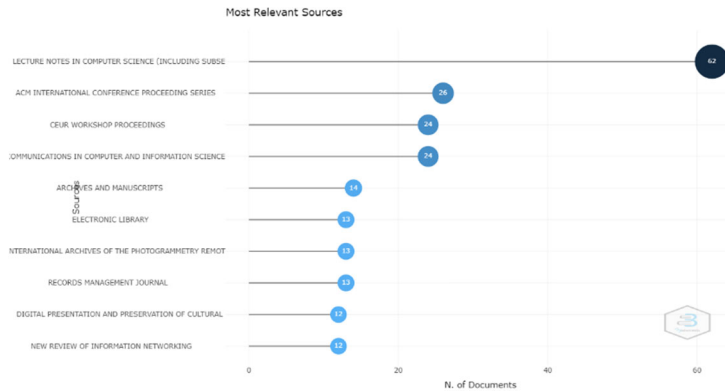


Fig. 2 Three Fields Plot.

**Most relevant sources**

The articles published by each publication are shown in the Figure 3 chart, which ranks them according to their importance to the digital archives study. Furthermore, the data displays the top journals and papers that have been published, with a ranking ranging from 0 to 62.

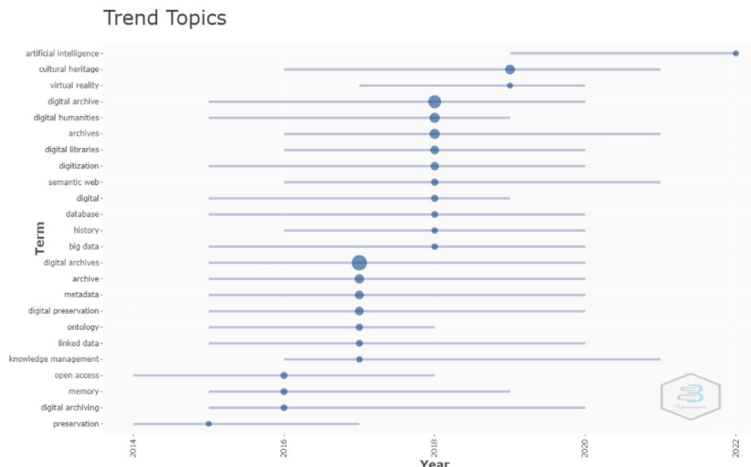


**Fig. 3** Most Relevant Sources.

**Trend Topics**

According to the term frequency found in this digital archives study, topics arise as a result. As shown in the accompanying chronology, these are the most frequently mentioned subjects. Keywords often used appear higher on the list, and their occurrence points are represented in the timeline.

Figure 4 depicts the topic's development since 2014 when it saw an accelerated rise. The timeline reveals that epistemology has been the most frequently discussed subject since 2014. Between 2015 and 2016, open access, memory, digital archiving, and preservation. digital archives, archive, metadata, digital preservation, ontology, linked data, ontology, knowledge management have interest in 2017. digital archive, digital humanities, archive, digital libraries, digitization, semantic web, digital, database, history, and big data began to explore digital archives topics in 2018. Between 2019 and 2020, cultural heritage and virtual reality were popular in digital archives research, especially in the digital archives topic numerous popular in 2019. From 2021, artificial intelligence became increasingly popular.



**Fig. 4** Topic historiography timeline.

## Thematic Map

This research examined a thematic map by separating it into four topic quadrants based on density and centrality, as shown in the picture to the right. Due to their great density and centrality, the topics in the upper right quadrant should be explored and researched further. In contrast, a unique, uncommon, but highly developing theme with high density and low centrality is in the top left quadrant. Themes with a downward tendency are located in the lower-left quadrant, while basic themes with a high centrality but a low density are located in the lower-right quadrant. In Figure 5, the timeline demonstrates that the subjects of archive, digital, memory, big data, and education in the upper-right quadrant of the timeline are the most promising for future study.

