# Digital Archives: An Overview on Digitization

**Lorenzo Ferrari, researcher**
**Research Center on Digital Humanities of the University of Modena and Reggio Emilia (DHMoRe)**

## 1.      What is an Archive?

An archive is an organized body of records created, received, or otherwise acquired by a creator – an organisation, institution, family, individual – in the course of its activities. These records are preserved because of their enduring value, whether administrative, legal, historical, or cultural. An archive is based on the relationship between the production, accumulation, and preservation of materials. Although it is difficult to define them briefly, archives, in the traditional sense of the term, are paper castles that humanity has always built to feel alive and not forget itself. In a way, they are granaries of facts and words and serve to sustain the societies that express them.

It is important to distinguish archives from libraries. While libraries collect published materials that exist in multiple copies, archives preserve unique and often irreplaceable documents. These may include letters, administrative files, photographs, maps, audio recordings, videos, and born-digital materials. There is therefore a great variety of documents, depending on their type and medium.

Archival work is guided by two key principles. The first is provenance, which means that the link between an archive and its creator should be maintained. The second is original order, which requires preserving the arrangement established during the active use of the records. These principles are fundamental because they preserve the context that gives meaning to archival materials, even though the advent of digital technology has called into question the very concept of physical order.

## 2.     Why Inventory an Archive?

In general, inventorying an archive means describing its contents in a systematic and structured way. The purpose of an inventory is not simply to list documents, but to establish intellectual and physical control over the archive.

A well-prepared inventory allows archivists and users to understand what the archive contains, how it is organized, and how the records were produced. It supports preservation planning, helps prevent loss or misplacement, and enables access for researchers, students, and the wider public.

In this sense, the inventory is the foundation of all other archival activities, including digitization.

## 3.     Hierarchical Levels of Description

Archival description is hierarchical. This means that records are described at different levels, from the most general to the most specific.

At the highest level, we find the fonds, which represent the entire body of records created by a single entity. Below the fonds are series, which group records related to the same function or activity. In some cases, series may be further divided into sub-series. At lower levels, we encounter files or units, and, only when necessary, individual items.

This hierarchical approach reflects the structure of the archive and avoids excessive item-level description, which is often impractical. International standards such as the General International Standard Archival Description, provide guidance on how to describe each level consistently. We can say that archives are tools of consistency because they stabilise thoughts and words by organising them hierarchically.

## 4.     From Inventory to Digitization

Digitization is the process of creating digital reproductions of archival materials. It is important to emphasize that digitization does not replace the original physical archive, nor does it diminish its value. Instead, it enhances access and supports preservation by reducing physical handling.

Digitizing an archive cannot be reduced to a restyling operation or an attempt to copy it and then read it on a computer. This is a delicate transition, which must be managed with great awareness and responsibility.

Digitization projects should always be based on a solid inventory. Without accurate descriptive data and metadata, digital files risk becoming isolated images with little contextual value. In other words, digitization without inventorying leads to digital disorder.

## 5.     Why Digitize Archive?

There are several reasons why institutions choose to digitize archives. Digitization improves accessibility, especially for users who cannot visit the archive in person. It supports teaching and research, and it increases the visibility of collections.

Digitization can also protect fragile materials by limiting the need for direct handling. However, it is not a neutral or purely technical process. It requires careful planning, sustainable funding, and awareness of legal and ethical issues, such as copyright and privacy. This is because, in the digital sphere, the concept of the archive expands, becoming fluid, delocalised, and streamlined, and cannot be a disjointed technological appendage of the original physical archive.

Digitizing an archive is not the task of a single professional figure. It requires an interdisciplinary team in which different forms of expertise converge. Archivists, conservators, restorers, scholars, IT specialists, project managers and other experts all contribute to the process, each addressing specific aspects of description, preservation, technology, and access.

Dematerialising an archive is not a mechanical act, a mere change of medium. Dematerialisation is a complex and pervasive process, of which digital objects are only the final consequence. In this sense, digitizing an archive means rethinking it, reconfiguring it as a whole, and even questioning its identity and the stories it tells through its documents.

## 6. Selection and Preparation

Not all archival materials can be digitized. Selection is a critical and interpretative phase, shaped by multiple factors such as research demand, historical and cultural value, physical condition, and institutional priorities. Digitization projects are rarely neutral or exhaustive: limitations in funding, time, and technical resources almost always require choices to be made.

