The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-M-9-2025
30th CIPA Symposium "Heritage Conservation from Bits:
From Digital Documentation to Data-driven Heritage Conservation", 25–29 August 2025, Seoul, Republic of Korea

Digital Preservation of "Historical" Archives. A case study of the Documentary Archive of the "disappeared" Museum of Ethnology of Porto[1].

Carolina Teixeira Sousa [1], Maria Leonor Botelho[2], Luís Sebastian[3]

[1] CITCEM, Faculty of Arts and Humanities of University of Porto, Portugal – carolt.sousa.up@gmail.com
[2] CITCEM, Faculty of Arts and Humanities of University of Porto, Portugal – mlbotelho@letras.up.pt
[3]Digital Transition Department, Património Cultural, I.P., Portugal – luissebastian@patrimoniocultural.gov.pt

Keywords: Cultural Heritage at Risk, Historical Archives, Archival Methodology, Digital Preservation, Digitization.

Abstract

The digitization of inactive archives, commonly referred to as "historical archives," has emerged as a strategic and fundamental tool for the preservation, accessibility, and safeguarding of cultural heritage at risk of degradation or disappearance. However, despite the digitization process adhering to international guidelines and standards, which are intended to guarantee the efficiency, security, and quality of physical preservation and digital accessibility, the public availability of a detailed methodological manual that standardizes the techniques applied is neglected. This study proposes a clear and replicable methodology for digitizing inactive archives, applied to the case of the Museum of Ethnology of Porto (MEP) (1945–1994), whose remaining documentation was severely affected by the precarious conditions of the building and the successive transfers and reallocations of the collections. The digitization methodologies, the criteria developed, and the logic adopted for the different digital equipment, the tools for exporting formats, applications for image processing, and utilities for automatic metadata extraction and document renaming are described and justified, as well as the final integration into the arquiv@ platform. This methodology — which in itself constitutes an action of documentary safeguarding and preservation — aims to contribute to the dissemination of the memory of the MEP, which, although physically gone, can be (re)constituted through its archive. Additionally, it is also expected that this methodological proposal will serve as a reference for future projects, promoting the standardization of practices and interoperability in the digitization of "historical" archives.

1. Contextual Background

The study and preservation of documentary archives play a central role in the development of collective memory and the valuation of the cultural heritage of a community, institution, or territory. In addition to administrative documents, these archives contain together material evidence of the practices, knowledge, and modes of organization specific to a particular era, community, or entity. When inactive — or, more often, designated as "historical", that is, consisting of documents relating to closed cases after the respective conditions for reopening have expired (Cupido, 2023) — they become particularly vulnerable. These archives are frequently susceptible to the threat of disappearance or complete loss due to a variety of factors, including physical support degradation, the inadequate conservation conditions, and disorganization, dispersion, relocation, or abandonment (Flinn, 2007; Cook, 2013). Their loss represents not only material damage but also a break in the continuity of collective knowledge, posing an urgent challenge in the field of memory and conservation policies (Cook, 2013).

In this context, digitization is a strategic response to the fragility of these archives, functioning simultaneously as a mechanism for safeguarding the preservation, democratization, and accessibility to documentary heritage (Conway, 2010; Ketelaar, 2001). The creation of digital representations of physical documents makes it possible not only to reduce direct handling - and, consequently, the wear and tear on the originals - but also to significantly extend their durability. This practice ensures the physical safeguarding of the media and contributes to the intangible preservation of institutional memory and the records that support its identity, facilitating access to information by diverse audiences (Barbedo, 2005). However, the effectiveness of this process hinges on adopting consistent methodologies that ensure the integrity, readability, and contextualization of digitized content.

Several international entities — such as UNESCO, IFLA (International Federation of Library Associations and Institutions), and Europeana — have published guidelines and recommendations to guide and ensure minimum standards of quality, efficiency, and security in digitization projects, both in technical terms (resolution, formats, digital preservation) and archival terms (metadata, description, accessibility). Nevertheless, a significant gap remains: the lack of a widely disseminated, detailed and certified manual describing the techniques and methodologies applied to the digitization of historical archives. This absence compromises the standardization of practices, hindering the replicability of processes and interoperability between archival platforms and systems. In a landscape marked by the diversity of documentary archives and the irregular conditions of their state of conservation, it is urgent to reflect on the production and sharing of clear, adaptable, and rigorous methodological proposals that guide technical and scientific work around the digitization of historical archives.