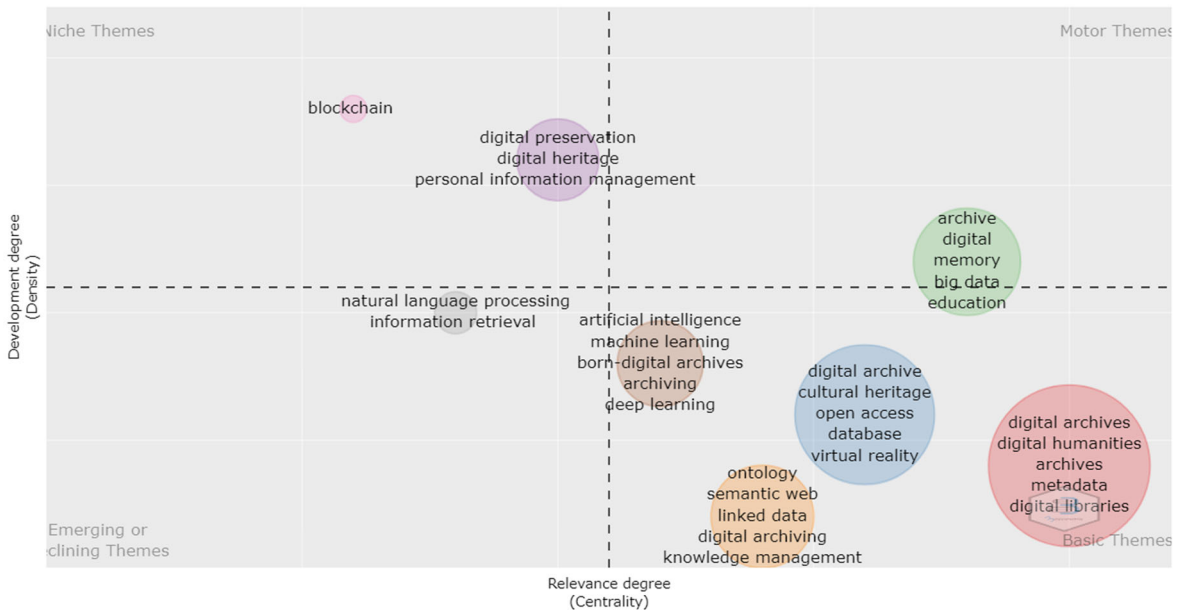


Fig. 5 Thematic Map.

## Structure of Digital Archives (DA) research

The research indicates that the Author and journal co-citation analysis and coupling bibliometric and network techniques are used in this study (Aria & Cuccurullo, 2017).

### Co-citation analysis of sources and documents

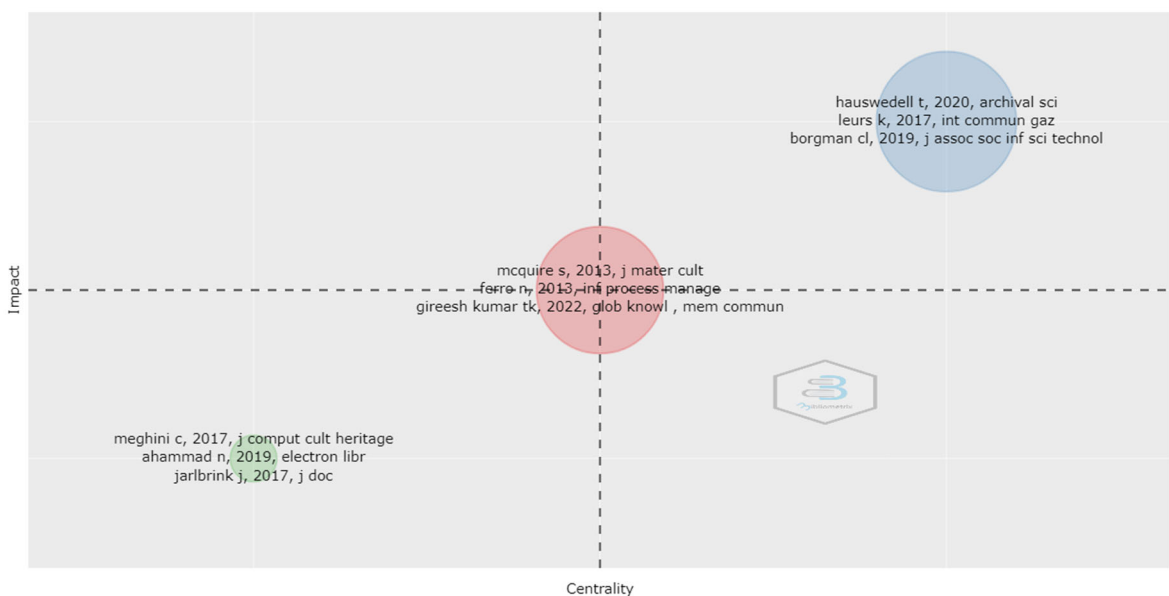
This research advanced a network of co-cited publications, with 49 of the 2,707 sources fulfilling the first criterion of a minimum of 20 citations. Four clusters emerged from a co-citation study of these 49 sources, including the top-10 sources with the most vital links. The first cluster (12 sources) is the largest and includes archival science (Betweenness = 305.9995701, Closeness = 0.012658228), archivaria, the american archivist, and etc. The second (18 sources) is the next largest cluster, consisting of studies from science (Betweenness = 42.46348263, Closeness = 0.010309278), nature, d-lib magazine, and etc. The third cluster (8 sources) includes new media & society (Betweenness = 45.63415056, Closeness = 0.010752688), journal of consumer research, theory, and etc. The fourth cluster (7 sources) consists of digital humanities quarterly (Betweenness = 12.25479859, Closeness = 0.009259259), debates in the digital humanities, and etc., as shown in Figure 6.



