Preservation of Digital Archives: Systematic Literature Review

Namira Rahmi Zahara[iD], Tamara Adriani Salim[iD]

Faculty of Humanities, Universitas Indonesia, Indonesia

Abstract

Background of the study: Digital archives in institutions or organizations has increased which is marked by the number of institutions that have established archive units. In the process of preservation or preservation of digital archives, it has its own difficulties compared to other media. Preservation is the most strategic activity, because the availability of records is closely related to the acquisition results that are maintained properly so that they can be processed and accessed for users. The implementation of archive preservation is not easy, there are several obstacles that cause the preservation to be less than optimal.

Purpose: This study aims to provide knowledge about the preservation of digital archives as an effort to save important information and data in archives, it requires preservation or preserving archives in digital format. So that further research is needed, so this research is to facilitate researchers in conducting further research.

Method: This study uses a systematic literature review method by combining bibliometric analysis of the literature from the Scopus database for the period 2018 to 2022.

Findings: The results of the study show that preservation is an effort to maintain and store archives, both content or media so that they are durable and longer.

Conclusion: Research topics in the search results regarding the preservation of digital archives are still found that have not been widely studied, namely, data preservation, long-term preservation, digital archives, and web archives, which can provide opportunities for further researchers to examine digital archive preservation

Keywords: archive preservation, digital archive, digital preservation

Introduction

Archives are an element of organizational information systems that are increasingly perceived as important in organizational life and must be managed properly. Problem archives have not been fully paid attention to by the general public, government, or private organizations. Many people still do not know or do not understand the importance and benefits of archives in daily life for individuals and organizations. Seeing the importance of archives as memory centcenters sources of information, the Indonesian government enacted Law Number 43 of 2009 concerning Archives as law on archives to ensure the safety of materials and national responsibility regarding the planning, implementation and, implementation of life nationality whose integrity, security and safety must be maintained and explain that what is meant by archives are recordings of activities or events in various forms and media under the development of information and communication technology made and accepted by state institutions, governments, educational institutions, companies, political organizations, community organizations and individuals in the implementation of social life, nation and state.

Technological developments also affect aspects of life, one of which is the field of archives, namely digitizing archives, and digitizing is done to create archives documents in digital form with the aim that the work carried out can be effective and efficient. One-way institutions or organizations manage archives digital is establishing an archive unit that aims to maintain availability. In Indonesia, the management of digital archives in institutions or organizations is experiencing problems marked by the number of institutions that set up archive units (Hakim, 2015). This means that these institutions recognize the value of archives. So that the usefulness of digital archives can take place continuously is also necessary efforts to preserve it. Preservation or preservation of digital archives is becoming part of digital archive management carried out for long-term preservation so that information can last a long time (Sugiharto, 2010). The process of preservation or preservation of digital archives has its difficulties compared to other media. Preservation is the most strategic activity because the availability of archives is related closely to acquisitions that are properly maintained so that they can be processed and accessed by the user. Preservation is an effort to maintain and store archives, contents, and media to make it durable and longer. The implementation of archive preservation is not easy. Several obstacles cause the preservation to be less than optimal. It is possible to store conventional archives in digital form. The emergence of a new era in information technology, namely the era of cloud computing, provides opportunities for the community, agencies, government institutions, and the private sector, especially archives, to preserve archives in digital form.

Preservation is the most strategic activity because the availability of records is closely related to the acquisition results that are maintained properly so that they can be processed and accessed by users. Preservation is an effort to maintain and store archives, content and media, so they are durable and longer. Implementing archive preservation is not easy. Several obstacles cause the preservation to be less than optimal. It is possible to store conventional archives in digital form. The emergence of a new era in information technology, namely the era of cloud computing, provides opportunities for the public, agencies, government agencies, and the private sector, especially archival institutions, to preserve archives in digital form.