This study aims to address this lacuna by presenting a clear, tested, and replicable methodology for digitizing historical archives, applied to the specific case of the Museum of Ethnology of Porto (MEP) (1945–1994), an institution that has physical and

[1] This research is part of the doctoral project "In Memoriam Museu de Etnologia do Porto: Salvaguarda, valorização e divulgação do arquivo documental e bibliográfico" (*In Memoriam Museum of Ethnology of Porto: Safeguarding, enhancing and disseminating the documentary and bibliographic archive*), developed in a non-academic context, with a fellowship awarded by the Fundação para a Ciência e a Tecnologia (FCT) in the 2022 competition. The research is hosted by Património Cultural, I.P.

materially “disappeared” and whose history remains largely unknown. The precarious conditions of the São João Novo Palace — the museum's only known headquarters — and the successive adversities to which the museum, its archive, and library were exposed — among which the fires of 1984 and/or 1987, the relocation and subsequent collapse of the roof of the Quartel de S. Brás, and repeated exposure to atmospheric conditions — led to the reallocation of the collections by various museum institutions, resulting in an unquantifiable loss of documentation and, consequently, in a progressive “disappearance” of this institution (Sousa, Botelho & Sebastian, 2024).

The Direção Regional de Cultura do Norte (DRCN) — succeeded, in 2023 by the Cultural Heritage Public Institute (PCIP) — the entity responsible between 2011 and 2023 for the management of the São João Novo Palace and the guardianship of the remaining documentary and bibliographic archive— proceeded in stages to transfer this archive to Casa Allen — the current headquarters of the PCIP, — with the aim of safeguarding what is considered to be the main vestige of the MEP's existence.

The scarcity of information about the MEP, from its foundation to its closure, represents one of the main weaknesses in the study of the institution. The lack of a rigorous inventory and the dispersion of records make the remaining documentary collection an indispensable resource for the reconstruction of its memory and for deepening knowledge about ethnographic museology in Portugal. In this context, the organization and digitization of this archive are priority tasks, not only as a preservation strategy, but also as an opportunity to produce knowledge and critically assess the history of the MEP.

In light of these concerns, the organization and digitization of the MEP's documentary archive is a priority task, not only as a methodological working tool, but also as an effective contribution to the preservation of the memory of this institution and to promoting the standardization of practices and the interoperability between “historical” archives.

Therefore, this article aims, on the one hand, to contribute to the preservation of a documentary archive at risk and, on the other, to offer a technical and scientific basis that can guide future practices in similar contexts.

The article is organized into three parts: contextualization and characterization of the archive; detailed description of the methodology developed for the digitization of inactive archives; and preliminary results.

2. Archive Context and Description

2.1 Scope of the Archive

The present study focuses exclusively on the MEP's documentary archive preserved at Casa Allen (Porto). While it is generally accepted that there are sporadic and individual documents in various institutions, this archive is, to date, the most significant and cohesive of the known documentary mass produced throughout the MEP's existence.

The archive, which consists of approximately 780 documents, is notable for its remarkable typological diversity, which reflects the functional and historical complexity of the institution. The most representative core corresponds to the administrative documentation relating to the museum's internal functioning and management, which includes the founding records such as the “Inauguration Act” and regulations, financial records, personnel records, official correspondence (files), reports on institutional activities, such as national projection inventories, and reports or projects on collaborations with external entities. Alongside this, it presents audiovisual documentation that incorporates a significant set of content related to cultural programming and mediation, such as photographs, projects, technical drawings, graphic and videographic materials produced and/or collected by the MEP. There is also a core of private documents associated with key figures in the foundation and development of the MEP, namely Pedro Vitorino, Fernando Pires de Lima, and Fernando Lanhas.

