Article

Michael Friedewald*, Iván Székely and Murat Karaboga

# Preserving the Past, Enabling the Future: Assessing the European Policy on Access to Archives in the Digital Age

https://doi.org/10.1515/pdtc-2024-0003
Received January 11, 2024; accepted March 20, 2024;
published online April 4, 2024

**Abstract:** In this study we investigate the implementation and challenges of accessing archives in the digital age. Through an online survey and expert interviews, data were collected from archival institutions, professional users, and civil society organizations in all Council of Europe member states. The findings reveal that digitization has had a positive impact on document accessibility, providing convenience, reducing workload, and improving user access. However, only a small portion of archival holdings have been digitized, highlighting the need for further efforts in this area. Data protection and copyright present significant obstacles, with the European General Data Protection Regulation leading to access restrictions and legal uncertainties. Preservation, interoperability, and organization of records pose challenges in the digital realm, and while artificial intelligence shows promise in automating metadata creation and identifying sensitive materials, concerns about algorithmic biases and the need for additional checks are highlighted. To enhance accessibility, user-friendly interfaces, cooperation among memory preserving institution, and outreach to new user groups are crucial, which can ensure the effective preservation and accessibility of digital collections in the digital age. We provide insights for policy-making and archival institutions in navigating the evolving landscape of accessibility.

**Keywords:** accessibility; archives; digitalisation; freedom of information; policy

***Corresponding author: Michael Friedewald**, Fraunhofer-Institut für System- und Innovationsforschung ISI, Breslauer Straße 48, 76139 Karlsruhe, Germany, E-mail: michael.friedewald@isi.fraunhofer.de. https://orcid.org/0000-0001-8295-9634
**Iván Székely,** Vera and Donald Blinken Open Society Archives, Central European University – Budapest Campus, Arany Janos u. 32, 1051 Budapest, Hungary
**Murat Karaboga,** Fraunhofer-Institut für System- und Innovationsforschung ISI, Breslauer Straße 48, 76139 Karlsruhe, Germany

## 1 Introduction

Following the radical changes in the political systems of Eastern Europe, there was a significant discussion in the early 1990s regarding the function and philosophy of archives (Körmendy 2007). The collapse of the authoritarian socialist regimes in East-Central Europe led to a demand for making previously inaccessible state documents available to the public, which aimed to lay the foundation for reckoning with the history of the Cold War era at both institutional and individual levels (Gruodytė and Gervienė 2015). However, the political system changes did not automatically result in the modernization of the legal system or the establishment of democratic rule-of-law societies. The guarantee of human rights, including legislative changes and judicial enforcement, remained a lengthy and complex process, still unfinished in certain countries today. Nonetheless, access to information and archives became a symbol of moving away from the dictatorial past during the significant political transformations of the early 1990s (Majtényi et al. 2005; Tyacke, Van Den Boeck, and Steendam 1995).

Recognizing the importance of access to archives, the Council of Europe (CoE) took a significant step in 2000 by adopting Recommendation No. R (2000) 13 on a European Policy on Access to Archives (Council of Europe 2000). This recommendation established the first international standard on a European policy for access to archives and, while not legally binding, it expressed common principles that all member states were expected to respect. It emphasized that access to information and archives should be considered a right rather than a privilege, aligning with the principles of the rule of law, the demands of historians, and the aspirations of civil society. In 2003/04, a pan-European survey and analysis were conducted to assess the implementation of the CoE recommendation, with the evaluation revealing significant shortcomings in meeting the recommended principles, particularly in new democracies in East-Central Europe, as well as that established Western democracies had traditions favoring archive secrecy or restricted user access in various ways (Kecskeméti and Székely 2005).

Over the past 20 years, the archival world has undergone significant changes, primarily driven by new and emerging technologies. The digitization of paper-based and audio-visual documents has become a widespread practice in archival institutions that have the necessary technical and financial resources, while the uploading of digitized documents to openly accessible platforms has further enhanced accessibility. Finally, the acceptance of born-digital documents, including formats different from traditional documents such as databases or ephemeral documents, has become essential. The impact of these technological advancements on the accessibility of archives cannot be overstated; digitization and online access have revolutionized the way archival materials are stored, preserved, and made available to users. It has opened up new possibilities for researchers, historians, and the general public to explore and engage with historical records (Hofman 2012; Székely 2015).

Since the 2004 pan-European survey, it was difficult to assess comprehensively how the access situation in Europe has changed and the role of new and emerging technologies in these changes. The data situation regarding the digitization of European archives is rather unsatisfactory; apart from regular statements by archives themselves about their progress in digitizing collections and processes, there have only been a few systematic and transnational studies. Following EU funded research such the ENUMERATE project,[1] the European Commission (2019) published a progress report on the digitization and online accessibility of cultural material and digital preservation in cultural heritage organizations a few years ago, which, however, does not specifically address the conditions of archives.

In light of these developments, the Council of Europe has commissioned a new study to answer, *inter alia,* the following two research questions:

(1) What progress have the Council of Europe member states made in implementing Recommendation R (2000) 13, and to what extent do the digitization of conventional holdings and the addition of digital documents in particular contribute to accessibility?
(2) What are current and future challenges for accessibility of archives, specifically considering the status of digitization efforts, online access, and the technical and curatorial challenges that may affect accessibility?

By responding to these questions our study does not only broaden the knowledge of researchers and archivists, but also that of CoE cultural policymakers, who can use it as a basis for complementing the instruments that promote archival accessibility in the digital world (Friedewald, Székely, and Karaboga 2024).

