

Pekka Henttonen<sup>1</sup>

## DO WE NEED TO KEEP ORIGINALS AFTER DIGITIZATION?

### Abstract

**Purpose:** *The paper examines reasons for preserving original documents after their digitization.*

**Methodology:** *The research is based on analysis of literature.*

**Results:** *There are many reasons for preserving the original paper documents after digitization. For example, information loss may be unacceptable, originals are needed to prove the authenticity of the records, or there are legal reasons for keeping the originals. These issues must be considered case by case before making decisions about the fate of the records.*

**Conclusions:** *One may destroy the originals after digitization. However, this requires careful consideration.*

**Key words:** *digitization, original records, information loss, authenticity*

## DOBBIAMO CONSERVARE GLI ORIGINALI DOPO LA DIGITALIZZAZIONE?

### Abstract

**Scopo:** *il documento esamina i motivi per cui è opportuno conservare i documenti originali dopo la loro digitalizzazione.*

**Metodologia:** *la ricerca si basa sull'analisi della letteratura.*

**Risultati:** *esistono molti motivi per conservare i documenti cartacei originali dopo la digitalizzazione. Ad esempio, la perdita di informazioni potrebbe essere inaccettabile, gli originali potrebbero essere necessari per dimostrare l'autenticità dei documenti o potrebbero esserci motivi legali per conservare gli originali. Queste questioni devono essere valutate caso per caso prima di prendere decisioni sul destino dei documenti.*

<sup>1</sup> Pekka Henttonen (DSocSci, MHist), associate professor, Faculty of Information Technology and Communication Sciences, Tampere University, Finland, pekka.henttonen@tuni.fi.

**Conclusioni:** È possibile distruggere gli originali dopo la digitalizzazione. Tuttavia, ciò richiede un'attenta valutazione.

**Parole chiave:** digitalizzazione, documenti originali, perdita di informazioni, autenticità

## ALI MORAMO OHRANITI ORIGINALE PO DIGITALIZACIJI?

### **Izveleček**

**Namen:** Članek preučuje razloge za ohranjanje izvirnih dokumentov po njihovi digitalizaciji.

**Metodologija:** Raziskava temelji na analizi literature.

**Rezultati:** Po digitalizaciji izvirnih papirnih dokumentov obstaja veliko razlogov za njihovo ohranitev. Na primer, izguba informacij je lahko nesprejemljiva, originali so potrebni za dokazovanje pristnosti zapisov ali pa obstajajo pravni razlogi za njihovo hrambo. O teh vprašanjih je treba razmisliti od primera do primera, preden se sprejme odločitev o usodi zapisov.

**Zaključki:** Originale je mogoče po digitalizaciji uničiti, vendar to zahteva pre-mišljen razmislek.

**Ključne besede:** digitalizacija, izvirni zapisi, izguba informacij, pristnost

## INTRODUCTION

The environment of archives has been revolutionized by digitalization. The change concerns all aspects of archival work. One of the areas affected is record copies. Today users often have access only to digital copies and the original documents are out of their reach. Generally, this is a positive change. Digital services liberate users and archives from constraints of time and space. The records are accessible everywhere. Digital services bring cost savings and efficiency. This raises the question: for what do we need the originals after digitization? Are they needless?

The question is relevant. Developments in Finland offer an example. The National Archives of Finland decided recently to transfer digitally 70 shelf kilometers of records from the custody of government agencies to the National Archives. The National Archives does not have storage space for the originals in analog form. Thus, the best solution is to digitize the documents and destroy the originals when the transfer is over.

Is this a good solution? Interestingly, historically permanence of information was first seen more important than permanence of the documents containing the information (see O’Toole, 1989). Nevertheless, mass destruction is a rare phenomenon (Hirvonen, 2017). Arguments for preserving original documents are discussed next one by one. Finally, I look at the concept of “intrinsic value” which has been traditionally used to evaluate the value of the originals and finally draw conclusions.

