

***Publication Trends on Digital Archive on the Scopus
Platform: A Bibliometric Study.***

Kamel CAID.

- University of Mohamed Khider; Biskra (Algeria).

- Email: k.eesaid@univ-biskra.dz

Received:03/03/2026; Accepted:05/04/2026; Published:15/06/2026.

Abstract:

The global transition toward digital governance has positioned the digital archive as one of the most consequential outcomes of this transformation, garnering sustained attention from public institutions and specialized organizations alike. This prominence has inspired scholars and practitioners to explore and debate diverse approaches to its creation, storage, accessibility, and preservation, all in service of safeguarding national memory. Such discourse has been disseminated through a substantial corpus of peer-reviewed scientific literature indexed in academic databases, thereby motivating the present bibliometric investigation within the humanities and social sciences. The study aims to measure and analyze open-access publishing trends in English on digital archiving, as indexed in the Socupis database, across the period 2000–2025. Findings indicate a paradigmatic shift in research focus: from narrowly technical concerns toward conceptualizations of the digital archive as an epistemic infrastructure for knowledge and memory management. Increasingly, scholarship emphasizes collective models that integrate technical expertise with theoretical perspectives. Nevertheless, this evolution remains geographically uneven, with research output disproportionately concentrated in the Global North, reflecting the influence of research policy orientations at the expense of contributions from the Global South.

Keywords: Digital Archive; Archival Governance; Bibliometric Analysis; Digital Sustainability.

Introduction:

One of the most distinctive features of states following the adoption of digital applications and the transition toward digital governance-aimed at making public services more accessible and efficient (Roux, 2010, p. 20)-is the production of archival documents by public and private institutions. These documents have evolved alongside administrative practices, transforming into intangible digital records generated by institutions in the course of their activities (Association de Professionnels de l'Information et de la Documentation en Europe). Others have been digitized from printed formats through scanning, stored as electronic copies in databases, and linked with metadata for retrieval (Person, 2017).

The shift toward digital archives and the development of related strategies was not merely a response to technological advancements. Rather, it reflected the administrative, social, cultural, and sovereign benefits of digital archiving, particularly in terms of control, speed, and quality. This prompted institutions to adopt proactive measures for collecting and managing digital archives to ensure accessibility both in the present and over the long term. Understanding the conceptual context and administrative processes in which these archives are created (Kallberg, 2012) has been enriched by theoretical and practical debates among experts and researchers, who have sought to advance, refine, and mitigate the risks associated with digital archival practices. In this regard, Hsu and colleagues, in their study *The Trend and Intellectual Structure of Digital Archives Research*, argue that digital archive management has become a rapidly growing field over the past decade, producing a substantial body of academic literature (Hsu, 2015).

The expansion of digital archiving and its theoretical foundations has been closely tied to the remarkable progress in digital archives since the 1990s (Hsu, 2015). This progress

has fostered unprecedented awareness in academic and professional circles of the importance of archival research (Gilliland, 2004), conducted within a multidisciplinary environment encompassing history, administration, computer science, and archival studies (Couture, 2005). Such research has increasingly adopted social, collaborative, and participatory approaches, encouraging user interaction and the integration of diverse tools, most recently social media platforms, into archival practices (Sukarman, 2026).

With the rapid proliferation and expansion of research in this domain, bibliometric analysis has become essential. By employing mathematical and statistical methods, bibliometrics describes scientific activity and assesses its significance over specific timeframes (Donthu & Kumar, 2021). It serves as a fundamental tool for mapping research trends, identifying dominant themes, and highlighting areas of concentrated scholarly attention. As Tupan (2023) notes, bibliometric analysis enables the construction of a comprehensive map of archival developments, the evaluation of prevailing research foci, and the establishment of a strategic foundation for future research agendas.