Digital archives are a form of recorded information stored in electronic media that can be accessed and changed easily, requiring a tool to view, and read or listen to digital archives (Read and Ginn, 2010). Just as archives are stored in collections, so digital archives are also stored in the form of collections, where the digital collection itself is everything that is grouped, given a file name, and stored in electronic form. Lazinger (2001) states that digital archive

collections can be divided into two groups: digital archive collections converted from printed archives into electronic or digital media and archive collections born in digital form (born-digital).

According to Pendit (2008), based on the source of information, the nature of the media, and its content, digital archive collections can be divided into 4 (four), namely: 1) Full text digital archive collections, which include e-journals, open digital collections. access), ebooks, e-newspapers, and digital theses and dissertations; 2) A collection of digital archives of metadata, including digital software in the form of catalogs, indexes, abstracts, or other resources that provide information; 3) Digital multimedia materials; 4) Various sites on the internet

According to Barthos (2003), preservation is an action that includes aspects of efforts to preserve library and archive materials which include policies on financial processing, labor methods, and storage techniques. Preservation is carried out on original library materials or archives that become collections, ancient manuscripts or manuscripts, and historical heritage books of high value from previous generations. Nufus (2017) preservation can also be understood as an activity that ensures that the information contained in the archive can still be used as needed. Meanwhile, digital archive preservation is a series of actions or activities to maintain digital collections in media that can be used and utilized for a long time (Kenny and Rieger, 2000). So, it can be interpreted that preservation is an effort to maintain by storing or transferring conventional archival material into digital form so that the available collections can be used later if needed without any damage. Several observations reveal that digital archive collections are very vulnerable to damage. It is because technology is constantly changing, including hardware and software used.

This study aims to provide knowledge about the preservation of digital archives as an effort to save important information and data in archives. It requires the preservation or preserving archives in digital format. So that further research is needed, this research is to facilitate researchers in conducting further research.

Method

This study uses the Systematic Literature Review method by combining bibliometric analysis from the SCOPUS database period 2018 to 2022. This method is one of the methods in reviewing literature reviews using the VOSviewer application, which is an application that is often used to conduct bibliographic research. VOSviewer is used to visualize bibliographies or data sets containing bibliographic fields (title, author, author, journal name, and so on). In the world of research, VOSviewer is used for bibliometric analysis, looking for topics that still have opportunities to be researched, finding the most widely used references in certain fields and so on.

The first stage. At this stage, the researchers carried out several processes, namely 1). Download the findings of journal articles sourced from Scopus. 2). data is saved in Excel 3 CSV format. Excel CSV data was analyzed using the VOSviewer application to get a visual. 4). The analysis results with VOSviewer are written on the paper.

Second stage. At this stage, researchers are running to get results about the network of authors who have contributed to digital archive preservation research in the 2018-2022 period. The results of data analysis are obtained in the form of images that show the map and the themes that appear based on the categorization in the VOSviewer program output, which contains data visualization such as 1) the size of the connecting line and the circle. This is related to the size of the numbers from the VOSviewer analysis, 2). some of these numbers are divided into links

(networks owned) by link strength (calculated based on full or fractional counting) and the number of occurrences. In addition, several types of analysis were carried out in this paper, including a). The citation will visualize the observed document. Documents tested/observed will be linked with other documents if they cite other articles that are equally observed. This analysis is useful for showing citations between documents, b). Bibliographic coupling articles are tested by visualizing and networking if they have the same reference. This analysis shows the closeness of the study between documents. c). Co-authorship, analyzing the author's collaboration with other authors. This analysis will visualize the results by author's name, author's organization, or author's country of origin. The output of VOSviewer has three visualization displays: network, overlay, and density.