The following section explains the approach taken in the study, followed by two sections that present the empirical results on digitization of archives and future challenges respectively. Finally, we draw conclusions for archives and policymakers.

## 2 Approach and Methodology

For the study, we conducted an online survey among archives in the CoE member states and their users. The design of the survey was based on the 2003/04 survey on the same topic to enable result comparison.

Representatives from various stakeholder groups, including national archives, regional and municipal archives, academics (historians in particular), and civil society organizations working for the protection of civil rights or freedom of information, were asked to fill out the questionnaire[2] for the survey. For the topic of this article, however, we only use the survey among the archives.

The questionnaire contained a total of 57 questions on 14 topics relating to accessibility of archives, including nine questions on digitisation, digital documents, and remote access. An invitation was sent to the 46 National Archives of all CoE member states and selected regional and municipal archives (see Table 1).[3]

The survey was conducted online using the professional survey tool EFS Survey by Tivian.[4] The fieldwork for the survey took place from October 10 to November 4, 2022 and, while the response rate among the National Archives was reasonably high (85 %), the sample size remained relatively small due to the limited population (see Table 1). Additionally, 23 local and regional archives from 14 countries (61 %) completed the questionnaire, with this article presenting only descriptive statistics of the survey.

Following a typical multi-method approach, the survey was complemented by six expert interviews, conducted in January and February 2023 with representatives of a regional and a national archive, academic experts from archivistics

1 https://cordis.europa.eu/project/id/270939/en.

2 The questionnaire is available online at https://doi.org/10.5281/zenodo.7676589.

3 While the aim was to reach as many national archives as possible, the survey of regional and municipal archives primarily served to highlight differences between Archives at different administrative levels.

4 https://www.tivian.com/.

**Table 1:** Response rates.

| | Contacted | | Responses | | Response rate (archives) |
|---|---|---|---|---|---|
| | Archives | Countries | Archives | Countries | |
| National archives | 46 | 46 | *N* = 39 | 39 | 39/46 = 85 % |
| Regional and local archives | 38 | 17 | *N* = 23 | 14 | 23/28 = 61 % |

and digital history, a civil rights lawyer, and a data protection commissioner. The aim of these interviews was to validate the quantitative survey results as well as identify and assess future challenges.

# 3 Findings on the Current Status of Digitization and Online Access to Archives

According to our survey, the main technical challenge for archives today is the digitization of information and the digitalisation of internal processes. With computer technology, the amount of information has multiplied,[5] which includes born-digital documents as well as traditional documents in digital form. Some digital documents mimic analogue ones while others are completely new for archives, like databases or social media messages.

Computer technology and networks offer powerful tools for storing, organizing, and retrieving information, making it accessible to users. However, it was not until the spread of the Internet in the early 2000s that this became practically relevant for archives, libraries, and other cultural heritage institutions, with these developments then starting to gain momentum after the CoE Recommendation was adopted in 2000.

In the digital age, accessibility of archives encompasses three primary dimensions: the availability of digital and digitized documents, the possibility to remotely find and access them, and the opportunity for users to receive online support.[6]

## 3.1 Digital Documents

According to the survey results on the current state of digitization, the majority of archives have a positive view of its impact on document accessibility. The primary reason for this positive assessment, cited by 73 % of archives, is the ability to access documents through the Internet. Digitized documents were also praised for their convenience, with 66 % of archives noting the advantage of eliminating the need for physical retrieval and transfer; furthermore, 61 % of respondents reported that digitization helps reduce the workload of archive staff. However, 55 % of archives agree that accessibility has only seen partial improvement thus far, as only a small portion of their holdings have been digitized or made remotely accessible.

Among the archives surveyed (42 in total, including 32 national and 10 regional/local archives), the proportion of born-digital text documents in their collections remains relatively low. Specifically, 86 % of archives reported that less than 5 % of their holdings consist of born-digital text documents, with an additional 9 % of archives reporting that the share of born-digital text documents ranges between 5 and 15 %. However, one national and one municipal archive reported that more than 50 % of their holdings are already born-digital.

Archives have also made significant efforts to digitize their analogue text documents (see Table 2). Among the responding archives, 56 % (18 national and five regional/local) have digitized less than 5 % of their analogue text documents, while another 34 % (10 national and four regional/local) have digitized between 5 and 15 % of their text documents, with a single national archive (2 %) achieving a digitization rate of between 25 and 50 %. Additionally, two national archives and one municipal archive reported that they have already digitized more than 50 % of their text documents.

Born-digital AV (audio, video and film) documents constitute a relatively small portion of the archives' collections, but their overall share is slightly larger than that of text documents. Out of the 40 responding archives (30 national, 10 regional and local), 27 (68 %) have less than 5 %

**5** According to official statistics, the volume of data created, captured, copied, and processed worldwide has been growing exponentially from two zettabytes in 2010 to 64 zettabytes in 2020; https://www.statista.com/statistics/871513/worldwide-data-created/.

**6** Among progressive contemporary archives, a good example is the Blinken OSA Archivum (https://www.osaarchivum.org/), which offers all three dimensions to its users.