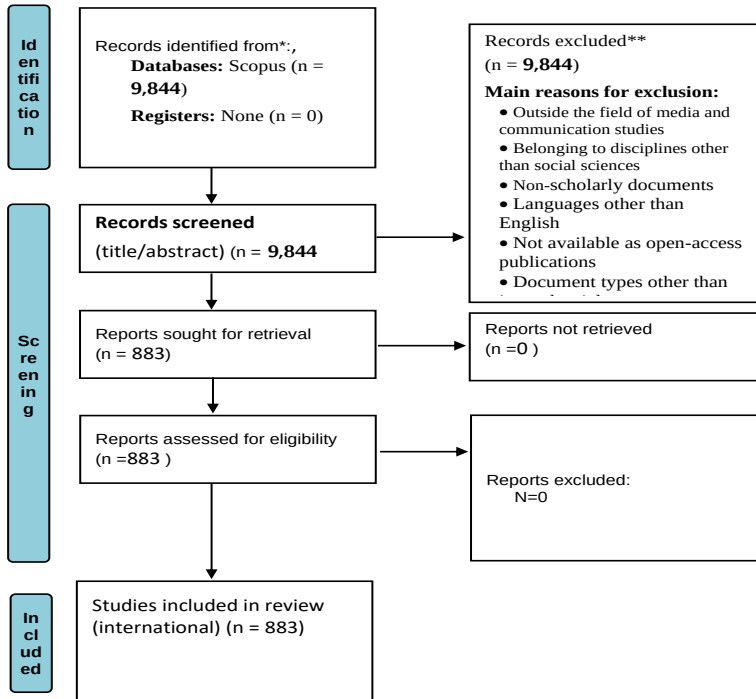
Given the importance of digital archives and the effectiveness of bibliometric methods, this study seeks to address a central question: To what extent are research variables related to digital archiving represented in the global Scopus database? In pursuit of this overarching inquiry, the study aims to answer the following specific questions:

1. What is the volume of scientific articles on digital archiving published between 2000 and 2025?
2. Which countries have produced the largest share of studies on digital archiving worldwide?
3. Which academic journals publish most frequently in this field, and who are the most influential authors?
4. What are the most prominent keywords and thematic trends in this research domain?

1- Methodology:

This study falls within the domain of quantitative research with a bibliometric orientation. Its primary objective is to trace and analyze the trajectory of scholarly production on digital archiving within the social sciences, employing a descriptive-analytical approach grounded in bibliometric indicators to uncover prevailing publication patterns and thematic trends. A structured methodological pathway was adopted, inspired by the general procedural principles of the PRISMA model-not as an evaluative framework for reviews, but as an organizational mechanism to regulate the stages of data collection and selection, thereby ensuring transparency, verifiability, and replicability.

The data source was restricted to the Scopus database, chosen for its extensive coverage of peer-reviewed multidisciplinary journals, its precision in classifying scientific fields, and its provision of rich metadata that supports bibliometric analysis in the social sciences. As illustrated in the subsequent figure, this design facilitated the construction of a robust analytical framework aligned with international standards.



1-1- Identification Phase:

Insert in this section the whole body of your text, start with the research procedures, then present your data, and discuss your results. Reduce the titles to the maximum, and do not number them. All tables, charts, diagrams or maps must have a title, number, and legend; and they must be embedded in the body of the text as an image; otherwise your article will be rejected. Times New Roman, Size-12, Interline simple.

In this stage, searches were conducted across the fields of title, abstract, and keywords (TITLE-ABS-KEY), employing a set of synonymous terms that capture the concept of the digital archive in its various formulations. The search equation was as follows:

TITLE-ABS-KEY (("digital archive*" OR "digital archive*" OR "electronic archive*" OR "digital records" OR "digital preservation"))

This initial process yielded 9,844 documents, representing the total body of scholarly output related to the topic, without restrictions on classification or time frame.