2.2 Initial Conservation State

Transported urgently to Casa Allen, after a process of cleaning and purging guided by the DRCN and followed by preliminary thematic organization by the technical team of the PCIP's Digital Transition Department — responsible for the digitization process of the archive under the supervision of that institution — the archive under analysis was in a fragile and disorganized state of conservation, without any internal coherence or functional reading sequence. The documents were randomly accumulated and showed clear signs of physical, chemical, and biological damage, notably the presence of moisture, mold, characteristic odors, sheets or documents stuck together, oxidation caused by hardware such as paper clips and staples, the presence of dead insects and traces of pests, partial loss of content due to tears or faded ink, and lost or damaged installation units. The existence of documentation that was completely compromised, with sheets stuck together and contaminated by fungi, was also identified.

2.3 Provenance and Conservation History

The lack of accurate records regarding their transfer makes it hard to piece together the document's journey. As mentioned earlier, the document archive has been moved around a lot, mainly because of the poor condition of the buildings and bad management of the spaces where it was kept. It is known that the archive was moved from the São João Novo Palace after a fire, followed by the collapse of the roof, in May of an uncertain year (1984 or 1987) to the Quartel de S. Brás, but it is unknown how it was transported and stored. Soares (2016) and Araújo (2021) later mention the activities carried out at the palace under the guidance of the DRCN. However, the chronology of the return to the palace is unknown.

However, it is these records, albeit very limited, do mention the existence of two phases (between 2017 and 2021) of organization, registration, and inventorying of the documentary and bibliographic archive prior to its transfer to Casa Allen. Despite these mentions, none of these records have come down to us, with the exception of the labeling carried out in the bibliographic archive.

2.4 Limitations

For this reason, it is not possible to accurately determine the full extent of the archive or to provide for its complete reconstruction. The lack of an original inventory and the scarcity of studies on the MEP make it impossible to quantify the total documentary universe produced by the institution. However, we are confident that, by cross-referencing with the remaining documentation that is stored separately in different institutions, we will be able to fill in the gaps on the subject in the field of historiography and make this museum, which has physically disappeared, known to the public.

The archive is still currently undergoing treatment, specifically in the process of being digitized and uploaded to the arquiv@© platform (PCIP's online archive). Following the conclusion of the process and the public defense of the project, it is anticipated that the project will be made fully available to the public, both physically and digitally.

3. Digitization Methodology for “Historical” Documentary Archives

3.1 Document Organization

Given the lack of knowledge and state of disorganization in which the archive was found, document organization proved to be fundamental not only for identifying the documentation that comprises it, but also for ensuring a rigorous (re)constitution of the archive’s structure, allowing for an informed reading of its content and technically oriented digitization.

The first methodological decision concerned the form of document organization. Classification by producer was rejected, since the MEP was, from the outset, interpreted as the unified producing entity. Although not all the documentation was originally produced by the MEP, it was this institution that, over time, gathered, organized, and preserved the archive, and was therefore responsible for its transmission.

Instead, a hybrid organizational methodology was chosen, based on the identification of document types and main themes, with subdivisions by subject and, where relevant, by chronology. This approach not only allowed us to reflect on the functional diversity of the archive, but also to reveal the MEP’s as yet little-known areas of activity, facilitated by the construction of a coherent archival structure based on the principles of provenance and original order.

The structure was materialized through the construction of an archive tree, which is a hierarchical representation of the archive structure organized, normally, by five levels of archival description:

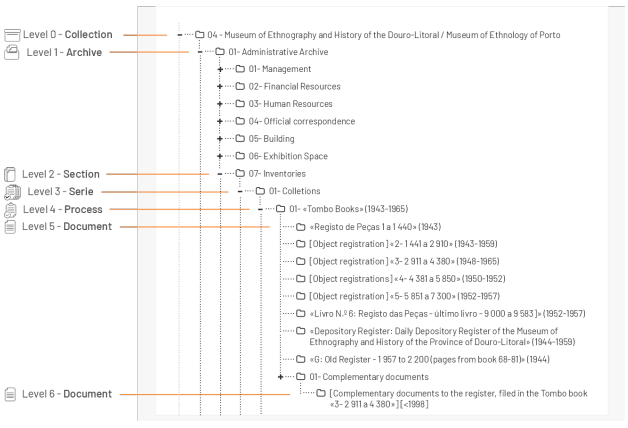

Figure 1. Partial view of the MEP archive tree structure, illustrating the levels of archival description. © 2025 Carolina Sousa

This model allowed for a functional and coherent organization of the archive, in line with the structure required for its future integration into the arquiv@© platform.