## WITH ORIGINALS SOMETHING IS LOST

A copy is never equivalent to the original, unless the original is a digital object. In that case, and for the first time in history, nothing is lost when making a copy. A digital copy of a digital object is identical with the original in every detail – except for its missing event history (DLM Forum Foundation, 2011, p. 81). In the case of paper documents something is lost in the process. This happens in more than one way.

## LOSING INFORMATION IN DOCUMENTS

Firstly, when a document is digitized, some information is lost. Materiality of the document disappears, and with it, its physical characteristics, like tangibility. When we approach artifacts via their digital simulacra we miss their three-dimensionality, physical characteristics such as paper quality, watermarks, the

ways a letter was folded, or objects inserted. (Ogilvie, 2016.) Loss of materiality means losing information. Materiality carries meanings that may not be reproduced in digital form.

Here are some examples. Brown and Duguid (2002) tell a story about a researcher who in a reading room sniffed documents instead of reading them. The odor in the letters allowed the researcher to identify disinfected letters, and thus, track how an infectious disease had spread around the world. Another example is given by Maryanne Dever. Nature of the relationship between late Swedish movie star Greta Garbo and her friend Mercedes de Acosta has raised discussion. When Garbo's letters to de Acosta became public, newspapers screamed that "they reveal nothing!" Dever looks what this "nothing" looks like and asks for "material literacy." She claims that when one looks how carefully de Acosta had put aside "any paper Garbo's hand could be presumed to have touched" (including blank florist's cards, and short refusals to meet) there is no doubt about the intimate connection between the two women. (Dever, 2010, 2013.) – Can there be material literacy in a digitized collection?

Paper watermarks are a common research interest. A paper texture or qualities of ink can be important for a researcher, too. Analysis of paper texture can be used in dating documents, and identification of cultural relics and ancient books (Lu et al., 2021). The watermarks are not usually visible in digital copies, but a watermark can be discernible, if it is strong and scanned in high resolution under bright or angled light. Besides digitization, there are other methods to reproduce watermarks and mould surface patterns. These include optical methods, and radiography (Zhang et al., 2019). These methods are separate from digitization. Thus, a digital copy is not enough to preserve this information. On the other hand, sometimes digitization renders more legible documents that have been damaged or whose ink has faded (Ogilvie, 2016).

Even when some information is lost, one may argue that this loss is acceptable. It has been suggested only "significant properties", "salient features" or "essential characteristics" are the ones that one needs to preserve. For instance, how an email message is rendered on the screen depends on the software that is used to read it, but regardless of the software the message is essentially the same. The challenge of significant properties is that defining them is culturally loaded, subjective and dependent on context. (Yeo, 2010.)

## **LOSING PHYSICAL RELATIONSHIPS AND PHYSICAL ACCESS TO RECORDS**

Disappearance of physical form of a record means losing physical relationships that a document has as a physical object with other objects in the fonds. Classical archival science emphasized the importance of principle of provenance (respect des fonds and the registry principle), because place of a record in the fonds inadvertently carries information about process, function, actors, and other records that the record is related with.

In contrast, and by default, a digitized document is itemized. When a person accesses it he typically gets a single file from an information system (Ogilvie, 2016). The files have no original order, and they do not automatically form larger collections. Because of this there must be a conscious effort to capture essential features of physical collection in digital form by adding metadata to the digitized items.

Even when this succeeds perfectly digitization changes the way the records are presented to a user and how the user accesses and interacts with them. For instance, instead of a bound folder the user gets a list of files on a webpage, for instance. One cannot easily browse digital documents. On the other hand, one may make searches in records metadata and possibly in their content. Thus, both digital and analog access offer possibilities for “serendipity”, that is, finding useful information unexpectedly (e.g., Duff & Johnson, 2002; Kumpulainen & Kautonen, 2017), but the process is different, and probably its results, too.

While a digitized record may lose its original context it may get a new one as a part of a digitized collection. This new context can be misleading (see Pashkeeva, 2024). “Digital surrogates” are more than copies because the representation of the information changes and the surrogate carries marks from the time it was created (Conway, 2015). Digitized documents have a provenance that is not the same as what the originals had (Podolsky Nordland, 2004; Yeo, 2007).