1-2- Screening Phase:

The retrieved results were subjected to a series of systematic filtering criteria. These included restricting the documents to the period 2000–2025, limiting the corpus to peer-reviewed scientific articles, written in English, classified under the social sciences (SOC), and available in open access. These constraints were applied using the database's filtering tools, as illustrated in the following equation:

TITLE-ABS-KEY (("digital archive*" OR "digital archive*" OR "electronic archive*" OR "digital records" OR "digital preservation")) AND (LIMIT-TO (SUBAREA, "SOC")) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (OA, "all"))

This stage reduced the sample size to 883 documents.

1-3- Eligibility Phase:

A qualitative review of document titles and abstracts was then undertaken to verify their substantive relevance to digital archiving in social, institutional, or epistemic contexts. Studies of a purely technical nature, or those focusing exclusively on engineering or information systems without analytical engagement with social or archival dimensions, were excluded.

1-4- Inclusion Phase:

The final verification process resulted in the selection of documents that met all inclusion criteria. This corpus constituted the bibliometric dataset for the study and was subsequently subjected to quantitative analysis.

2- Analytical Tools and Indicators:

The study employed a comprehensive set of bibliometric indicators, encompassing: the temporal evolution

of scholarly output, the distribution of publications across journals and authors, the geographical spread of research contributions, and the analysis of keyword frequency to extract the conceptual structure and identify dominant research trends in the domain of digital archiving within the social sciences.

This stepwise methodological design enabled the construction of a coherent analytical framework characterized by objectivity and precision, fully aligned with internationally recognized standards in bibliometric research. Such alignment enhances the reliability of the findings and strengthens their potential utility for future studies in the field.

2-1- Temporal Evolution of Scholarly Output (2000–2025):

The bibliometric data spanning the years 2000 to 2025 reveal a gradual yet non-linear upward trajectory in scholarly output on digital archives, a trajectory that reflects the structural transformations experienced by the field at both conceptual and technical levels. The initial phase (2000-2006) was characterized by relatively modest publication volume and marked fluctuations in the number of studies, underscoring the foundational nature of the field. During this period, the digital archive was largely conceived as a technical extension of the traditional archive, with limited attention to issues of long-term preservation or the broader social and epistemic dimensions of digitization.

Beginning in 2007, a phase of gradual expansion in publication activity emerged, coinciding with the introduction of concepts such as digital preservation and digital curation, alongside growing institutional concern with the challenges of digital sustainability. This expansion was particularly driven by the acceleration of national digitization projects and the increasing reliance on born-digital data. The period from 2012 to 2019 witnessed a degree of relative stability in scholarly production, signaling the field's transition from theoretical experimentation to the consolidation of methodological and normative frameworks, supported by the establishment of

specialized research centers and international collaborative programs.

In the most recent phase (2020–2025), the data indicate a significant quantitative surge in publications, attributable to the convergence of several factors: the acceleration of global digital transformation, the impact of the COVID-19 pandemic in reasserting questions of open access and digital memory, and the rise of debates surrounding community archives, big data, and artificial intelligence in records management. This surge reflects the attainment of a new level of epistemic maturity in the field of digital archiving, which has moved beyond purely technical solutions to engage with critical questions of power, ethics, and digital governance.

2-2-Geographical Analysis of Scholarly Production on Digital Archives:

The geographical distribution of scholarly output on digital archiving reveals an uneven epistemic structure, reflecting imbalances of power in knowledge production within the fields of information science and archival studies. The Global North-particularly the United States and Western European countries such as the United Kingdom, Germany, and Canada-dominates the majority of publications. This predominance can be attributed to advanced digital infrastructures, robust national archival institutions, and research policies that actively support innovation in digital preservation and data management.

This geographic concentration positions digital archiving as part of the broader domain of “infrastructure sciences,” which demand substantial technical and financial resources, thereby conferring structural advantages on research institutions in the North. By contrast, the Global South exhibits a more limited presence, both in scale and influence, often confined to individual contributions or applied studies linked to externally funded projects, with little capacity to shape the theoretical agenda of the field.