It is important to acknowledge that this representation of the archive corresponds to an interpretation that has been formed through a critical analysis of the documentation.

3.2 Digitization Criteria

In contradistinction contemporary archives, with standardized formats and predefined management systems, this documentary archive is characterized by remarkable material heterogeneity, reflected in the diversity of formats, support, dimensions, and states of conservation of the documents. The inherent complexity of the material imposed specific technical and operational requirements on the digital reproduction process, making it impossible to apply the “standard” model.

Concurrently, the absence of previously established uniform criteria, in conjunction with the presence of numerous sheets devoid of informational content, also posed significant challenges. The complete and indiscriminate reproduction of this support would entail a substantial increase in the volume of digital files, thereby generating unnecessary “visual noise” during reading. Furthermore, it would necessitate digital storage resources that were disproportionate to the objectives of accessibility and informational utility.

Added to these constraints was a critical variable: the time available for digitization *versu*s the unexpectedly vast and diverse volume of the document mass and the unpredictability of its state of preservation. Consequently, a meticulous and methodical approach was imperative to ensure the optimal utilization of available time and resources while maintaining the representativeness, legibility, and informational value of the archive.

In this particular context, a hybrid digitization methodology was formulated as a response to the aforementioned set of requirements. This methodology was guided by the principles of treating inactive (“historical”) and intermediate/active archives in a differential manner, contingent upon the type of document, its state of preservation, and its informational content.

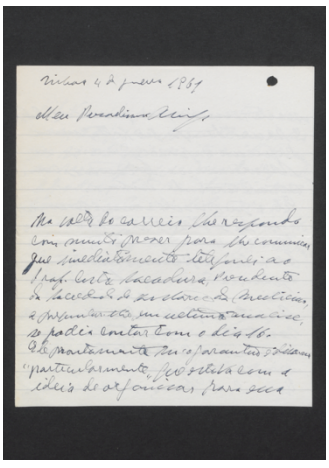

Figure 2. Exemple of document digitization following the inactive archive methodolody. Scnned using a planetary scanner Zeutschel OS 12000 at 300 dpi in RGB mode. © 2024 arquiv@

In the case of documents treated as belonging to an inactive archive — that is, completed processes with no prospect of reopening — full digital reproduction was carried out. All pages were digitized, including front and back, regardless of the presence of informational content, in order to ensure complete

capture of the support and preservation of its original physical layout. Visually, this type of digitization is identified by the presence of bleed (additional black or white margin), which guarantees the completeness of the captured image and respects the material limits of the document, while ensuring the fidelity of the digital record to the physical object.

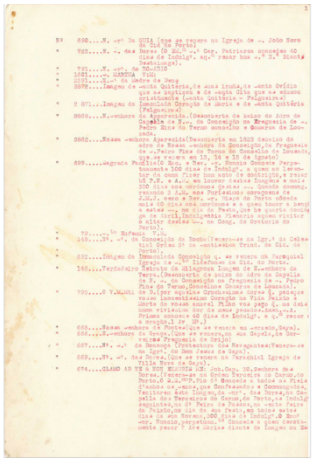

Figure 3. Exemple of document digitization following the intermediate/ative archive methodolody. Scnned using a flatbed scanner Epson WorkForce DS-6000N at 300 dpi in RGB mode. © 2024 arquiv@

Conversely, when designated as part of an intermediate/active archive — that is, content susceptible to reinterpretation, reorganization, or administrative reuse—a selective logic was implemented. Digitization was conducted exclusively on pages containing pertinent informational content, with exclusionary criteria applied to exclude blank sheets or pages deemed to lack content or informational value. This approach, which aligns more closely with the operational criteria of an active archive, enabled the optimization of resources and time, thereby ensuring the integrity of essential data. The digitization resulting have margins adjusted to the physical limit of the document, without additional bleed, accurately reflecting the size of the document.