## **LOSING EMOTIONAL ASPECTS OF WORK IN ARCHIVES**

With physical form we also lose something intangible that is hard to define: emotional aspects of work in archives. A Finnish historian once noted: “I visit archives with same enthusiasm as a movie freak goes to a cinema.” Perhaps digital archives may raise similar sentiments, but clearly the experience of accessing documents is different. Many researchers and archivists can bear witness to what

touching an original document may feel about. Having the original document in your hand gives one a feeling of immediate access to the past. This feeling of unmediated access is to some extent misleading. Both researchers and archivists prefer to ignore processes of arrangement, description, and appraisal that have brought archival documents to their reach (Schwartz & Cook, 2002). Still, it is part of the research and work in archives

Siiri Savinotko (2025) suggests concept of “tangible archive.” She notes that concepts like “digital archive” and “physical archive” hide bodily and emotional aspects of user interaction with archival documents. There is reciprocity: a user touches a physical document, and the document touches emotions of the user.

## **LOSING BACKUP FOR DIGITAL COPIES**

If digitization now means losing information, could future technological development save us from this loss? Certainly, technology has improved all the time. There are several issues complicating digitization and processing of digitized historical records. These issues include uneven illumination, contrast variation, bleed-through degradation, faded ink or faint characters, smear and smudge, blur, and deteriorated documents. (Sulaiman et al., 2019.) When records are digitized perhaps OCR (Optical Character Recognition) or HTR (Hand-Written Text Recognition) in mind, who can say that tomorrow we might not get better results than today? Perhaps there will be a new technology capturing features paper watermark, mould surface patterns, and information about fiber structure on one stroke? If originals exist one may reprocess them, but if they are not there, the quality of current copies sets the limit for future improvements.

One may argue that the originals are always a good backup for digital versions. The information is there in analog form, that is, in a form that is independent from any catastrophe that might hit the digital environment. Previously digital information was considered more fragile and more difficult to preserve for long times than traditional paper documents (Gertz, 2000). The persuasiveness of this argument seems to have lessened in recent years. In the aftermath of the war in Ukraine, digitization has emerged as the best way to protect cultural heritage because it enables having copies outside the country (Pawłowska et al., 2023). However, originally digitization was not a preservation method, but a method for making records broadly

accessible, like microfilming. Because it was expensive, digitization served mainly academic research. Other potential uses have appeared later. (Zaagsma, 2023.).

## **ONE NEEDS ORIGINALS TO ENSURE PROPERTIES OF RECORDS**

According to ISO 15489 standard properties of records are authenticity, reliability, integrity, and usability (International Organization for Standardization, 2001, 7). The standard says that

“An authentic record is one that can be proven

- a) to be what it purports to be,
- b) to have been created or sent by the person purported to have created or sent it, and
- c) to have been created or sent at the time purported.” (International Organization for Standardization, 2001, p. 7.)

The emphasis above on the words “one that can be proven” is mine. If the originals are not there, how can we prove that they were authentic? And in addition, there is the challenge of proving authenticity of digital copies.

Professionally this is a moot point. The issue is complex and challenging both from a practical and theoretical point of view. Traditionally the records profession has guaranteed the authenticity of the records, and possibility that one may not be able to do it makes one uneasy. However, in practice, the question is less burning. Usually, there is trust, and one does not question the authenticity of archival records. Nevertheless, if authenticity of records is crucial and likely to be challenged, destroying originals after digitization is hardly a good solution.

## **ONE NEEDS ORIGINAL DOCUMENTS FOR LEGAL REASONS**

It is possible that a copy is not sufficient for legal purposes, and, thus, one needs the original document. Luciana Duranti notes (1989) that in diplomatics an “imitative copy”, like a photocopy

“...reproduces, completely or partially, not only the content but also the forms including the external ones (layout, script, special signs, medium and so on), of the original: a modern example is the photocopy. The probative or evidentiary value of an imitative copy is close to that of the original itself, but it does not confer on the copy legal validity in court [Emphasis mine].” (Duranti, 1989.)