2-3-Analysis of Journal Distribution in Academic Publishing:

From the perspective of quantitative performance and scholarly influence, sharp disparities emerge in publication volume and citation rates across journals. Sustainability (Switzerland) occupies a leading position, with a substantial number of documents (55,730) and total citations (483,826), alongside a high Cite Score (8.7). This dominance reflects its multidisciplinary scope and capacity to situate digital archiving within broader frameworks such as digital transformation, artificial intelligence, and knowledge sustainability. Yet, this quantitative weight does not necessarily indicate deep specialization in archiving; rather, it suggests the integration of the topic into cross-disciplinary debates with wide citation bases. By contrast, specialized journals such as Archival Science (145 documents, Cite Score 3.5) and D-Lib Magazine (126 documents, Cite Score 2.4) exemplify medium-scale output but with greater conceptual focus, contributing to the theoretical and methodological foundations of digital archiving, albeit within narrower citation networks.

Heritage-oriented journals with a technical emphasis, such as npj Heritage Science, demonstrate a notable balance between volume and impact (1,554 documents, Cite Score 5.6), reflecting the rise of algorithmic and AI-driven approaches to visual and heritage archives. This positions them at the intersection of applied sciences and cultural studies. Conversely, professional or association-based journals such as American Archivist and Archives and Manuscripts record lower Cite Scores (0.7 and 1.5, respectively), a pattern attributable to their primarily local professional readership and their focus on applied practices and archival policy rather than internationally citable knowledge production.

Second, from the standpoint of thematic analysis and epistemic shifts, selected article titles illustrate the field's transition from an emphasis on the technical infrastructure of knowledge exchange (e.g., the Scholix framework in D-Lib

Magazine) to more complex issues related to memory, trauma, and epistemic justice, as seen in Archival Science through the incorporation of concepts such as “trauma” and “archival legacy.” Simultaneously, the notion of the “archive as a digital platform” emerges in journals such as Preservation, Digital Technology and Culture and Internet Archaeology, where the archive is reconceptualized within the contexts of smart cities, digital museums, and digital twins, signaling the field’s expansion into new urban and cultural domains.

Third, at the institutional level of publishing, the table highlights the strong presence of major commercial publishers-particularly Springer Nature, MDPI, and Taylor & Francis-who dominate journals with the highest impact and widest citation reach. In contrast, academic publishers and professional associations (e.g., the Society of American Archivists and the Australian Society of Archivists) sustain journals with lower citation influence but stronger ties to professional discourse, archival ethics, access issues, and cultural ownership, especially in relation to Indigenous archives and intangible heritage.

2-4- Author Distribution and Intellectual Dynamics:

From the perspective of scholarly actors, two principal categories of researchers can be distinguished. The first category represents what may be termed Big Digital Heritage Science. This pattern is exemplified in the work of Christoph Becker (University of Toronto), who records the highest citation count (1,513), a strong h index (19), and prolific output (79 publications). His leadership position reflects the integration of digital preservation into agendas of software sustainability, systems design, and digital infrastructure policies, positioning the researcher as a coordinator of multidisciplinary knowledge networks spanning computer science, archival studies, and institutional sustainability. A similar pattern is observed in Michael Seadle (Humboldt University, Berlin), whose high productivity (114

publications) and strong citation record (497) confirm his pivotal role in linking archival studies with digital access and spoken document retrieval.

Second, the category of Niche Experts emerges, represented by scholars such as Mpho Ngoepe (University of South Africa). His intellectual influence derives from a single highly cited reference article (748 citations) with a moderate h index (16). This underscores the centrality of institutional and governance issues in African national archives. Here, impact is not based on volume of output but on the normative function of a foundational text that serves as a benchmark for evaluating archival systems and records management in developing contexts.