**Table 2:** Digitization of text and AV documents. Question: What percentage of your (analogue) text and AV holdings have been digitized?

| | | <5 % | 5–15 % | 15–25 % | 25–50 % | >50 % |
|---|---|---|---|---|---|---|
| *Text* | National archives (*n* = 31) | 18 (58 %) | 10 (32 %) | | 1 (3 %) | 2 (6 %) |
| | Regional and local archives (*n* = 10) | 5 (50 %) | 4 (40 %) | | | 1 (10 %) |
| *AV* | National archives (*n* = 29) | 20 (69 %) | 5 (17 %) | 1 (3 %) | 2 (7 %) | 1 (3 %) |
| | Regional and local archives (*n* = 10) | 5 (30 %) | 3 (30 %) | | 2 (20 %) | |

Source: Questionnaire for Archives, Question 13.2, variables 340 and 341.

of their AV documents in born-digital format. In eight other archives (23 %), the share of born-digital AV documents is between 5 and 15 %, while two national Archives reported that more than 50 % of their AV documents were digitally created.

Regarding the digitization of analogue AV holdings, the situation differs from that of text documents (see Table 2). Among the responding archives, 64 % have digitized up to 5 % of their analogue AV documents, while another 21 % have digitized between 5 and 15 % of their AV documents. One national archive has digitized between 15 and 25 % of their AV documents, with another archive falling within the same range, while two more archives have digitized between 25 and 50 % of their AV documents. The leading national archive has already digitized more than 50 % of their AV documents, while the two most advanced municipal archives have digitized between 25 and 50 % of their AV documents.

According to Figure 1, 81 % of archives prioritize the protection of vulnerable documents for digitization, followed by 79 % wanting to make documents more accessible to a wider range of users, often in response to user demands. Efficiency gains in archival management were mentioned by only 31 % of respondents, with regular preservation and migration of digital documents of primary importance to 26 % of archives. Only 31 % mentioned efficiency gains in archival management, while 26 % digitize documents in the course of regular preservation and migration activities.

The possibility to provide digital documents over the Internet has posed new challenges for archives and other cultural heritage institutions, including user involvement in tasks like indexing archival material (Benoit, III and Eveleigh 2019; Ridge 2014). This "participatory archives" approach allows user contributions by enriching archival holdings and fostering a collaborative community. The Blinken OSA Archivum for instance has implemented two interactive projects: the first, Yellow Star Houses,[7] commemorates a forgotten chapter of the Holocaust in Budapest, offering an interactive website with a map and related documents, where users can contribute photos and personal memories; the second project, Parallel Archive, enables researchers to upload and share archival documents, facilitating collaboration and context enhancement.

The archives were asked if they involve users in describing and/or commenting on archival documents through crowdsourcing, with such user involvement practiced in 23 % of the National Archives and 33 % of the regional and local Archives. This includes tasks like transcribing and indexing civil records, handwritten documents, and providing content descriptions for photographs. Archives may rely on volunteers, including amateur genealogists, or collaborate with platforms like Wikimedia.

Regarding the risks associated with increased online accessibility of documents, data protection and copyright violations were the most commonly mentioned risks, with both mentioned by 40 % of participants. Risks related to national security and document forgery or modification were considered less relevant, with only 12 % of participants mentioning each of these risks.

## 3.2 Remote Access

In addition to digital documents, remote access to finding aids and documents through the Internet is the second important component of a digital archive. Among the surveyed archives, 93 % provide online access to catalogues and finding aids but, although this is a high percentage, the extent of digital cross-references in the catalogues should also be taken into account (Figure 2). In national archives, only 11 % of online catalogues provide direct links to individual digital documents, while 59 % offer only descriptions of the documents. In contrast, larger regional and local archives tend to have a higher percentage of catalogues linking to digital documents, however, this may not be representative of small and medium-sized archives in regions and cities. Finally, we also asked whether the archives offer artificial intelligence-based applications for finding documents, with as many as 20 % of the national Archives having such applications in use, with none of the regional or local archives surveyed responding positively to this question.

7 http://www.yellowstarhouses.org/.

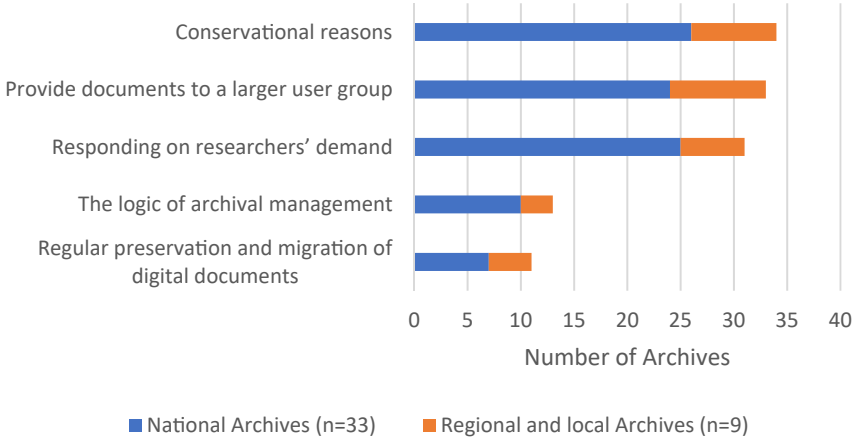


**Figure 1:** Priority setting for digitisation. Question: When defining priorities in digitizing archival documents, which is of primary importance in your institution? Source: Questionnaire for Archives, Question 13.3, variables 376–80.