Besides imitative copies, there are “authentic copies”, among others. In diplomatics, an authentic copy is certified by officials authorized to execute such a function, so as to render it legally admissible in evidence (Duranti, 1989). To me this suggests that a certified official could create a digital copy that is from a legal perspective as good as the original document. Nevertheless, I think that legal validity and question of authentic copies are issues that must be examined and solved locally in the framework of its legislation. To my understanding a digital copy is not inherently valueless in eyes of the law. What weight is given to a digital copy depends on the legal system.

Finnish conceptualization of this issue seems to be influenced by modelling in the field of libraries. Conceptual model of Functional Requirements for Bibliographic Requirements (FRBR) state that there are “works” of art (like book about Harry Potter) that may have various “expressions” and “manifestations” (e.g., paperback and hardback editions and a movie) that appear physically in “items” that may be found on the shelves of a library as books or DVDs (e.g., Carlyle, 2006). In similar vein, but without analyzing this in detail or taking similarities to FRBR to extreme, instructions of the National Archives of Finland suggest that a record may have an analog and a digital “expression.” Both expressions are considered equal. One may destroy analog expressions if “digitization process has not weakened evidentiality, integrity, and authenticity of records”, and the National Archives has decided that the analog expressions do not have “cultural-historical value” that would necessitate preserving them. (Kansallisarkisto, 2019, 2023.)

## DISCUSSION

In English-speaking literature, term “intrinsic value” has been used in conjunction with “cultural-historical value” to describe historical materials that should be retained in their original form rather than as copies. One can find the term in an American glossary in year 1974 (McRanor, 1996). In year 1979 the term gained particular importance when one began in the United States to consider possible large-scale replacement of paper records with miniaturized or other copies. Therefore, a Committee began to work to define the qualities and characteristics of records having intrinsic value. (NARA, 1982.)

Qualities and characteristics of records having intrinsic value reflect to my understanding the arguments that I discussed above. The Committee on the Intrinsic

Value noted that all records in their original physical form have qualities and characteristics that would not be preserved in copies. Records with intrinsic value have them to such “a significant degree” that the originals must be saved. Therefore, records in their original physical form is the only archivally acceptable form for preservation. (Emphasis mine, NARA, 1982.) Thus, the borderline between records with intrinsic value and those without intrinsic value is blurred and not definite.

The Committee suggested following criteria for intrinsic value (NARA, 1982):

- Physical form that may be the subject for study if the records provide meaningful documentation or significant examples of the form (e.g., a wax-cylinder sound recording)
- Aesthetic or artistic quality (e.g., watercolor sketches, photographs)
- Unique or curious physical features (an exemplary selection)
- Age that provides a quality of uniqueness (age can be a factor even with comparatively recent records, e.g. the earliest records concerning, for example, the development of the radio industry may have intrinsic value)
- Value for use in exhibits
- Questionable authenticity, date, author, or other characteristic that is significant and ascertainable by physical examination (records are of doubtful authenticity or have informational content that is open to question)
- General and substantial public interest because of direct association with famous or historically significant people, places, things, issues, or events
- Significance as documentation of the establishment or continuing legal basis of an agency or institution
- Significance as documentation of the formulation of policy at the highest executive levels when the policy has significance and broad effect throughout or beyond the agency or institution.

The criteria are mostly easy to understand. Interestingly they are filled with value-laden terms that make the criteria largely subjective. McRanor (1996) has noted that there is a contradiction between what seems outwardly like “objective” criteria and the ultimate de facto relativity of the intrinsic value. If “meaningful documentation” (1), “significant example” (1), “aesthetic or artistic quality” (2), “unique or curious physical features (3), “value of use in exhibits” (5), “general and substantial interest” (7), and “famous and historically significant

people” (7) are criteria the line between records having intrinsic value and others is drawn in water. This seems to be exactly the point of the committee: there are no records that are absolutely without value. The only question is to what extent they are valuable.