Formulating Research Questions

At the planning stage, the research questions for this paper are structured based on the PICOC criteria (Population, Intervention, Comparison, Outcomes, and Context). The scope identification will be arranged in table 1:

Table 1: Identify the scope for structured questions

Scope	Criteria
Population	Preservation of digital archives, preservation of archives, digital archives
Intervention	Digital archive preservation intervention
Comparison	Comparison not applied
Outcomes	Digital preservation, preservation of digital archives
Context	Review of all literature related to digital archive preservation

(Source: Research Data, 2022)

The following are research questions in this paper:
1. RQ1: How many Items and digital archive preservation clusters were found from mapping keyword research from VOSviewer?
2. RQ2: How many studies on the research topic have been carried out about preserving digital archives from Scopus?
3. RQ3: How is the development of digital archive preservation publications based on the results search from Scopus?

Search Strategy

The research strategy is a search strategy to find sources that match the research question. This stage includes the execution stage. This research uses the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) approach; namely, the data search strategy is shown through the search terms used, data sources from online databases, inclusion criteria, and the exclusions used, assessing the quality of search results and explaining the results of data searches (Handayani, 2017).

Search structure

(digital AND archive AND preservation)

The data collected in the previous stage will be analyzed at this stage. The results that have been analyzed will answer all the research questions that have been previously determined.

Literature Source

The research literature sources are based on the Scopus portal electronic database to extract data by title and abstract information. Several papers were cited using the snowball search technique through electronic databases. Relevant papers are then saved for later selection. The literature cited in this review paper considers publications between 2018 and 2022.

Inclusion and Exclusion Criteria

Inclusion and inclusion are used to select based on criteria and focus on journal articles that suit the needs of researchers. The inclusion and inclusion criteria given will be shown in table 2:

Table 2: Inclusion Criteria and Exclusion Criteria

Inclusion	Exclusion
All literature included in journal articles	All literature not included in journal articles category
All journal articles published from 2018-2022	Journal articles published before 2018-2022
	Journal articles that have duplicates in the database
	All literature that cannot be accessed in full

(Source: Research Data, 2022)

Result and Discussion

In the results of previous studies related to the research to be carried out, researchers seek to find comparisons, to find new inspiration for further research, in addition to helping to show the originality of the research.

The first research conducted by Nuri Ifka Bengi (2021), "Preservation of digital archives as an effort to save information in the era of cloud computing" analyzes that digital archives are formed as a result of rapidly developing information technology. Digital preservation has a goal so that anyone who needs archives in digital format now or in the future can access the information. If there is a problem in the organization or individual accessing files, they can still access the stored files without encountering any problems, such as lost or damaged. Digital archives can be evidence of the problems faced to become decisions regarding the problems faced by the institution, person, agency or organization. Meanwhile, according to Ade Nufus, et al (2018) in his research,"Archive Preservation" analyzes that the value of the information contained in the archive is valuable information, then the archive must be maintained, handled properly and preserved, in accordance with the basic archive materials to maintain sustainability. Archive life is carried out throughout the archive life cycle. Preservation aims to prevent the risk of damage from occurring in a fast time. Proper handling can repair damaged parts and keep information accessible. Nurul Fikriati Ayu Hapsari and Catharina Labore Tatiek Ariyani (2018) in their research,"The Urgency of Digital Archive Preservation"

analyzes that in preservation, digital archive storage media is very at risk of experiencing a degradation process because it is not designed to last for a long time. Preservation of digital archives is not only considered from storage media, it is also necessary to pay attention to preserving the technology used to access digital media and intellectual preservation used for long-term preservation so that the important information contained can still be utilized by the next generation.