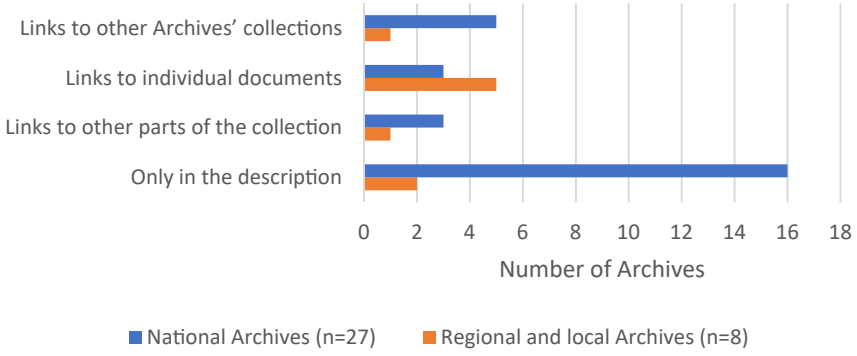


**Figure 2:** Cross-references in online catalogues. Question: If your catalogues/finding aids are available online, are there digital cross-references? Source: Questionnaire for Archives, Question 14.1.1, variable 351–54.

Regarding online access to digital documents, the majority of archives (88 %) offer this service for free. However, there are restrictions in place. In nearly half of national archives (48 %) and the majority of regional and local archives (90 %), not all digital documents are accessible online; instead, only certain collections/series or selected samples are available for online access (Figure 3).

### 3.3 Online Communication

Communication channels with archive staff are vital for accessibility. As digitization and remote access increase, new digital communication channels have emerged to facilitate user requests and support, which include professional consultation, reading room reservation, and on-demand digitization requests (see Figure 4). Nearly all archives, both national and regional/local, provide support through their institutional website (90 % of national archives/100 % of regional and local archives), direct email with the reference archivist (79 %/90 %), and social networks (76 %/80 %). However, real-time chat with the reference archivist remains less common, available in only 15 % of the responding national archives.

## 4 Challenges for the Future Accessibility of Archives

In this section, we go beyond the description of the current status and analyze the future challenges and opportunities for the accessibility of archives in the digital age on the basis of the literature and, above all, our interviews with experts.

Each step of the archiving process, as depicted in Figure 5 below, is faced with new challenges due to the use of digital technologies. These steps are crucial for archival institutions and other memory-preserving organizations, as they involve collecting, preserving, and providing access to documents for research and other purposes.

Experts for digital archivistics such as Jaillant (2022) or Andreas Fickers (interview) point out that discussions on born-digital and digitized archives have been primarily led by scholars from archival studies, with limited input from other disciplines, particularly the users of archives. As a

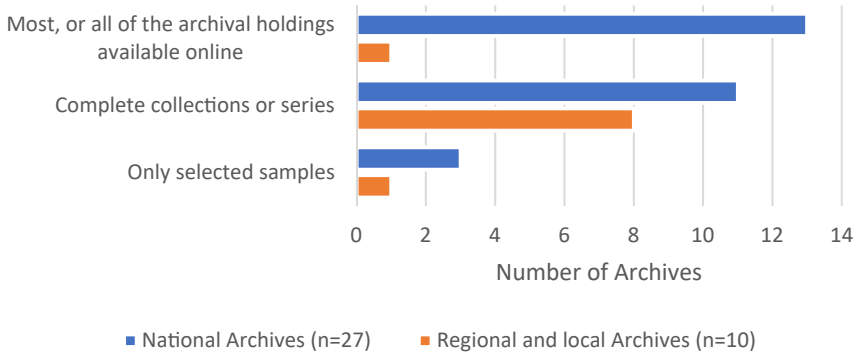


**Figure 3:** Free online access to digital documents. Question: If you provide free access to digital/digitized documents online, which parts can be accessed free of charge? Source: Questionnaire for Archives, Question 14.2.1, variable 357.

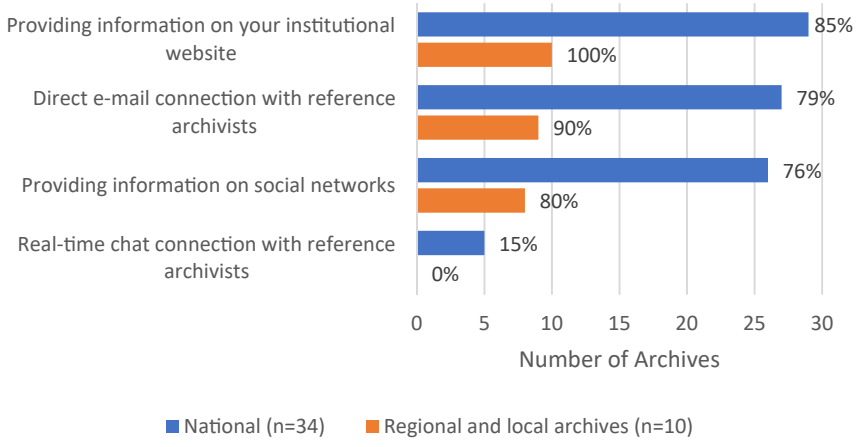


**Figure 4:** Possibilities for online communication with archives. Question: What kind of online channels does your institution use for communicating with archival users? Source: Questionnaire for Archives, Question 14.3, variable 363–66.

