

# Scientific publication reading sheet

## Bibliographic references

---

**Title:**

[Artificial Intelligence and the Preservation of Historic Documents](#)

**Author:**

Gaute Barlindhaug, University of Tromsø.

## Summary

---

Various initiatives and policies have been set to ensure the conservation of artifacts, manuscripts, and documents that contributes to Norway's rich cultural heritage. Preserving it faces some tough challenges, especially when it comes to digitizing.

Critical digitization, often obstructed by resource constraints and high costs poses a significant obstacle. The selection process in slow digitization can inadvertently introduce bias, providing a distorted view of history. On the other hand, mass digitization with its automated procedures, may get to results that lack the thoughtful consideration needed.

The National Library of Norway has undertaken a monumental task of digitizing the entirety of the country's cultural heritage, driven by the urgency to safeguard visual and sonic materials at risk of deterioration taking a sharp turn on archivists' conventional wisdom. In the initial phase of digitization, the focus is on adding basic retrieval information. Moving into the second phase, artificial intelligence (AI) and machine learning (ML) will be deployed to enhance information and efficiently organize the vast array of digitized material. The Ministry anticipates this task to span on thirty years.

**Keywords:** Digitization, cultural heritage, National Library of Norway, Artificial intelligence, challenges

## Identification of the reading sheet

---

**Author:**

OUAFI Mohammed Zakaria

**Date of creation:**

21 December 2023

## Publication plan

---

The aim of this text is not to go into the technical details of such technology, but rather look at how it is proposed to be used in the context of the Norwegian National Library in ways that enforce political and cultural ideas about cultural heritage preservation.

## Analysis and comments

---

### Challenges to overcome

The challenges of digitizing historic material involve two main approaches: critical (slow) digitization and mass digitization. Critical digitization aims for in-depth transformation of material into the best digital format, but faces challenges like high costs and biased [\[1\]](#) [\[2\]](#) selection processes. Mass digitization, while avoiding biased selection, encounters issues such as reduced technical quality with its automated process [\[3\]](#) and the potential loss of original context [\[4\]](#) when organizing diverse materials into a unified system. Both approaches highlight concerns [\[5\]](#) about preserving historical authenticity and avoiding biases in shaping our digital history.

### The National Library's Unconventional Approach

The National Library of Norway, since 2006, has been digitizing all printed material published in the country [\[6\]](#), aiming to create a comprehensive online library. This digital collection includes books, magazines, newspapers, and more. The library's recent project extends beyond published material to digitize the entire cultural heritage of Norway, addressing the need to preserve visual and sonic material at risk of deterioration. The Ministry of Culture allocated funds for a 30-year project, establishing a Centre for Cultural Heritage Digitization. The digitization process faces challenges in organizing diverse materials and metadata for effective public access.

The National Library's unique approach involves minimal metadata in the initial phase, with plans to enrich it using artificial intelligence (AI) and machine learning (ML) in the second phase. This approach contrasts with traditional archivist strategies, aiming to give equal importance for all cultural heritage and aligning with Norway's cultural policy [\[7\]](#) goals of democratizing access to diverse historical materials. The use of AI and ML is seen as a means to overcome challenges in organizing vast and varied digitized content. However, this approach deviates from some archival community perspectives that highlight the potential loss of context in digital search tools and emphasize the importance of metadata provided by creators.

## Artificial Intelligence and the Enrichment of Metadata

AI and ML have undergone testing for the organization and analysis of digitized archival material within the ongoing 30-year digitization project. In its initial phase, the project primarily concentrates on digitized photography [7]. These technologies are considered potential solutions to the vast diversity and lack of predefined categories for metadata. The library's experimental "Maken" search function allows users to find similarities between documents based on content interpretation rather than metadata. The aim is to connect information across different sources and improve context for materials with limited metadata. Future possibilities include linking digitized photography with newspapers to provide richer insights. Challenges include establishing links between visual and textual material and creating a functional asset manager for the heterogeneous digitized material. The "Maken" search engine offers a user-friendly and flexible approach, allowing users to shape their engagement with the material based on their specific needs and interests. AI and ML have the potential to enable users to build new connections and interpretations of historic material.

## Author's perspective

The author holds the view that there is potential to discover connections between published pictures and unmarked negatives. He also expresses his perspective is shaped by the knowledge that professional photographers often shoot an entire analog roll in a day and develop negatives the same day, where a roll captures images taken in a succession which can provide insights into who, when and where other pictures on the same roll were taken even if there's no clear visual similarity.

---

[1] Schwartz, J. M., & Cook, T. (2002). Archives, records, and power: The making of modern memory. *Archival science*, 2(1), 1–19.

[2] Breckenridge, K. (2014). The politics of the parallel archive: digital imperialism and the future of record-keeping in the age of digital reproduction. *Journal of Southern African Studies*, 40(3), 499–519.

[3] Hughes, L. M. (2004). *Digitizing collections: strategic issues for the information manager* (Vol. 2). Facet Publishing.

[4] Oliver, G. (2012). The digital archive. In Hughes, L. M. (Ed.) *Evaluating and measuring the value, use and impact of digital collections*. Facet Publishing.

[5] Dahlström, M., & Hansson, J. (2019). Documentary Provenance and Digitized Collections: Concepts and Problems. *Proceedings from the Document Academy*, 6(1), 8.

[6] Takle, M. (2009a). The Norwegian national digital library. *Ariadne*, 60.

[7] Colavizza, G., Ehrmann, M., & Bortoluzzi, F. (2019). Index-driven digitization and indexation of historical archives. *Frontiers in Digital Humanities*, 6, 4.