Systematic Review Selection Results (PRISMA Approach)

Based on the search strategy results from scientific journals (Scopus) from 2018-2022, there were 1,660 journals. Furthermore, the researchers selected based on exclusion and inclusion criteria so articles were not included. The results of journal articles that will be reviewed further based on the selection are 11 articles
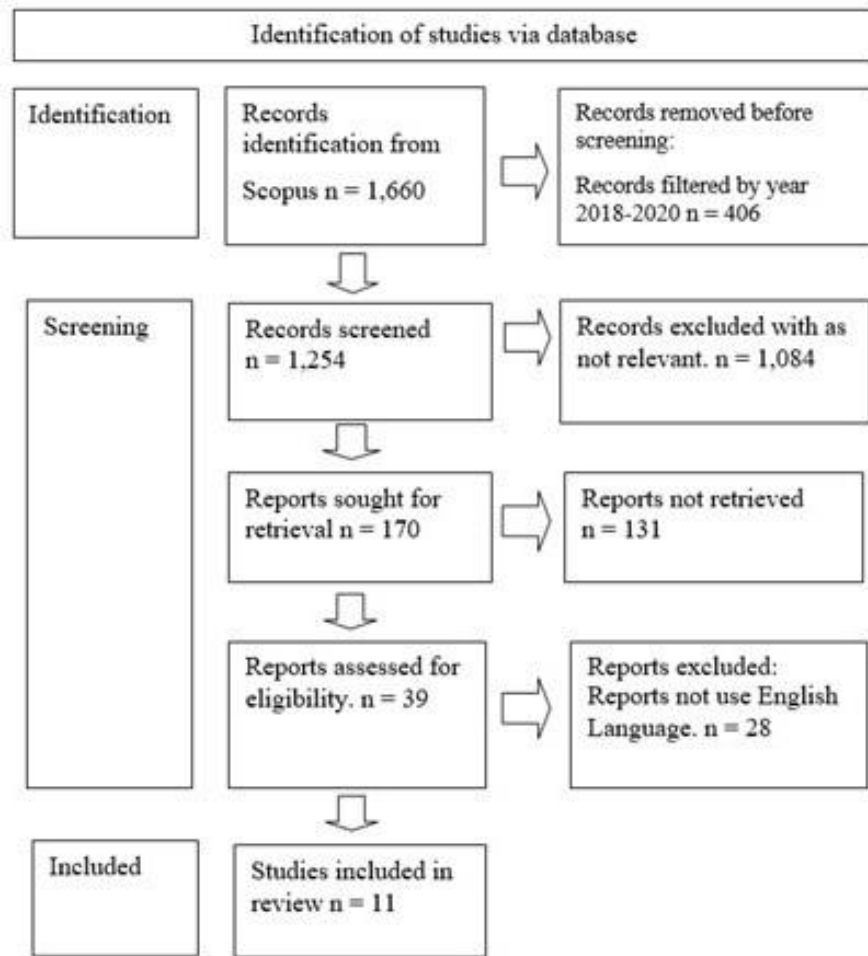

Figure 1. Article Selection Results based on PRISMA Approach
(Source: Research Data, 2022)

At this stage the data that has been found will be evaluated based on the following questions:
1. RQ1: How many items and clusters were found from the research keyword mapping of VOSviewer?

Based on the results of a search conducted from selected journals and used in the assessment of Digital Archive Preservation, 56 items in 6 clusters were found.

Table 3. Mapping Keyword

Clusters	Keywords that appear
1	Archaeology, archive, archives, born digital, digital preservation, digital storage, history, libraries, museums, photography, web archives, web archiving, website
2	Article, digitization, education, heritage conservation, human, information center, information processing, internet, log-term preservation, open access, preservation, procedures
3	3d modeling, 3d modeling, architectural design, architectural heritage, building information model-bim, digital technologies, historic preservation, image reconstruction, photogrammetry, restoration, three dimensional computer graphics
4	Big data, data management, data preservation, data preservations, database, digital archives, information management, information systems, information use, visualization
5	Cultural heritage, cultural heritages, digital archives, digitization, virtual reality
6	Digital humanities, information retrieval, semantics, surveys