This methodological definition also extends to the degree of granularity of the digitization. In documents treated as inactive, the complexity of the support, the physical fragility, and overlapping layers of information have, in several instances, necessitated the production of multiple images per sheet. In documents treated as intermediate/active, formal simplicity and greater structural homogeneity allowed for a one-to-one correspondence between the physical and digital documents, thereby ensuring that each physical document was reflected in a corresponding digital image.

Nevertheless, regardless of the model applied, all digitizations were performed at actual scale, respecting the orientation of the informational content and using the RGB color mode, ensuring the visual fidelity of the digital reproduction.

The selection of documents to be digitized followed cumulative criteria of informational relevance, legibility, and state of preservation, combined with operational factors such as the time available for execution and the technical complexity of the support. Priority was given to the documentation of the administrative archive — namely the sections relating to management, human and financial resources, the constitution of the collections, etc. — followed by the collections linked to the MEP's cultural mediation.

Each document was evaluated individually to determine the most appropriate digitization methodology: in cases where there was internal consistency, physical fragility of the support, informational fragility, and preservation value as a closed record — that is, it must be preserved in its entirety without intervention — the inactive archive approach was chosen, with full reproduction; on the contrary, in cases where the documentation had standardized formats and less material complexity, the intermediate/active archive methodology was favored.

Documents with no informational value or consisting entirely of blank pages were excluded from the digitization process. Nevertheless, all these cases were duly recorded and justified in the context of the document evaluation, ensuring methodological transparency and traceability of the process. This practice reinforces the commitment to a critical and informed preservation policy that prioritizes archival utility and future accessibility without compromising document integrity.

3.3 Equipment and Software

Consequently, the application of this methodology required close coordination between the criteria defined and the technical resources available. The selection of equipment and software was based on criteria of compatibility with the archive's various physical formats, the fragility of the support and their state of conservation. The heterogeneity of the material thus required a diversified and flexible technological response.

In this regard, three digitizers, characterized by distinct functionalities and applications, have been utilized:

- EPSON WorkForce DS-60000N: flatbed scanner with ADF (Automatic Document Feeder), used mainly for standard format documents (A4 and A3) and in good condition. This equipment enables efficient scanning of large volumes of sheets, with the possibility of automatic double-sided capture, and is widely used for documents processed according to the criteria of the intermediate/active archive.
- CONTEX IQ Quattro X A0: roll-feed scanner for large formats (A1, A0 and above), particularly technical plans, maps and posters. Its use was essential to guarantee the reproduction integrity of documents whose size and fragility prevented them from being handled by conventional equipment.
- ZEUTSCHEL OS 12000 Advance: overhead camera planetary scanner, suitable for fragile and non-standard documents. By avoiding direct contact with the support, it guarantees safe scanning. This equipment was reserved for documents treated as inactive files.

The quality of the digitization, in turn, was ensured by strict technical parameters defined according to the fidelity of the reproduction and its suitability for long-term preservation. The standard resolution used was 300 dpi (Dots Per Inch), which was increased to 600 dpi whenever the legibility of the information content required it. The digital matrices were produced in TIFF (Tagged Image File Format), given their stability and suitability for preservation.

Complementary software was used for post-digitization processing and file management:

- Adobe Acrobat Pro: used for optical character recognition (OCR) and for creating multipage PDF files from the TIFF matrices;

- ImageMagick: an open-source tool that enables the automatic conversion of formats, particularly from TIFF to PDF, speeding up the production of versions for dissemination;
- A-Ren: a specific application for automatic batch renaming, which is essential for ensuring uniform identifiers and the sequential organization of digital documentation.