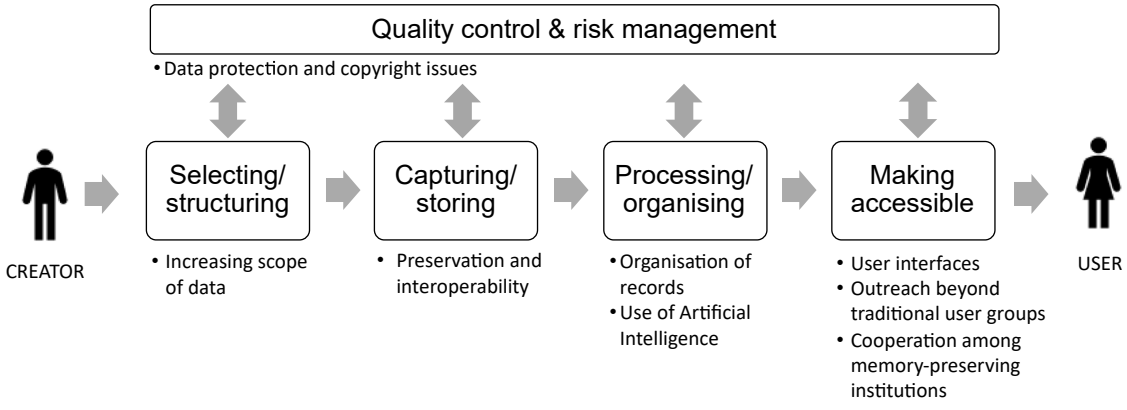


**Figure 5:** Challenges for accessibility along the archival process chain.

result, the focus has been more on capturing and preserving documents than prioritizing access to digital archives.

## 4.1 Increasing Scope of Data

Already at the stage of appraisal and selection, archives face a dilemma: should they preserve everything due to cheaper storage space and the potential future interest of even seemingly irrelevant documents and data, or should they continue to assess the current and future value of all available data and preserve only the valuable ones? Some argue that with the growth of the global information infrastructure, all information can and will be stored forever and accessible anytime, anywhere (Bell and Gemmell 2009).

If you follow this though the question arises as to whether society really needs specialized institutions like archives and other memory institutions that collect, preserve, and make information accessible. In response to this, Székely (2017, 11–14) identifies six key arguments supporting the continued relevance of archival institutions in the digital age:

(1) Institutional traditions and their embeddedness in the cultural fabric of society;
(2) Ensuring the long-term functionality of documents and data;
(3) The ability to preserve physical, non-digital copies;
(4) The importance of preserving the historical and information technology context;
(5) The ongoing task of migrating document formats; and finally,
(6) The important role of archival institutions in taking institutional responsibility.

Experts generally do not consider "total recall" as an effective approach. Simply storing documents and information without proper processing, cataloguing, and metadata enrichment does not guarantee easy accessibility, an issue which has been challenging for some time, even in the pre-digital era, resulting in the development of procedures for targeted selection or deletion (Mayer-Schönberger 2011).

Contemporary archives typically collect digital data and documents from the recent past, making it challenging to place them in a historical context. It is estimated that up to 20 % of digital documents will be archived, compared to 5 % in analogue contexts (Convery 2014, 159; Moss and Gollins 2017, 7). However, there is a concern that the size of digital collections may overwhelm smaller archives, as they may lack the resources to properly catalogue and index them.

The volume of data raises the question of prioritizing materials based on their importance and value. For example, the Federal Archives in Germany prioritize digitizing holdings from the National Socialist era due to high user demand (Hänger, interview). Similarly, in the Netherlands, there is a focus on digitizing holdings related to colonial history (Jeurgens, interview). While prioritization is necessary to ensure completeness for important topics in the digital collection, there is a risk of marginalizing less prioritized topics, leading to digitally inaccessible related documents.

## 4.2 Data Protection and Copyright Issues

The interviewed experts emphasized that the European General Data Protection Regulation (GDPR)[8] is a primary challenge and threat to archive accessibility to date. This perception is influenced by the practical considerations and the stance of the competent data protection authorities. In response to GDPR, archives have often chosen to restrict access to potentially sensitive documents entirely, fearing legal violations and penalties, despite the previous successful handling of sensitive data by researchers using the archives (Sipos, interview).[9] Archives feel uncertainty due to varying interpretations of GDPR requirements across the EU, find existing guidelines (EAG 2018) too abstract for their daily practice, and struggle to reconcile the tensions between data protection laws, freedom of information laws, and archive legislation.

Data protection concerns have led to the closure of entire digital collections, as highlighted by Jaillant (2022, 420f.) in an interview study. For example, the email correspondence of novelist Ian McEwan was transferred to a private Archive in the US in 2014 and is not even listed in finding aids. Even researchers who do have access to such closed collections face difficulties in publishing their research due to the time-consuming process of obtaining consent or rights to use the information from individuals involved. As a result, scholars tend to focus on more accessible document collections when choosing their research topics, a problem which affects not only born-digital documents but also digital copies of paper documents that were previously accessible in reading rooms.

Even if documents are (digitally) sanitized, combining them may still lead to the identification of individuals, violating data protection rules. Artificial intelligence tools can also uncover hidden patterns in data, potentially exposing sensitive personal or business information.

In addition to data protection, archives hesitate to make digital collections accessible for reasons beyond data protection. Copyright issues, especially when authors cannot be located, pose significant challenges;[10] even when authors are known, acquiring usage rights can be time-consuming, causing archives to refrain from providing access to relevant documents (Hänger, interview).

Finally, accessibility may be restricted not only due to data protection and copyright concerns, but also because there may be "records that are culturally sensitive, or have offensive language in it, or are simply unpleasant to look at" (Jaillant 2022, 425).

8 https://eur-lex.europa.eu/eli/reg/2016/679/oj.

9 Sipos mentioned the case of Hungarian swim trainer László Kiss in 2016; https://www.reuters.com/article/csports-us-hungary-swimming-coach-rape-idCAKCN0X41PO (accessed December 14, 2023).