Third, at the level of thematic transformations, the most highly cited articles illustrate a shift in the field from the question “How do we preserve?” to “How do we govern and sustain?”. Paul Conway’s (2010) article on digital preservation in the age of Google represents a pivotal moment in redefining digital archives as a socio technical and political issue, beyond purely technical solutions. This explains its high citation count (462) and respectable h index (11). Similarly, Peter Burnhill’s (2005) work on the Digital Curation Centre marked an institutional turning point, establishing digital preservation as a national system requiring coordination across policies, infrastructures, and standards.

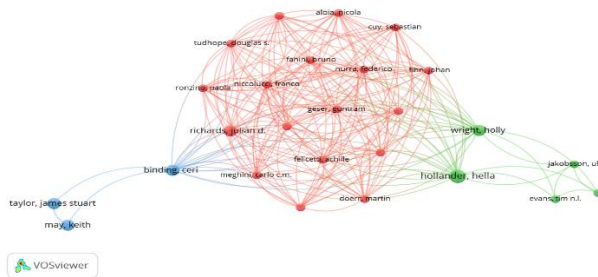
In the same vein, the works of Cooper, Becker, and Seadle reflect the rise of discourses on big data and digital sustainability as new epistemic axes. These integrate digital archives into large scale projects concerning identity, heritage, and collective memory, particularly in European and North American contexts.

Fourth, in terms of the geopolitics of knowledge, there is a clear concentration of high impact output in universities and institutions of the Global North (United States, Canada, Germany, United Kingdom), where research infrastructures, funding sources, and international publishing networks are

abundant. By contrast, the Global South-exemplified by South Africa-produces applied, institutionally oriented knowledge with high normative impact, but less integration into “big science” networks. This reflects the persistence of epistemic inequalities between centers of analysis and regions of application.

Finally, bibliometric indicators such as publication counts and h index suggest that scholarly impact in digital archives is not linearly correlated with productivity. Instead, it depends on the extent to which researchers engage with major structural issues (governance, sustainability, digital infrastructure) or produce foundational texts repeatedly cited in the literature. The case of James Doig (National Archives of Australia) illustrates this divergence: despite recent publication and low citation counts (3), his work reflects a professional-procedural orientation aimed at re engineering record systems rather than achieving international academic influence.

2-5- Interpretive Analysis of the Co Authorship Network:



The co authorship network reveals a clear structural transformation in patterns of knowledge production within the field of digital archives. The centrality of the individual researcher is diminishing in favor of clustered collaborative structures characterized by high degrees of internal cohesion. At the core of the network, a solid epistemic nucleus emerges,

represented by repeated and systematic collaboration among scholars such as Douglas Tudhope, Bruno Fanini, Federico Nurra, and Guntram Geser. This nucleus is distinguished by dense interconnections, reflecting a stable institutional collaboration model based on long term projects in areas such as digital heritage, three dimensional modeling, and semantic infrastructures for archives. This pattern signals a shift from individually descriptive archival studies toward knowledge infrastructure projects requiring coordination across archival science, information science, and computer science.

In contrast, another cluster with a critical-institutional orientation includes researchers such as Hella Hollander and Holly Wright. Their work focuses on issues of preservation ethics, community participation, and active archives. This cluster demonstrates good internal cohesion but connects to the central nucleus through a limited number of intermediary nodes, indicating a relative separation between the dominant technical trajectory and the humanistic–critical trajectory within the field.

At the margins of the network, a group of specialized technical actors appears, including Ceri Binding, James Stuart Taylor, and Keith May. Their contributions involve the development of specific tools and models related to metadata, interoperability, and the semantic web. However, their limited integration into larger social networks positions them more as “knowledge providers” than as central actors shaping the theoretical agenda.

This structure reflects the emergence of epistemic stratification within the field of digital archives:

- A central nucleus producing high impact normative knowledge,
- A critical-social periphery re examining concepts of memory and authority,
- And a technical margin developing tools without monopolizing theoretical discourse.