and exclusion of digitised newspapers in and from online archives. We draw attention to providers' emphasis on meeting the needs of their end-users and how this is shaping the form and function of digital archives. At the same time, Leurs (Leurs, 2017) study re-conceptualises and empirically grounds communication rights. The focus is on the usage of social media among young refugees, who operate from the margins of society. He focuses on digital performativity as a means to address unjust communicative power relations and human right violations. Borgman and colleagues (Borgman et al., 2019) defined archives are preferred means for open access to research data. Little is known about how they mediate information exchange between stakeholders. Archivists devote about half their time to aiding contributors with curation processes and half to assisting consumers. Human assistance in curation and search remains essential.

Scott McQuire (McQuire, 2013), author argues that the current shift of the photograph from 'picture' to 'data' is driving a related shift in the image archive. Operability has implications for developing new protocols for managing image archives relevant to historically oppressed groups such as Aboriginal peoples in Australia represents cluster two (40 authors). Meghini and his associates (Meghini et al., 2017), which conducted a research infrastructure for archaeology (ARIADNE), that is the Advanced Research Infrastructure for Archaeological Dataset Networking in Europe. This EU-funded network has developed an e-infrastructure that enables data providers to register and provide access to their resources (datasets, collections). The portal also enables discovery, access, and other services across the integrated resources cluster three (23 authors).

#### Clusters by Documents Coupling

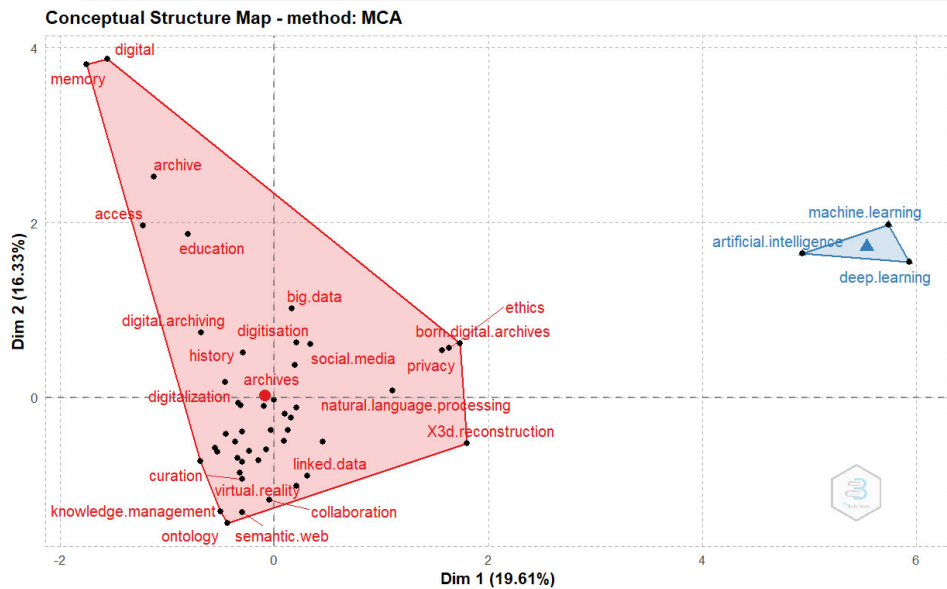


**Fig. 8** Bibliographic Document Coupling.

#### Digital archives publications concept structure

The study conducted the Multiple Correspondence Analysis (MCA) of the keywords included in my dataset. The conceptual structure of the keywords associated with the resilience articles included in this study was presented in Figure 9. It compresses extensive data with multiple variables into a low-dimensional space to form an intuitive two-dimensional (or three-dimensional) graph that uses plane distance to reflect the

similarity between the keywords. Keywords approaching the centre point indicate that they have received close attention in recent years (Xie et al., 2020).