From the TIFF matrices, multipage PDF files were generated, with a resolution of 150 dpi for access purposes, or 300 dpi in cases where the correct and complete reading of the content requires greater definition. Whenever technically feasible, OCR was applied to enhance textual search and facilitate access to information.

3.4 Image Processing

The image processing was deliberately kept to a minimum, in keeping with the principle of fidelity to the original. The only systematic intervention consisted of manually correcting the orientation of the images in cases where there were misalignments resulting from the digitization process, especially in documents processed with the planetary scanner, whose suspended capture system, although minimally intrusive, did not always guarantee perfect alignment of the support.

In documents scanned with EPSON equipment, attention was focused on the appropriate configuration of the automatic margins and the precise definition of the size of the support, in order to avoid involuntary cuts in the information content. No contrast, brightness or sharpness filters were applied, since all the images produced were fully readable, clear and unobstructed.

It is imperative to acknowledge that a meticulous examination of all images was conducted post-digitization, ensuring adherence to the stipulated technical and methodological criteria. To ensure the integrity of the original record, any excessive automation that could potentially compromise its visual or informational authenticity was deliberately avoided. Consequently, the authenticity and legibility of the original record were preserved in all cases.

3.5 Metadata, Metainformation and Digital Organization

The digital organization, metadata and assignment of metainformation were also structured using a hybrid methodology, combining an automatic component (metadata) - generated and extracted directly by the scanning equipment and software - with a manual component of descriptive and administrative metadata, produced during the document appraisal process. This approach ensured effective integration between the technical data and the archival description required for integration into arquiv@©.

The automatic component was ensured by the scanning devices themselves, whose software embeds a set of essential technical metadata in the TIFF files. This includes: resolution, bit depth, color mode (RGB), image dimensions (width, height, total number of pixels), file format, software and its version, model and manufacturer of the capture equipment, operating system, specific scanning settings, production date and file identifier. This information, which remains associated with the digital matrix, constitutes a fundamental technical record for the reproducibility and future validation of the process.

In addition to this, the digitization of each file — which resulted in the creation of a multipage PDF file with optical character recognition (OCR) — was manually supplemented with technical data, such as the author of the scan, the reproduction rights (Património Cultural, I.P.© - arquiva.patrimoniocultural.gov.pt), and the identifier (ID) assigned in the document assessment process (e.g. MEP_00000_0000).

Each digitized document was also associated with an individual technical file, inserted as the first page of the PDF, containing essential additional information, such as identification of the entity responsible (Património Cultural, I.P.), archive nucleus (Museum of Ethnology of Porto), destination platform (arquiv@©), identification code (ID), title and description of the document, geographical area, dates of production and digitization, names of those responsible for document evaluation and digitization. To this information was added an explicit indication of the methodology applied to digital reproduction — “Digital reproduction entirely in accordance with the physical file” or “Digital reproduction limited to physical support with information content” — clarifying the archival framework (inactive or intermediate/active) of each file.

This assignment was complemented by support tools, namely the A-Ren application, which ensured consistent nomenclature. The standard defined for the identifiers was based on the fund's acronym — MEP — followed by a five-digit sequential numbering assigned at the time of the document assessment (MEP_00000), plus four additional digits indicating the order of the images within the file (MEP_00000_0000). Whenever the backs of the pages were digitized, the letter “V” was added (MEP_00000_0000_V). This system ensured uniform nomenclature, correspondence between physical and digital documents and a functional organization, compatible with the archival structure.

In terms of the digital structure, the organization of the files replicated the archive tree — made up of five main levels — by creating a hierarchical directory structure with “chained” sub-folders that reflect the different levels of description. Within these, the TIFF files were preserved at maximum quality (resolution of 300 or 600 dpi), ensuring correspondence between the logical organization and their digital representation.

Finally, taking into account the limitations set by the arquiv@© platform - which imposes a maximum limit of 199MB per file - the PDFs were divided, where necessary, into sequential volumes (e.g. MEP_00000_0000_P1, MEP_00000_0000_P2), guaranteeing in all cases informational continuity and the link to the original identifier.