These dynamics confirm that contemporary archival knowledge production is increasingly conditioned by participation in multidisciplinary teams that constitute the epistemic infrastructure for managing the complexities of digital memory at technical, ethical, and institutional levels.

2-6- Analysis of the Keyword Co Occurrence Network

The keyword co-occurrence map reveals the emergence of a dominant epistemic paradigm centered on the concept of Digital Preservation. Positioned at the heart of the network with a large node size and high linkage density, digital preservation is no longer a secondary technical activity or a supportive procedural function. Rather, it has become the “central epistemic doctrine” around which diverse research axes in the field of digital archives are organized.

This centrality is evident in the strong and direct associations between digital preservation and concepts such as Digital Records, Records Management, Archival Systems, and Long term Storage. Such linkages reflect the field’s transition from conceiving the archive as a static repository of documents to understanding it as a dynamic system for managing temporality, ensuring continuity, and controlling long term technical and institutional risks.

At the same time, a humanistic-social semantic cluster emerges around the concept of Human, connected to terms such as Collective Memory, Citizen Curation, Right to be Forgotten, and Activist Archives. This indicates that scholarly debate has expanded beyond the sustainability of data per se, toward interrogating its relationship with collective memory, identity, and the rights of individuals and communities. In this way, the digital archive is repositioned within an ethical-political space that addresses questions of power, representation, and community participation.

On another level, the network highlights a notable technical-visual cluster encompassing concepts such as Photogrammetry, 3D Modelling, 3D Visualization, and Born digital Data. This reflects the rise of three-dimensional

archiving and the growing reliance on modeling and imaging technologies for preserving both tangible and intangible heritage. However, the positioning of this cluster in a non-dominant but connected relationship with the central nucleus suggests that technology functions as a cognitive mediator rather than the ultimate goal in shaping archival discourse.

In addition, an institutional-applied cluster appears, including terms such as Archival Studies, Applied Research, Audiovisual Archives, and Europeana Libraries. This reflects the practical dimension of the field, translating theoretical frameworks into policies, operational platforms, and institutional systems, particularly within the European context.

Overall, the keyword co-occurrence network illustrates a dual paradigmatic transformation in the field of digital archives: From the archive as a technical system to the archive as a socio-cultural practice, and from preservation as a limited technical procedure to preservation as a long term ethical-political act.

This multi-level model integrates technical, humanistic, and institutional dimensions into a unified epistemic framework, redefining the meaning of memory, sustainability, and knowledge in the digital age.

3- Discussion:

In this study, we analyzed the scholarly output related to digital archives within the Scopus database during the period spanning roughly from the beginning of the twenty first century to 2025. The analysis employed a bibliometric approach that combined temporal analysis, geographical distribution, co authorship networks, and the semantic structure of keyword co-occurrence. The findings demonstrate that the field of digital archives has undergone a profound structural transformation, shifting from a supportive technical domain for traditional archives into a complex, interdisciplinary intellectual field where technological

dimensions intersect with institutional, social, and ethical aspects of managing digital memory.

Temporally, the dynamics of publication reveal a non linear upward trajectory that can be divided into distinct phases. The initial stage was marked by low quantitative output and fluctuating publication rates, reflecting the exploratory and foundational character of early research, which focused primarily on technical digitization and electronic document preservation. Beginning in the second decade of the twenty first century, the field entered a phase of relative stability and methodological expansion, coinciding with the consolidation of concepts such as long term digital preservation, descriptive organization, and international standards. This indicates the transition of digital archives from experimentation to institutionalization. In recent years, a clear quantitative surge in scholarly output has been recorded, reflecting the intellectual maturity of the field and the expansion of research questions to encompass issues such as open access, community archives, born digital data, and the impact of accelerated digital transformation on institutional memory.