10 It is important to emphasize that access for research and educational purposes only may fall into the category of fair use and this possibility can be offered in both physical and virtual (online) research rooms for registered users.

## 4.3 Preservation and Interoperability

More than 25 years ago, when the discussion about digital archives started to spread, experts expressed concern about the possibility of significant data loss. They warned about a potential "digital dark age" due to the deterioration of storage media like magnetic tapes and floppy disks, as well as the aging of computer hardware (Hedstrom 1997; Kuny 1997). Additionally, the obsolescence of software and data formats often posed a significant barrier to accessing born-digital documents (Jaillant 2022, 423).

The current challenge is no longer focused on creating computer-readable documents in preservation-friendly formats like PDF/A. Instead, it lies in capturing and providing content and data from outdated systems, including obsolete email systems. Additionally, there is the question of determining the "original" version of a document, like a database, and whether the authenticity relies on the original IT environment and its technical properties (Friedewald and Leimbach 2011).

Effective and efficient data capture relies on personnel with the necessary skills, however, there is a shortage of appropriate IT knowledge in this area (Fresa, Justrell, and Prandoni 2015, 191; Kim 2018; Hänger, interview). Furthermore, ensuring the traceability of authenticity for born-digital objects poses an additional obstacle (Poole 2015, 116).

## 4.4 Organization of Records

Organizing digital documents also poses challenges for archives. The traditional concept of *respect des fonds*, where archivists passively preserve documents, is challenged in the digital realm, as digital collections often lack structure and proper metadata, requiring a more active curatorial approach for effective collection management (Jaillant 2022, 427f.).

Preservation and accessibility requirements are often in conflict with each other, as storing documents in non-future-proof formats necessitates repeated migration to ensure future accessibility, while obsolete systems and fragile data carriers, especially in the audio-visual field, pose additional challenges. The use of coding-decoding algorithms (codecs) can also impact the preservation of artistic qualities in video recordings; to address these challenges, an international community, known as "No Time To Wait!", is developing new open standards and technical tools.[11]

Standardized metadata schemes and adherence to international standards are crucial for effective access to digital documents across Archives. Open standards like the Open Archival Information System (OAIS) Reference Model[12] and the Metadata Encoding and Transmission Standard (METS)[13] are recommended. Elsewhere, the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)[14] enables the collection of metadata descriptions from various archives, enabling the creation of services that utilize metadata from multiple sources. Metadata repositories act as aggregated information sources, allowing users to search within the repository and redirecting them to the original archive for relevant results. Aggregator portals like Archives Portal Europe[15] and European[16] provide integrated search capabilities across cultural institutions and collections, following similar principles.

In theory, the significant time gap between the creation of records and their opening up could encourage archivists to make access to collections as user-friendly as possible, while having ample time to process and organize records should also allow for designing strategies to make non-sensitive parts of the collections available sooner. However, without a clear understanding of how a future digital archive should be structured and the potential new tasks it may have, the long periods of closure can result in a hesitant "wait-and-see" attitude among archivists (Jaillant 2022, 429f.).

Within this context, experts highlight the notable differences in internal procedures followed by Archives for preservation and accessibility purposes (Hänger and Fickers, interviews). Preservation of documents can often be automated through mass processing techniques (scanning, text recognition), but ensuring accessibility requires time-consuming and challenging manual indexing and cataloguing.

New concepts of descripting standards have been developed, such as the "Records in Context" (ICA 2021), which includes a conceptual model and an ontology, and aims at enabling archival description to better capture the complex relationships archival documents have with each other and their creators, holders, and subjects. Linked Open Data is another concept of the Semantic Web, offering a common space for sharing data from repositories, including archives of a certain kind (Bizer, Vidal, and Skaf-Molli 2018). These concepts and the resulting standards may modernize existing descripting standards in the future, as well as inter-organizational cooperation,

11 https://mediaarea.net/NoTimeToWait.

12 http://www.oais.info/.

13 https://www.loc.gov/standards/mets/.

14 https://www.openarchives.org/pmh/.

15 https://www.archivesportaleurope.net/.

16 https://www.europeana.eu/.

resulting in semantic interoperability and better accessibility of data and document repositories.

## 4.5 Automation and the Use of Emerging Technologies

As in other fields, there are high expectations for what artificial intelligence and machine learning (AI/ML) can achieve in the archival context. This was illustrated by Colavizza et al. (2022), who reviewed six years of English-language literature in the intersection of AI and archives, and analyzed 53 relevant publications using the Records Continuum model. Elsewhere, the journal *AI and Society* devoted a thematic section on the topic, "AI use in Archives" (e.g., Stapleton and Jaillant 2022). In this literature as well as in our expert interviews, two aspects in particular were mentioned where AI can improve the accessibility of (digital) archives:

1) AI/ML can automate the creation of metadata, particularly for large unstructured collections, which involves capturing and organizing the content of documents to provide a rough classification (Colavizza et al. 2022). However, it is important to consider algorithmic bias, as AI/ML learns biases inherent in the data. For instance, if documents from the colonial period are classified, there is a risk of reproducing European colonial terminology and racist stereotypes (Luthra et al. 2023; Jeurgens, interview).
2) AI/ML algorithms can automatically detect sensitive materials and differentiate between problematic and unproblematic content. They can also aid in indexing and disclosing digital documents, as well as identifying images like faces or buildings (Decker 2022). However, given the probabilistic nature of detection, there is a possibility that problematic documents may not be identified. Consequently, additional checks are required to prevent or minimize risks before making documents public.