(Source: VOSviewer June 22, 2022)
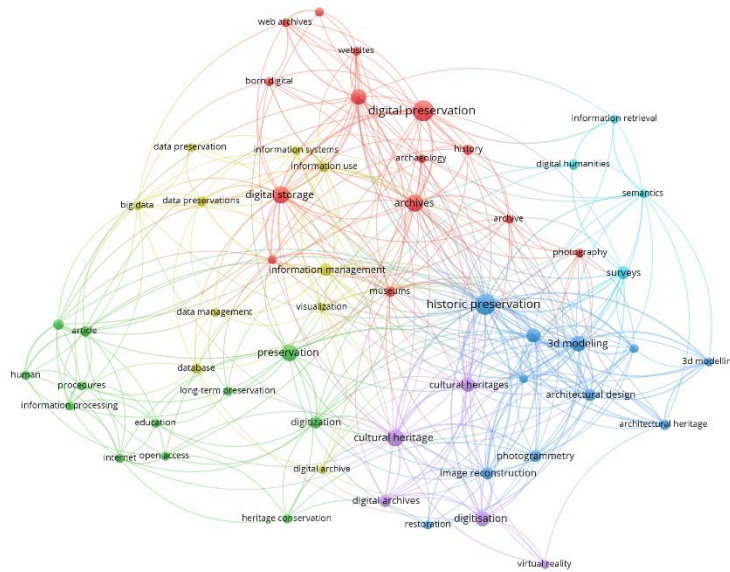
Figure 2. Network Visualization
(Source: VOSviewer June, 2022)

From Figure 2, we get a network of themes that have been researched by researchers related to Digital Archive Preservation, the bigger the circle, the more studied, such as digital preservation, historic preservation, preservation, cultural heritage, archives, digital libraries, 3d modelling, digital storage, the farther and smaller the circle is, the less attention is given to these

themes such as web archiving, web archives, born-digital, websites, history, archaeology, data preservation, semantics, digital humanities, information retrieval, big data, data preservations, information systems, information use, data management, articles, databases, long-term preservation and others.
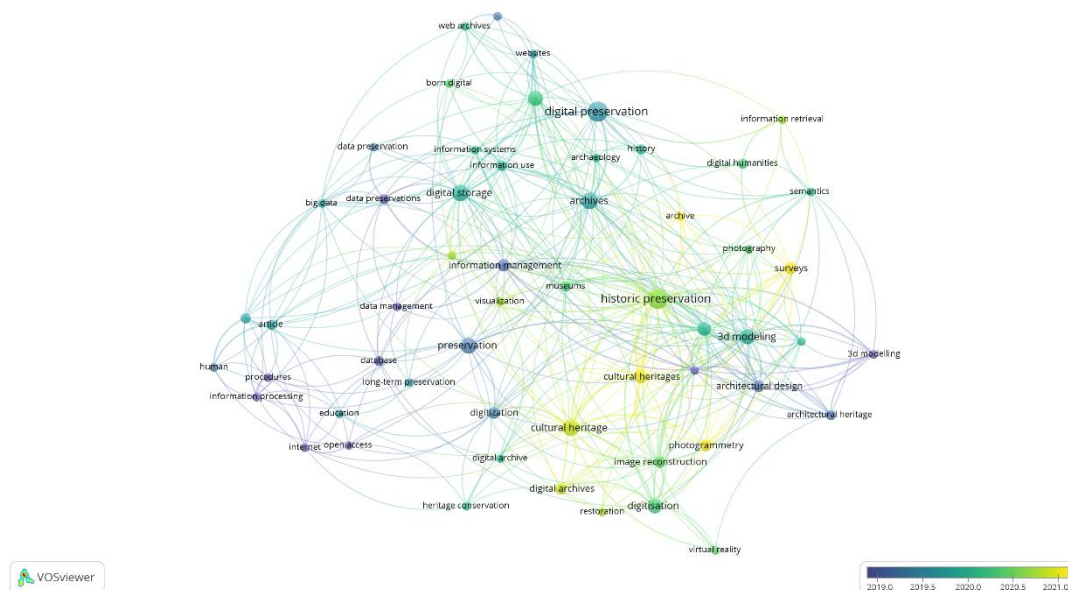