## CONCLUSIONS

Neither the criteria for intrinsic value, nor the arguments for preserving the originals that I have discussed suggest that one might not destroy the originals after digitization. However, this requires careful consideration.

## REFERENCES

- Brown, J. S., & Duguid, P. (2002). *The social life of information*. Harvard Business School Press.
- Carlyle, A. (2006). Understanding FRBR as a conceptual model. FRBR and the bibliographic universe. *Library Resources & Technical Services*, 50(4), 264–273.
- Conway, P. (2015). Digital transformations and the archival nature of surrogates. *Archival Science*, 15(1), 51–69. Retrieved at <https://doi.org/10.1007/s10502-014-9219-z> (accessed on 1. 8. 2025).
- Dever, M. (2010). Greta Garbo’s foot, or, sex, socks and letters. *Australian Feminist Studies*, 25(64), 163–173. Retrieved at <https://doi.org/10.1080/08164641003762461>(accessed on 25. 7. 2025).
- Dever, M. (2013). Provocations on the pleasures of archived paper. *Archives and Manuscripts*, 41(3), 173–182. Retrieved at <https://doi.org/10.1080/01576895.2013.841550> (accessed on 25. 7. 2025).
- DLM Forum Foundation. (2011). *MoReq2010. Modular requirements for records systems. Volume 1. Core services & plug-in modules. Version 1.0*.
- Duff, W. M. & Johnson, C. A. (2002). Accidentally found on purpose: Information-seeking behavior of historians in archives. *Library Quarterly*, 72(4), 472–496.
- Duranti, L. (1989). Diplomatics: New uses for an old science part I. *Archivaria*, 28, 7–27.
- Gertz, J. (2000). Selection for preservation in the digital age. *Library Resources & Technical Services*, 44(2), 97–104. Retrieved at <https://doi.org/10.5860/lrts.44n2.97> (accessed on 1. 8. 2025).

- Hirvonen, L. (2017). *Survey of digitization in archives*. The National Archives of Finland. Retrieved at [https://kansallisarkisto.fi/uploads/Viranomaisille/Mas-sadigitointi%20alasivujen%20liitteet/Liite\\_2\\_Digitization\\_Survey\\_2017.pdf](https://kansallisarkisto.fi/uploads/Viranomaisille/Mas-sadigitointi%20alasivujen%20liitteet/Liite_2_Digitization_Survey_2017.pdf) (accessed on 1. 8. 2025).
- International Organization for Standardization [ISO]. (2001). *ISO 15489-1. Information and documentation. Records management. Part 1: General*. ISO.
- Kansallisarkisto. (4. 12. 2019). *Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin KA/3357/07.01.02.04.02/2019 [National Archives' requirements for digitization aimed at disposal]*.
- Kansallisarkisto. (12. 5. 2023). *Julkishallinnon pysyvästi säilytettävien digitoitujen ja digitoitavien asiakirjojen säilytysmuoto ja analogisen säilytysmuodon kulttuurihistoriallisen arvon arviointitietojen pysyvä säilytys. Päätös KA/15927/28.02.01/2023 [Preservation format for permanently retained digitized and to-be-digitized documents in public administration, and permanent storage of cultural-historical value assessment data of analog preservation formats. Justification memorandum]*.
- Kumpulainen, S. W. & Kautonen, H. (2017). Accidentally successful searching: Users' perceptions of a digital library. *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, 257–260. Retrieved at <https://doi.org/10.1145/3020165.3022124> (accessed on 25. 7. 2025).
- Lu, Q., Zhu, Z., Li, Z., & Lian, Z. (2021). *Paper dating analysis based on paper texture image feature*. 13–17. Retrieved at <https://doi.org/10.1109/ICOIAS53694.2021.00010> (accessed on 25. 7. 2025).
- McRanor, S. (1996). A critical analysis of intrinsic value. *The American Archivist*, 59(4), 400–411.
- National Archives and Records Administration [NARA]. (1982). Intrinsic value in archival material. *A Modern Archives Reader*, 91–99.
- Ogilvie, B. (2016). Scientific archives in the age of digitization. *Isis*, 107(1), 77–85. Retrieved at <https://doi.org/10.1086/686075> (accessed on 25. 7. 2025)
- O'Toole, J. (1989). On the idea of permanence. *The American Archivist*, 52(1), 10–25. Retrieved at <https://doi.org/10.17723/aarc.52.1.3x85283576r43387> (accessed on 25. 7. 2025).