3.6 Integration into the *arquiv@* plataform

The final stage of the technical treatment process entailed the integration of the MEP’s documentary archive into the arquiv@© platform. This integration ensured the long-term digital preservation of the archive and its future public accessibility. This stage was based on three fundamental operations: the design of the archive tree, the archival description according to international standards and the sequential uploading of the PDF files and their metadata.

The data is uploaded to the platform manually, document by document, based on the structure defined in the archive tree. The description fields follow the normative standards of the DGLAB (Direção-Geral do Livro, dos Arquivos e das Bibliotecas) and other international archival description standards, including areas such as identification, context, content and structure, conditions of access and use, associated documentation, notes, access points, description control and administrative area. Whenever possible, we have tried to fill in these fields exhaustively, in order to

guarantee the completeness, reliability and future retrievability of the information by users.
Once the description has been completed, the corresponding multipage PDF file is uploaded, maintaining the link between each level of description and the respective digital content.

Although uploading is currently being carried out manually, an automation script based on CSV files is being developed to optimize the process and reduce the technical processing time, while maintaining compliance with the standards required in the archive@©.

All this work is being carried out as part of the ongoing PhD project, which is why, although arquiv@© allows documentation to be searched, viewed and downloaded publicly, full access to the entire digitized archive will only be available after the public exams have been defended. In addition, the project adheres to the principles of open science, which promote free access to scientific knowledge, and the FAIR principles (Findable, Accessible, Interoperable and Reusable), which ensure that research data can be easily found, accessed, integrated and reused. To this end, it is also planned to make the database available in the future and to draw up a data management plan to ensure that it is organized, preserved and shared responsibly.

The integration of the MEP's documentary archive into this digital archive represents a concrete contribution to the dissemination of knowledge about a "disappeared" institution at risk of oblivion, its historical and scientific revaluation, long-term digital preservation and the interoperability of archival information, by allowing different audiences — researchers, cultural institutions and citizens — to access, reuse and critically contextualize the museum's material memory.

4. Contribution and Preliminary Results

The digitization methodology presented here is one of the main contributions of this study. Developed from the critical and empirical analysis of a document archive with atypical characteristics — in terms of the diversity of formats, types and state of preservation - this methodological proposal is a concrete response to the lack of public and systematized guidelines for the digitization of historical archives. By combining principles applied to inactive archives with flexible approaches inspired by intermediate/active archives, a hybrid model was built, adjusted to the material and informational specificities of the MEP archive.

The model adopted demonstrates the effectiveness of digitization — with around 70% of the documentation digitized so far — through clear criteria of legibility, informational relevance and material integrity. The delineation between full and partial reproduction, in conjunction with the definition of strict naming and digital organization, have simultaneously ensured document fidelity, efficient structuring and the future reuse of information. The integration of metadata — compatible with arquiv@© — ensures interoperability, accessibility and long-term digital preservation.

From a scientific point of view, the preliminary results of applying this methodology already reveal a strong potential for (re)constituting the history of the MEP, particularly its previously unknown museological and administrative practices. The process has also made it possible to identify gaps, continuity and internal dynamics in the archive, making visible a hitherto dispersed or largely unknown institutional operating logic.

Thus, the methodological proposal presented in this article is not limited to the treatment of a specific case, but offers a replicable model for other institutions faced with historical archives at risk, fragmented or lacking technical treatment. In this way, it is believed to have contributed to filling a recognized gap in the field of archives and to promoting good practices in the digitization and management of documentary heritage, particularly in the context of the digital transition of cultural institutions. By proposing a systematized methodology that is technically feasible and sensitive to the specificities of historical archives, this study reinforces the importance of digital preservation strategies that not only safeguard institutional memory, but also make it accessible to future generations.

5. Conclusion

In summary, the methodology delineated herein is the result of an applied archival practice, adapted to a real case of heritage emergency, and is based on principles of technical rigor, methodological coherence and future accessibility. Its replicability in similar contexts reinforces its value as a guiding tool for institutions faced with documentary archives at risk of disappearing or disintegrating. By digitizing, structuring and making accessible the archive of the Museum of Ethnology of Porto — an institution whose memory survives today mainly on paper — this proposal contributes not only to preserving its history, but also to strengthening a more critical, participatory archival culture that is prepared for the challenges of the digital age.