**Fig. 9** Digital archives publications conceptual structure in map view.

The results are interpreted based on the relative positions of the points and their distribution along the dimensions; as words are more similar in distribution, the closer they are represented in the map (Aria & Cuccurullo, 2017). Cluster 1 (red colour) of 24 keywords comprises papers regarding digital, memory, archive, access, education, big data, digital archiving, and etc. Cluster 2 (blue colour) is the most significant consists of three keywords that focus on the documents related to artificial intelligence, deep learning, and machine learning.

## CONCLUSION

A quantitative and systematic literature review is used in this research to assess the digital archives' reach and effect. It's a great way to reflect on the past while also contributing to the advancement of future knowledge. The study found broad trends in digital archives research based on the two research topics, such as the most prolific authors, institutions, nations, important literature, and citation structure. The series Lecture Notes in Computer Science (LNCS), including its subseries Lecture Notes in Artificial Intelligence (LNAI) and Lecture Notes in Bioinformatics (LNBI), the ACM International Conference Proceeding Series (ICPS), and CEUR Workshop Proceedings (CEUR-WS.org) are the top three journals in this field. United States, United Kingdom, and Canada are the top three countries on the list.

In addition to digital archives and collaboration, there is also a major topic on digital technology included in this list. This analysis shows that digital archives research focused on five major themes: archive, digital, memory, big data, and education. Therefore, understanding how the digital archives knowledge base has evolved is made easier with the help of this tool. Second, by establishing source and document co-citation networks, this research examined the intellectual underpinnings of the digital archives research base. Third, a network of the most prolific authors, the most influential texts, and a thematic map were constructed. These contributed to a better understanding of digital archives' future (emerging topics).



In addition, future the advance technology to digital archives research may focus on artificial intelligence, deep learning, and machine learning, according to the findings of this study. These results add to the growing body of work in digital archives. Fourth, by creating a network of co-authors, the research showed how publishing academics are now collaborating. Digital archives have a slew of publishing teams, as seen by the graph. These may be beneficial to authors in terms of looking back at past partnerships and creating new ones.

This article provides a thorough overview of the digital archives knowledge base, allowing scholars to quickly access a variety of trends, topics, and streams. However, this research examined just the bibliometric data associated with the articles, not the whole text. While it is anticipated that the keywords used will accurately represent the primary substance of the text, some restrictions may apply. As a result, integrating this evaluation with more traditional methods would significantly raise the cost of such research. This contribution is intended to assist academics in advance technology to digital archives research and focus their efforts on future highly relevant research.

## REFERENCES

1. Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of informetrics*, 11(4), 959-975.
2. Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
3. Borgman, C. L., Scharnhorst, A., & Golshan, M. S. (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science and Technology*, 70(8), 888-904.
4. Cook, T. (1993). The concept of archival fonds and the post-custodial era: Theory, problems and solutions. *Archiviaria*, 35, 24–37
5. Duranti, L. (1998). *Diplomatics: New uses for an old science*. Lanham, Maryland, USA: Society of American Archivists and Association of Canadian Archivists in association with Scarecrow Press.
6. Gilliland-Swetland, A. J. (2000). *Enduring paradigm, new opportunities: The value of the archival perspective in the digital environment*. Washington, DC, USA: Council on Library and Information Resources.
7. Hauswedell, T., Nyhan, J., Beals, M. H., Terras, M., & Bell, E. (2020). Of global reach yet of situated contexts: an examination of the implicit and explicit selection criteria that shape digital archives of historical newspapers. *Archival Science*, 20(2), 139-165.
8. Leurs, K. (2017). Communication rights from the margins: Politicising young refugees' smartphone pocket archives. *International Communication Gazette*, 79(6-7), 674-698.
9. McQuire, S. (2013). Photography's afterlife: Documentary images and the operational archive. *Journal of Material Culture*, 18(3), 223-241.
10. Meghini, C., Scopigno, R., Richards, J., Wright, H., Geser, G., Cuy, S., ... & Vlachidis, A. (2017). ARIADNE: A research infrastructure for archaeology. *Journal on Computing and Cultural Heritage (JOCCH)*, 10(3), 1-27.
11. Pearce-Moses, R. (2005). *Glossary of archival and records terminology*. Society of American Archivists.
12. Van Eck, N. J., & Waltman, L. (2014). CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *Journal of informetrics*, 8(4), 802-823.
13. Xie, H., Zhang, Y., Wu, Z., & Lv, T. (2020). A Bibliometric Analysis on Land Degradation: Current Status, Development, and Future Directions. *Land*, 9(1), 28. <https://doi.org/10.3390/land9010028>