- Pashkeeva, N. (2024). Building ignorance by disseminating “evidence”: An agnotological look into the digital archives of the Islamic Republic of Iran. *Archival Science*, 24(3), 455–479. Retrieved at <https://doi.org/10.1007/s10502-024-09447-9> (accessed on 25. 7. 2025).
- Pawłowska, A., Pabich, M., Ivashko, O., Dmytrenko, A., Kozłowski, T., & Bigaj, P. (2023). Problems of preservation, protection and restoration of cultural heritage objects and museum collections in the conditions of war. *Muzeológia a Kultúrne Dedičstvo*, 11(2), 17–40. Retrieved at <https://doi.org/10.46284/mkd.2023.11.2.2> (accessed on 25. 7. 2025).
- Podolsky Nordland, L. (2004). The concept of secondary provenance: Re-interpreting Ac ko mok kis map as evolving text. *Archivaria*, 58(August), 147–159.
- Savinotko, S. (2025). Käsinkoskettava arkisto: Tiedontuotantoa arkistojen materiaalis-affektiivisissa sommitelmissa [The tangible archive: knowledge production in the material-affective assemblages of archives]. *Elore*, 32(1), Article 1. Retrieved at <https://doi.org/10.30666/elore.156232> (accessed on 25. 7. 2025).
- Schwartz, J. & Cook, T. (2002). Archives, records, and power: The making of modern memory. *Archival Science*, 2(1), 1–19.
- Sulaiman, A., Omar, K., & Nasrudin, M. F. (2019). Degraded historical document binarization: A review on issues, challenges, techniques, and future directions. *Journal of Imaging*, 5(4), 48. Retrieved at <https://doi.org/10.3390/jimaging5040048> (accessed on 25. 7. 2025).
- Yeo, G. (2007). Concepts of record (1): Evidence, information, and persistent representations. *The American Archivist*, 70(Fall/Winter), 315–345.
- Zaagsma, G. (2023). Digital history and the politics of digitization. *Digital Scholarship in the Humanities*, 38(2), 830–851. Retrieved at <https://doi.org/10.1093/llc/fqac050> (accessed on 25. 7. 2025).
- Zhang, Z., Ewert, U., Barrett, T., & Bond, L. (2019). *Paper watermark imaging using electron and low energy x-ray radiography*. Retrieved at <https://dr.lib.iastate.edu/handle/20.500.12876/1925> (accessed on 25. 7. 2025).

## SUMMARY

The paper explores whether original paper documents should be preserved after digitization. It argues that while digitization offers accessibility, efficiency, and cost savings, it also results in the loss of several important aspects of original documents:

### Key Points:

#### Loss of Materiality and Information

Digitization removes physical features like texture, watermarks, and even smells, which can carry historical or research value.

#### Loss of Context and Provenance

Originals are part of a physical archival structure that conveys relationships and context. Digitized items often become isolated unless carefully curated with metadata.

#### Emotional and Experiential Value

Physical documents evoke emotional responses and a sense of direct connection to history, which digital versions cannot replicate.

#### Technological Limitations and Future Potential

Current digitization may not capture all details. Originals allow for future re-digitization with better technology.

#### Authenticity and Legal Validity

Proving authenticity is harder without originals. Legal systems may require originals or certified copies.

#### Intrinsic Value

Some documents have unique qualities (e.g., age, artistic value) that make their original form irreplaceable.

#### Conclusion

While digitization is beneficial, it cannot fully replace originals in terms of materiality, context, emotional impact, and evidential value. Decisions to discard originals should be made carefully, considering archival principles, legal frameworks, and cultural significance.

*Typology: 1.03 Short Scientific Article*