In addition to these application areas, there will be further applications in the field of automatic person-, object-, and location recognition, which will help researchers to search and interactively enrich large archival collections. New types of sorting, tagging, extraction, translation, text-to-speech, and speech-to-text conversion of archival documents according to criteria defined by researchers will bring new possibilities to the work of users of archives.

## 4.6 User Interfaces

Creating a suitable interface for digital collections is vital for document accessibility. This includes addressing archival requirements to safeguard against modification or deletion, while it is also important to conceal sensitive data based on the user's location and authorization, whether in a reading room or online.

In addition to viewing the document itself, users should have access to its broader or narrower context if available, which facilitates "scalable reading" which is especially useful for professional users. It enables them to seamlessly switch (or zoom) from distance to close reading (Fickers, interview; Fickers, Tatarinov, and Van Der Heijden 2022).

## 4.7 Cooperation of Memory Preserving Institutions and Outreach Beyond Traditional User Groups

Our survey primarily focused on traditional public archival institutions, particularly national archives. In these archives, it is important (though expensive and time-consuming) to ensure interoperability at a higher level; the survey results indicate a growing preference for online access among archive users, which necessitates the presence and online availability of digitized or born-digital documents. If researchers only find a portion of the documents they seek, they may expect the first archival institution to redirect them to another institution with supplementary material on the research topic, redirection which should be specific to the direct continuation of the document series in question. Developing such interoperable systems and standards, and making entire digital collections available in this manner, requires collaboration among institutions, however, these institutions often compete for resources and professional prestige, even if the benefits of these developments are realized elsewhere or contribute to improving user access in general. In the era of new types of archives, such as community archives or post-custodial archives, this collaboration becomes even more necessary.

New services that enhance accessibility to archival collections are attracting more users, but the demand is not primarily from traditional user groups. Instead, it comes from individuals with limited prior experience with archives who have new expectations. It is crucial to create appropriate accessibility methods for these users, which may differ from those for professional users, and experts

hope that archivists, librarians, and museum curators will explore ways to generate increased interest in their collections and specific topics in the future (Sipos, interview). While exhibitions of original artifacts like handwritten manuscripts have been successful in capturing interest, digital documents lack the same materiality and aura. Memory institutions must develop new approaches to provide broader access to their collections for the public.

## 5 Conclusions

New and emerging technologies present both challenges and opportunities for archives. Our pan-European study revealed that public archives have only digitized a small portion of their analogue records, and the number of born-digital records remains low in most European public archives. The management of digital records and non-traditional documents, including databases, websites, social media messages, and ephemeral documents, presents archives not only with theoretical problems but also concrete practical, technical, and financial challenges.

In the archival paradigms of the present and the future, providing access has become a primary objective. However, fast and convenient access is leading to increased user demands, and meeting these demands presents archives with new challenges. A new generation of users may emerge, some of whom may not even be able to read handwritten documents (Roberto 2017) – even if they are digitised, and for whom conversion facilities and special interfaces need to be provided. The use of AI in archives holds promise for various tasks such as document identification, document retrieval, identification of places and people, automatic metadata assignment, and the selection of documents containing sensitive data, but it cannot and should not completely replace human control.

To adequately address these challenges and leverage the benefits of emerging technologies, archives need to clarify their future goals, tasks, and visions. This will enable them to make necessary developments and provide the expertise required to navigate the evolving landscape of archival practices in the digital age.

## References


Bell, C. Gordon, and Jim Gemmell. 2009. *Total Recall: How the E-Memory Revolution Will Change Everything*. New York: Dutton.

Benoit, III, Edward, and Alexandra Eveleigh. 2019. *Participatory Archives*. London: Facet Publishing.

Bizer, Christian, Maria-Esther Vidal, and Hala Skaf-Molli. 2018. "Linked Open Data." In *Encyclopedia of Database Systems*, edited by Ling Liu, and M. Tamer Özsu, 2096–101. New York: Springer.

Colavizza, Giovanni, Tobias Blanke, Charles Jeurgens, and Julia Noordegraaf. 2022. "Archives and AI: An Overview of Current Debates and Future Perspectives." *Journal on Computing and Cultural Heritage* 15 (1): 1–15.

Convery, Nicole. 2014. "From Reactive to Proactive Appraisal." *Archives and Manuscripts* 42 (2): 158–60.

Council of Europe. 2000. *Recommendation No. R (2000) 13 of the Committee of Ministers to Member States on a European Policy on Access to Archives*. Strasbourg: Council of Europe. https://rm.coe.int/16804cea4f.

Decker, Stephanie. 2022. "Finding Light in Dark Archives: Using AI to Connect Context and Content in Email." *AI & Society* 37: 859–72.

EAG. 2018. "Guidance on Data Protection for Archive Services. EAG Guidelines on the Implementation of the General Data Protection Regulation in the Archive Sector: European Archives Group. https://commission.europa.eu/system/files/2018-10/eag_draft_guidelines_1_11_0.pdf (accessed March 20, 2024).

European Commission. 2019. "Cultural Heritage: Digitisation, Online Accessibility and Digital Preservation: Consolidated Progress Report on the Implementation of Commission Recommendation (2011/711/EU) 2015-2017." Working Document. Brussels.

Fickers, Andreas, Juliane Tatarinov, and Tim Van Der Heijden. 2022. "Digital History and Hermeneutics – Between Theory and Practice: An Introduction." In *Digital History and Hermeneutics*, edited by Andreas Fickers, and Juliane Tatarinov, 1–20. Berlin: De Gruyter.