Figure 3. Overlay Visualization
(Source: VOSviewer June, 2022)

From Figure 3, you can see the strengths and weaknesses of the network of themes by paying close attention to the size of the lines connecting the themes in the Network that were created and a description of the period of the article and the research was carried out with a periodization between 2018-2020. colored blue to green with keywords such as digital preservation, archives, web archiving, web archiving, data preservation, data preservations, information management, preservation, database, human, procedures, information processing, internet, open access, digitization, long-term preservation, education, internet, architectural design, 3d modeling, 3d modeling, semantics, photography, information management, building information model-b, information use, information systems, architectural heritage, architectural design, then green to yellow 2020-2022 with keywords such as historic preservation, surveys, cultural history heritages, cultural heritage, image reconstruction, digital archives, digitation, virtual reality, surveys, archives, digital technologies, libraries, visualization.

Figure 4. Density Visualization
(Source: VOSviewer June, 2022)

From Figure 4, it can be illustrated that the themes that are still far from being a big circle are those that are still under-explored, especially with the preservation of digital archives. These themes in terms of the results of the VOSviewer are keywords or variables that can still be, open to the inside to be used as research material in the future, and possibly can be used as research gaps by further researchers.

2. RQ 2: How many studies relevant to the research topic have been carried out about the preservation of digital archives?

Based on the findings of articles relevant to the research topic, there were 11 journals regarding the preservation of digital archives.

Table 4. Findings of relevant articles

Code	Author	Title	Journal	Year
A1	Moles, Nathan	Preservation for diverse users: digital preservation and the "Designated Community" at the Ontario Jewish Archives	Journal of documentation, 03/2022, Volume 78, Issue 3	2022
A2	Ismail, Adila; Bullah Affandy,	Conceptual Paper: Digital Preservation Strategies in Archival Institution	MATEC Web of Conferences, 2018, Volume 150	2018
A3	Oyelude,	Galleries, libraries, archives	Library hi tech news,	2019

	Author	Title	Source	Year
	Adetoun A	and museums (glams) and digital preservation activities	09/2019, Volume 36, Issue 6	
A4	Buragohain, Dibanjyoti; Deka, Manashjyoti; Kumar, Amit	Documentation and Preservation of Endangered Manuscripts through Digital Archiving in North-Eastern States of India	Library Philosophy and Practice; Lincoln (Jan 2022): 1-23.	2022
A5	Anderson, C.B	Preservation and archiving of digital media	Information Services and Use. 40(3), pp. 201-208	2020
A6	Pütt, Karin	Documentation and Digital Preservation of Syrian Heritage: A German Archive Project for Syria	The Public historian, 2018, Volume 40, Issue 4	2018
A7	Acker, Amelia	Emulation practices for software preservation in libraries, archives, and museums	Journal of the Association for Information Science and Technology, 09/2021, Volume 72, Issue 9	2020
A8	Tripathi, S.	Digital preservation: Some underlying issues for long-term preservation.	Library Hi Tech News.	2018
A9	Bralić, V., Stančić, H., & Stengård, M.	A blockchain approach to digital archiving: digital signature certification chain preservation.	Records Management Journal.	2020
A10	Woodall, A., & Ringel, S.	Blockchain archival discourse: Trust and the imaginaries of digital preservation.	new media & society, 22(12), 2200-2217.	2020
A11	Matlala, E.	Long-term preservation of digital records at the University of KwaZulu-Natal archives.	Journal of the South African Society of Archivists, 52, 95-109.	2019

(Source: Research Data, 2022)

3. RQ3: How is the development of digital archive preservation publications based on the results search from Scopus?

Based on the search results on the Scopus database, it shows that the development of digital archive preservation research during the 2018-2020 period is always changing, as shown in table 5 and figure 5.