Digitization can therefore represent an investment in both the preservation and accessibility of an archive. At the same time, selection decisions influence which materials become more visible and which remain less accessible, with long-term implications for research and interpretation.

Before digitization begins, materials must be prepared. This includes verifying inventory information, checking conservation needs, and ensuring that documents can be safely handled and reproduced.

## 7. The Digitization Workflow: Images

The digitization workflow typically includes image capture using scanners or digital cameras, followed by quality control procedures. These checks ensure that the digital files are complete, accurate, and faithful to the originals.

Usually, institutions create high-quality master files for long-term preservation, along with smaller access files for online use. Secure storage, backup systems, and clear file-naming conventions are essential components of this workflow. Technical standards, such as resolution, file formats, and color management, play a crucial role in ensuring the long-term usability of digital materials.

The International Image Interoperability Framework, commonly known as IIIF, is a set of shared standards designed to improve the way digital images are delivered, accessed, and reused on the web. Rather than functioning as a single platform, IIIF provides a framework that allows different institutions to make their digital images interoperable. This means that images preserved in one digital archive can be viewed, compared, annotated, and reused across multiple environments and tools, without being copied or duplicated.
However, the workflow is not limited to the creation of digital copies.

### 8.    The Digitization Workflow: Metadata

Digitized documents require metadata to remain meaningful and accessible. Metadata, literally data about data, provide information about what a digital object is, how it was created, and how it can be used.

We can distinguish between descriptive metadata, used for identification; administrative metadata, which include the information to help manage a resource; and structural metadata, that is to say metadata to describe the internal structure of documents and manage the relationships between the various parts of digital objects. Common standards such as Dublin Core are widely used in digital archival systems. Generally, the development of a metadata template, which must be completed by cultural institution staff or other researchers, is a fundamental step in the workflow.

Ideally, digital metadata should be derived from and aligned with the original archival inventory. However, it is worth remembering once again that the digital archive is not so much the archive of the physical archive, but a new web of dynamic relationships between digital objects that were previously unthinkable. Unlike archives in the traditional sense, digital archives are not simply created by the actions of a creator, but rather by specific objectives and information needs, which only apparently do not require the mediation of an archivist or other professional.

## 9.    Digital Archives

The digital archive is a broad category, not always in line with the requirements of traditional archives. It demands a degree of methodological flexibility that cannot be taken for granted. There are electronic archives that are document management systems; invented archives, that is thematic digital collections independent of their creators; and again participatory archives, created through collaborative practices and where the need to build a collective, shared memory prevails; living archives, conceived as a social resource for civil and cultural progress.

And finally, learning archives, where the potential of artificial intelligence is amplified. Like archives themselves, artificial intelligence relies on data to generate new data and can be affected by cognitive bias. But here, the original spontaneity of the archive and its sedimentation gives way to complex algorithmic structures, opening up scenarios that are still largely unpredictable. This shift raises critical questions about transparency, accountability, and control. When algorithmic processes intervene in archival description and access, it becomes essential to understand how decisions are made, which data are prioritized, and which perspectives may be marginalized. This transformation invites a profound reflection on the nature of archival work.

In addition, digital archives often blur the boundaries between creation, preservation, and access. Actions such as digitization, metadata enrichment, and interface design become intertwined, demanding collaboration among specialists in archival theory, information technology, conservation, and user experience. The digital archive is therefore both a technical and intellectual ecosystem.

## 10. Digital Preservation and Sustainability

Digitization is not a one-time activity, but an ongoing responsibility and a work in constant improvement. In many ways, digital copies are fragile in their own way: they depend on hardware, software, file formats, and institutional continuity.

Digital preservation means ensuring that digital copies remain accessible, authentic, and usable over time. This requires regular monitoring of storage systems, the use of redundant backups, and the planned migration of files to new formats or platforms as technologies evolve.

Equally important is documentation. Every digitization project should record the standards, tools, and workflows adopted, so that future archivists can understand how digital objects were created and managed. Without this contextual knowledge, digital files risk becoming silent data, detached from their origins and difficult to interpret.

Sustainability also has an institutional dimension. Long-term digital preservation requires clear policies, stable funding, and trained personnel. In this sense, preserving digital archives is not only a technical task, but an ethical responsibility toward future generations.