Fresa, Antonella, Börje Justrell, and Claudio Prandoni. 2015. "Digital Curation and Quality Standards for Memory Institutions: PREFORMA Research Project." *Archival Science* 15 (2): 191–216.

Friedewald, Michael, and Timo Leimbach. 2011. "Computersoftware als digitales Erbe: Probleme aus Sicht der Technikgeschichte." In *Neues Erbe. Aspekte, Perspektiven Und Konsequenzen Der Digitalen Überlieferung*, edited by Caroline Y. Robertson-von Trotha, and Robert Hauser, 201–19. Karlsruhe: KIT Scientific Publishing.

Friedewald, Michael, Iván Székely, and Murat Karaboga. 2024. *Access to Archives in the Digital Age: Implementation of the Committee of Ministers of the Council of Europe Recommendation No. R(2000)13 on a European Policy on Access to Archives*. Strasbourg: Council of Europe Publishing.

Gruodytė, Edita, and Silvija Gervienė. 2015. "Access to Archives in Post-Communist Countries: The Victim's Perspective." *Baltic Journal of European Studies* 5 (2): 147–70.

Hedstrom, Margaret. 1997. "Digital Preservation: A Time Bomb for Digital Libraries." *Computers and the Humanities* 31 (3): 189–202.

Hofman, Hans. 2012. "Rethinking the Archival Function in the Digital Era." *Comma* 2012 (2): 25–34.

ICA. 2021. *Records in Contexts, Conceptual Model, Consultation Draft v0.2*. Paris: International Council on Archives. https://www.ica.org/sites/default/files/ric-cm-02_july2021_0.pdf (accessed March 20, 2024).

Jaillant, Lise. 2022. "How Can We Make Born-Digital and Digitised Archives More Accessible? Identifying Obstacles and Solutions." *Archival Science* 22 (3): 417–36.

Kecskeméti, Charles, and Iván Székely. 2005. *Access to Archives: A Handbook of Guidelines for the Implementation of Recommendation No R(2000)13 on a European Policy on Access to Archives*. Strasbourg: Council of Europe Press.

Kim, Julia Y. 2018. "Researcher Access to Born-Digital Collections: An Exploratory Study." *Journal of Contemporary Archival Studies* 5 (1): 7.

Körmendy, Lajos. 2007. "Changes in Archives' Philosophy and Functions at the Turn of the 20th/21st Centuries." *Archival Science* 7 (2): 167–77.

Kuny, Terry. 1997. “A Digital Dark Ages? Challenges in the Preservation of Electronic Information.” In *Paper Presented at 63rd Annual Conference of the International Federation of Library Associations and Institutions, Copenhagen, Denmark, August 31–September 5, 1997*. Den Haag: International Federation of Library Associations and Institutions. http://archive.ifla.org/IV/ifla63/63kuny1.pdf (accessed March 20, 2024).

Luthra, Mrinalini, Konstantin Todorov, Charles Jeurgens, and Giovanni Colavizza. 2023. “Unsilencing Colonial Archives via Automated Entity Recognition.” *Journal of Documentation*. https://doi.org/10.1108/JD-02-2022-0038.

Majtényi, László, Péter Molnár, Lukás Ádám Petri, and Máte Dániel Szabó, eds. 2005. *Az Elektronikus Információszabadság [Electronic Freedom of Information]*. Budapest: Eötvös Károly Policy Institute.

Mayer-Schönberger, Viktor. 2011. *Delete: The Virtue of Forgetting in the Digital Age*. Princeton: Princeton University Press.

Moss, Michael S., and Tim J. Gollins. 2017. “Our Digital Legacy: An Archival Perspective.” *Journal of Contemporary Archival Studies* 4 (2): 3.

Poole, Alex H. 2015. “How Has Your Science Data Grown? Digital Curation and the Human Factor: A Critical Literature Review.” *Archival Science* 15 (2): 101–39.

Ridge, Mia, ed. 2014. *Crowdsourcing Our Cultural Heritage. Digital Research in the Arts and Humanities*. Farnham: Ashgate.

Roberto, Claude M. 2017. “General Reflections on Improved Access and Public Engagement in the Past, Present and Future in Canada.” *Comma* 2015 (2): 49–53.

Stapleton, Larry, and Lise Jaillant. 2022. ““Born Digital” Shedding Light into the Darkness of Digital Culture.” *AI & Society* 37 (3): 819–22.

Székely, Iván. 2015. “The Four Paradigms of Archival History and the Challenges of the Future.” In *Management and Participation in the Public Sphere*, edited by Mika Markus Merviö, 1–37. Advances in Public Policy and Administration. Hershey: IGI Global.

Székely, Iván. 2017. “Do Archives Have a Future in the Digital Age?” *Journal of Contemporary Archival Studies* 4 (1): 17. https://elischolar.library.yale.edu/jcas/vol4/iss2/1 (accessed March 30, 2024).

Tyacke, Sarah, Jan Van Den Boeck, and Ellen Steendam. 1995. “Archives in a Democratic State.” *Journal of the Society of Archivists* 16 (2): 133–8.

## Interviews

Andreas Fickers (Professor for Contemporary and Digital History at the University of Luxembourg), interviewed by the authors, January 20, 2023.

Andrea Hänger (Vice President of the German Federal Archives), interviewed by the authors, February 7, 2023.

Charles (K.J.P.F.M.) Jeurgens (Professor of archival studies at the University of Amsterdam, the Netherlands), interviewed by the authors, January 26, 2023.

András Sipos (Main archivist of the Budapest City Archives, Hungary), interviewed by the authors, January 26, 2023.