Table 5. Number of publications research by year

Year of Publication	Quantity
2022	8

2021	21
2020	16
2019	13
2018	16
Total	74

(Source: Research Data, 2022)

Fig. 5: Pie chart of publication progress research by year
(Source: Research Data, 2022)

From table 5 and figure 5 from 2018 to 2022, changes in the development of digital archive preservation research began to occur since 2018 which reached 16 publications and in 2019 decreased to 13 publications and then increased again in 2020, namely 16 publications and an increase back in 2021 to reach 21 publications and in 2022 a total of 8 publications.

Conclusion

Research findings on VOSviewer show 6 clusters and 56 research items on digital archive preservation. Research searches using Scopus from articles that are relevant to the research topic obtained as many as 11 articles, and this digital archive preservation research has changed the number of research developments in the 2018-2022 period. The limitation of this study is in the research process, some of the limitations experienced and can be several factors, in order to be able to pay more attention to future researchers in further perfecting their research because this research itself has shortcomings that need to be improved in future studies. Some of the limitations in this study are, the time required is still relatively short, 2018-2022, research topics in search results regarding the preservation of digital archives are still found that have not been widely studied, namely, data preservation, long-term preservation, digital archives, and web archives, which can provide opportunities for further researchers to examine the preservation of digital archives. So, researchers have suggestions that are expected to provide input for further digital archive preservation researchers.
1. Continually strive to preserve the preservation of digital archives, where information and

communication technology is the main element of changes in preservation systems and strategies.

2. Study more in the preservation of digital archives on data preservation, long-term preservation, digital archives, and web archives

The results of these two studies states that archive preservation can help save current and future information if handled properly, and it is necessary to pay attention to a preservation system that can keep up with changes in technological developments.

Implications of theorists, the findings in this study provide empirical support for the theories that have been presented, including;

1. The results of this study prove that digital preservation has a goal so that anyone who needs archives in digital format now or in the future can access the information. This is in line with the research by Pendit (2008: 248) which states that digital preservation is a planned and managed activity to ensure that digital materials can be used if possible (Nurrohmah Hidayah, et al. 2019)

2. The results of this study prove that technology can change the preservation system in order to save the information contained. This is in line with research by Hitchcock & Tarrant (2010) in Oehlerts & Shu (2013) which shows that digital archive preservation plans must consider the possibility of changes in access and can guarantee their use from time to time or in the long term. To accommodate all digital archival content, preservation plans must also consider changes in various formats and various versions of formats. Ideally, the digital archive preservation plan should allow format flexibility, namely the ability to be widely used in all existing and future formats (Hapsari & Ariyani, 2018)

Managerial Implications, this research is expected to provide various benefits for further researchers who want to research the preservation of digital archives, including;

1. The results of this study can be used as material for consideration and evaluation regarding motivation, opportunities, processes, problems in digital archive preservation

2. Identification of various obstacles or problems faced by digital archive preservation systems against technological developments so that they can be used as guidelines for future improvements.

If the preservation of digital documents is carried out properly in accordance with the procedures of the digitization system and provisions, then digital documents will not be fragile or damaged due to rapid technological changes. This makes the preservation of digital archives important to pay attention to.

References

Acker, A. (2021). Emulation practices for software preservation in libraries, archives, and museums. *Journal of the Association for Information Science and Technology, 72*(9), 1148-1160.

Anderson, C. B. (2020). Preservation and archiving of digital media. *Information Services & Use, 40*(3), 201-208.