Acknowledgements

This research was conducted as part of a doctoral project entitled "Museum of Ethnography and History of the Douro Litoral: Safeguarding and creating a memory space", funded by the Fundação para a Ciência e a Tecnologia (FCT), under the individual doctoral scholarship with the reference 2022.13991.BDANA.
https://doi.org/10.54499/2022.13991.BDANA

The project is also supported by Transdisciplinary Research Centre «Culture, Space and Memory» (CITCEM) — financed by National Funds through FCT, under the project reference UIDB/04059/2020.
https://sciproj.ptcris.pt/157505UID

References

Araújo, J., 2021. O "livro antigo" na biblioteca do Museu de Etnografia e História do Douro-Litoral. *História*. Porto: Faculdade de Letras da Universidade do Porto. IV Série, 11(1), 229-275.
https://ojs.letras.up.pt/index.php/historia/article/view/8766/9793

Barbedo, F., 2005. Arquivos digitais: da origem à maturidade. *Cadernos BAD,* 2, 6-18.
https://doi.org/10.48798/cadernosbad.810

Conway, P., 2013. Preserving Imperfection: Assessing the Incidence of Digital Imaging Error in HathiTrust. *Dig. Tech. Cult. Magazine*, 42(1), 17-30.
http://hdl.handle.net/2027.42/99522

Cook, T. (2013). Evidence, Memory, Identity, and Community: Four Shifting Archival Paradigms. *Archival Science*, 13(2-3), 95–120. https://doi.org/10.1007/s10502-012-9180-7

Cupido, A., Vieira, J.E.F, Pavão, L., Sebastian, L., Brandão, M., Costa, M., Oliveira, M., Rodrigues, M., Martins, O., Seabra, P. & Miguel, R., 2023. *Arquiv@: Objetivos, Conceitos, Conceção e Produção***.** Porto: Direção Regional da Cultura do Norte – Ministério da Cultura, (Património a Norte, 14). https://culturanorte.gov.pt/patrimonio-a-norte-14-construcao-do-arquivo-online-da-drcn/

Duranti, L., & Rogers, C. (2012). *Trusting Records in the Cloud: The Creation, Management and Preservation of Trustworthy Digital Content*. UBC.

Europeana, 2021. *Publishing Framework and Guidelines*. https://pro.europeana.eu/page/europeana-publishing-framework

Flinn, A., 2007. Community histories, community archives: Some opportunities and challenges. *Journal of the Society of Archivists,* 28(2), 151-176. https://doi.org/10.1080/00379810701611936

Gilliland, A. J. (2014). *Conceptualizing 21st-Century Archives*. Chicago: Society of American Archivists.

Harvey, R. (2011). *Preserving Digital Materials* (2nd Ed.). Berlin: De Gruyter Saur.

IFLA, 2002. *Guidelines for Digitization Projects*. The Hague: International Federation of Library Associations and Institutions.

Ketelaar, E., 2001. Tacit Narratives: The Meanings of Archives. *Archival Science*, 1(2), 131–141. https://doi.org/10.1007/BF02435644

Soares, C., 2016. *A casa de S. João Novo e o Museu de Etnografia e História da Província do Douro-Litoral (1710-2016): estudo histórico-integrado, problemas e reflexões para a sua salvaguarda.* Porto: Faculdade de Letras da Universidade do Porto, 2016

Sousa, C., Botelho, M. L., Sebastian, L., 2024. "In Memoriam" Museu de Etnologia do Porto: digitalização e divulgação do arquivo documental através do arquiv@. Proposta metodológica. *Antologia de Ensaios. LABORATORIO COLABORATIVO: Dinâmicas Urbanas, Património, Artes. X Seminário de Investigação, Ensino e Difusão.* 155-173

UNESCO, 2003. *Charter on the Preservation of the Digital Heritage*. Paris: UNESCO.