At the level of research networks, co authorship analysis revealed the decline of the “individual researcher” model and the rise of clustered structures based on multidisciplinary teams. A central nucleus of researchers and institutions with high internal cohesion has emerged, leading long term structural projects, particularly in digital heritage, semantic modeling, and interoperability of archival systems. This transformation reflects the integration of digital archives into what may be termed “knowledge infrastructure projects,” requiring coordination across archival science, information science, and computer science. In contrast, other research clusters with a critical-social orientation focus on preservation ethics, community participation, and memory policies. However, these remain less integrated into the dominant

technical nucleus, indicating a relative separation between technical and humanistic trajectories within the field.

Geographically, the distribution of scholarly output reveals a clear dominance of the Global North, particularly the United States and Western Europe, in terms of publication volume and scientific impact. This concentration is explained by the possession of advanced digital infrastructures, strong archival institutions, and research policies supportive of innovation.

4- Conclusion:

The findings of this study indicate that scholarly production on digital archives within the Scopus database is clearly moving toward greater structural complexity and interdisciplinary integration. The field has shifted from focusing on partial technical solutions to adopting holistic approaches that conceptualize the digital archive as an epistemic infrastructure for managing knowledge and memory in the digital age. This transformation reflects a growing recognition that issues of digital preservation, accessibility, and authenticity cannot be addressed in isolation from the institutional and social contexts in which digital documents are produced.

The results also highlight that impactful scholarly output is increasingly conditioned by integration into international, multidisciplinary research networks. This confirms what may be termed the “end of the individual archival author” and the rise of collective work models based on the intersection of technical and theoretical expertise. However, this development remains geographically uneven, as structural dominance by the Global North continues to shape the research agenda, while the Global South maintains only a relatively marginal presence.

Accordingly, the study recommends adopting an integrative approach that combines technical and critical-social dimensions in digital archival studies, while strengthening research capacities in underrepresented

geographical contexts. Expanding the scope of international collaboration, supporting community archives, and embedding ethical and political considerations into the design of archival systems represent promising pathways for overcoming epistemic imbalances and building a more equitable and inclusive digital archival field.

Bibliography:

- 1-Association de professionnels de l'information et de la documentation en Europe. (2020, May 5). *L'archivage électronique*. ADBS. <https://www.adbs.fr/ladbs>
- 2-Couture, C., & Ducharme, D. (2005). Research in archival science: A status report. *Archivaria*, 59, 41–67.
- 3-Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- 4-Gilliland, A., & McKemmish, S. (2004). Building an infrastructure for archival research. *Archival Science*, 4(3), 149–197. <https://doi.org/10.1007/s10502-005-2596-0>
- 5-Hsu, F. M., Lin, C. M., & Fang, C. T. (2015). The trend and intellectual structure of digital archives research. In *PACIS 2015 Proceedings (Paper 128)*. Association for Information Systems (AIS) Electronic Library. <https://aisel.aisnet.org/pacis2015/128>
- 6-Hsu, F. M., Lin, C. M., & Fun, C. T. (2015, June). The trend of researches in digital archives. In *Proceedings of the International Conference on Information Technology & Society* (pp. 8–9).
- 7-Kallberg, M. (2012). The records continuum model as a framework for managing digital archives. *Records Management Journal*, 22(2), 98–115. <https://doi.org/10.1108/09565691211268162>
- 8-Person, D. A., & Plumb, T. K. (2017). Digital archives from the ground up. *Digital Library Perspectives*, 33(3), 253–263. <https://doi.org/10.1108/DLP-07-2016-0019>
- 9-Roux, L. (2010). *L'administration électronique: Un vecteur de qualité de service pour les usagers?* Informations sociales, 158(2), 20–29.
- 10-Sukarman, S., Harianto, H., & Amalia, N. (2026). Mapping digital archives research trends: A bibliometric study using the Dimensions database (2020–2024). *Inkunabula: Journal of Library Science and Islamic Information*, 5(1).
- 11-Tupan, T. (2023). Analisis bibliometrik publikasi penelitian kearsipan di Indonesia berbasis data Scopus. *Media Pustakawan*, 30(3), 224–234.