Bengi, N.I. (2021). Preservasi arsip digital sebagai upaya penyelamatan informasi di era cloud computing. *IJAL (Indonesian Journal of Academic Librarianship), 5(*1), 35–41. Retrieved from http://journals.apptisjatim.org/index.php/ijal/article/view/93

Barthos, B. (2003). *Manajemen Kearsipan: untuk Lembaga Negara, Swasta, dan Perguruan Tinggi.* Jakarta: Bumi Aksara

Bralić, V., Stančić, H., & Stengård, M. (2020). A blockchain approach to digital archiving: digital signature certification chain preservation. *Records Management Journal,*

30(3), 345-362. https://doi.org/10.1108/RMJ-08-2019-0043

Buragohain, D., Deka, M., & Kumar, A. (2022). Documentation and Preservation of Endangered Manuscripts through Digital Archiving in North-Eastern States of India. *Library Philosophy and Practice, 1-23.*

Handayani, P.W. (2017). *Systematic Review dengan PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses.* Makalah Workshop Riset Sistem Informasi Fakultas Ilmu Komputer UI, 1-3 Agustus

Hakim, H. A. B. (2015). Omeka: aplikasi pengelola arsip digital dalam berbagai format. *Jupiter, 17*(2), 31–37

Hapsari, N. F. A., & Ariyani, C. L. T. (2018). Urgency Preservation of Digital Archives. *Record and Library Journal, 4*(2), 127-138.

Hidayah, N., & Saufa, A. F. (2019). Preservasi Digital Arsip Naskah Kuno: Studi Kasus Preservasi Arsip di Badan Arsip dan Perpustakaan Provinsi Jawa Tengah. *JIPI (Jurnal Ilmu Perpustakaan dan Informasi), 4*(1), 41-51.

Ismail, A., & Affandy, H. B. (2018). Conceptual Paper: Digital Preservation Strategies in Archival Institution. *In MATEC Web of Conferences (Vol. 150, p. 05052). EDP Sciences.*

Kenney A. & Rieger O., (2000). Introduction: Moving Theory into Practice. in Kenney A. & Rieger O. (eds.), *Moving Theory into Practice: Digital Imaging for Libraries and Archives*. Mountain View, CA.

Lazinger S. (2001). Digital Preservation and Meta Data: History, Theory, Practice. Englewood: Libraries Unlimited.

Matlala, E. (2019). Long-term preservation of digital records at the University of KwaZulu-Natal archives. *Journal of the South African Society of Archivists, 52*, 95-109.

Moles, N. (2021). Preservation for diverse users: digital preservation and the "Designated Community" at the Ontario Jewish Archives. *Journal of Documentation, 78*(3), 613-630. https://doi.org/10.1108/JD-02-2021-0041

Nufus, A. (2018). Preservasi Arsip. *Libria, 9*(2), 211-226.

Oyelude, A. A. (2019). Galleries, libraries, archives and museums (glams) and digital preservation activities. *Library Hi Tech News, 36*(6). 8-9. https://doi.org/10.1108/LHTN-06-2019-0039

Pendit, P. L. (2008). *Perpustakaan digital dari A sampai Z.* Jakarta: Cita Karyakarsa Mandiri

Pütt, K. (2018). Documentation and Digital Preservation of Syrian Heritage: A German Archive Project for Syria. *The Public Historian, 40*(4), 107-128.

Read, J., Ginn, M. L. (2010). *Records management*. South-Western USA: Cengage Learning.

Sugiharto, D. (2010). Penyelamatan informasi dokumen/arsip di era teknologi digital. *Baca, 31*(1), 51–64. http://dx.doi.org/10.14203/j.baca.v31i1.96

Tripathi, S. (2018). Digital preservation: Some underlying issues for long-term preservation. *Library Hi Tech News, 35*(2), 8-12. https://doi.org/10.1108/LHTN-09-2017-0067

Tupan, T. (2016). Pemetaan Bibliometrik Dengan Vosviewer Terhadap Perkembangan Hasil Penelitian Bidang Pertanian di Indonesia. *VISI PUSTAKA: Buletin Jaringan Informasi Antar Perpustakaan, 18*(3), 217-230.

Woodall, A., & Ringel, S. (2020). Blockchain archival discourse: Trust and the imaginaries of digital preservation. *New Media & Society, 22*(12), 2200-2217