
Digital Archives: Elements of Definition

1.1. Key concepts of digital archives

Before talking about digital archives, it is appropriate to present in the foreword some elements of definition relating to key concepts related to archiving in general, namely, archives, archival tools and procedures for sorting, transfer and disposal, among other things.

1.1.1. *Archives*

It has now been more than 70 years since the International Council on Archives (ICA), representing archival professionals from around the world, was founded, and efforts are being made to develop and implement both a body of global archival legislation and archival training and research programs. Among the first formal definitions, we cite that of ICA:

Archives are the documentary by-product of human activity retained for their long-term value. They are contemporary records created by individuals and organizations, as they go about their business and therefore provide a direct window on past events. They can come in a wide range of formats including written, photographic, moving image, sound, digital and analogue. Archives are held by public and private organizations and individuals around the world. [ICA 16]

Today, the majority of countries have legal texts defining and organizing archives as “[...] documents, whatever their date, form and material support,

produced or received by any natural or legal person, and by any public or private organization, in the exercise of their activity”¹.

Archives can therefore be public, coming from the activities of the State, public institutions (industrial and commercial [EPIC] or administrative [EPA]) and any other legal entity under public law or legal persons under private law managing a public service (financed by a public fund for a general interest).

They can also be private, coming from natural persons or persons with a private status such as families, unions, political parties and associations.

1.1.2. *Archive management*

Records management today is the basis for all actions related to good governance, respect for the law and the collective memory of humanity, the rights of citizens to access information and administrative transparency [ICA 16].

Indeed, in all public or private organizations, information recorded on various media is created and/or received “involuntarily” in the course of people’s activities. Over time, this information accumulates and increasingly hinders work, requiring intervention. What should be retained? Why keep archives? What is their purpose? For how long? These are in addition to many other questions related to the value of these documents, distribution, access rights and places of conservation, among other things.

1.1.2.1. *Conservation objects*

A priori all administrative documents are concerned by the conservation for different periods according to their nature and their value, which could be administrative, legal or historical. A specific retention period is therefore assigned to each document.

1.1.2.2. *Conservation objectives and utility*

We are obliged to preserve archives primarily for their administrative or legal value. Indeed, archives are an integral part of an organization’s

¹ The French Heritage Code (art. L. 211-4); Tunisian Law No. 88-95 of August 2, 1988 on archives (art. 1).

information system and represent the backbone of its proper functioning on the administrative, and also financial and legal levels. They may also have historical value as witnesses to the past. François Mitterrand (President of France from 1981 to 1995) summed up the answer to these questions about the values of archives in the following terms: “Archives in all countries, by keeping track of yesterday’s acts and their paths, illuminate but also command the present. Those who act responsibly are well aware that one does not set directions in ignorance of the past” [MIT 88]. We also keep archives for the good governance of organizations, saving space, time and money.

In other words, archives are evidential documents that allow the continuity of administrative services, historical research and economic, social and cultural development.

1.1.2.3. *Shelf life*

As noted above, each document is assigned a retention period. This retention period represents the continuous process that a document must go through from its creation to its final disposition, which may be destruction or deposit in an archive for its historical value. Each retention period varies according to the informational, administrative or legal value of the document. Archives are therefore successively called “current”, “intermediate” and then “permanent”, which is known as the lifecycle of archives or the theory of three ages [PER 61]:

- *Current* or *active archives* represent documents that are regularly used in day-to-day work and are generally used to manage ongoing business. They are kept in offices close to the users;

- *Intermediate* or *semi-active archives* represent documents that no longer have an immediate and daily use, but which must be saved because of a possible reopening or legal prescriptions. Since the frequency of use is low, these archives can be moved to another location for consultation if necessary or entrusted to an archive service that manages access to them on demand;

- *Permanent* or *historical archives* are those archives that are no longer useful for the conduct of administrative activity and which are of historical or patrimonial interest. These archives must be kept for an unlimited period of time. It should be remembered that a very large number of archives that are not intended to enter the active age must be disposed of after agreement from the archive service.

1.1.2.4. *How to keep archives*

Preserving archives also depends on the location, nature and value of the documents. It is a matter of keeping a trace of all documents at every stage of their lives. Three principles are to be respected in this sense:

- the provenance of the collection, which consists of not mixing documents;
- the order of the collection, which consists of keeping the documents according to the classification made by the original organization;
- the integrity of the collection, which consists of not splitting a collection that has already been set-up.

1.1.3. *Archival management tools*

1.1.3.1. *Inventory of documents*

The document inventory is a basic tool that consists of a good knowledge of all the documents (whatever their form, age and place of conservation) of the organization and its various administrative structures, as well as their modes of creation, use and conservation. This inventory may be general, covering all the organization's documents regardless of the unit that has custody of them, or specific, covering only the documents of a particular department, directorate or sector.

1.1.3.2. *Directory of typical files*

One of the most important tools for setting up an archiving system is the “directory of typical files“ or “nominal list of typical files”. This tool helps the records manager identify all types of documents and records produced or received by the institution in the course of its administrative activities. It can be summary, analytical and/or regular.

1.1.3.3. *Functional classification scheme*

A classification scheme (or file plan) is an organization tree of the main functions of an organization that classifies recorded information from the general to the specific, by subject of activity, using a tree-like structure. In this system, a rating is assigned to each subject to facilitate the retrieval of information.

Developing a records classification scheme is essential to the smooth running of all administrative activities in each organization. Indeed, the logical organization documents produced and received by each institution and their grouping into structured folders help to locate them easily. This classification, based on the principle of the hierarchy of subjects, consists of presenting the classes from the general to the specific (see Figure 1.1).

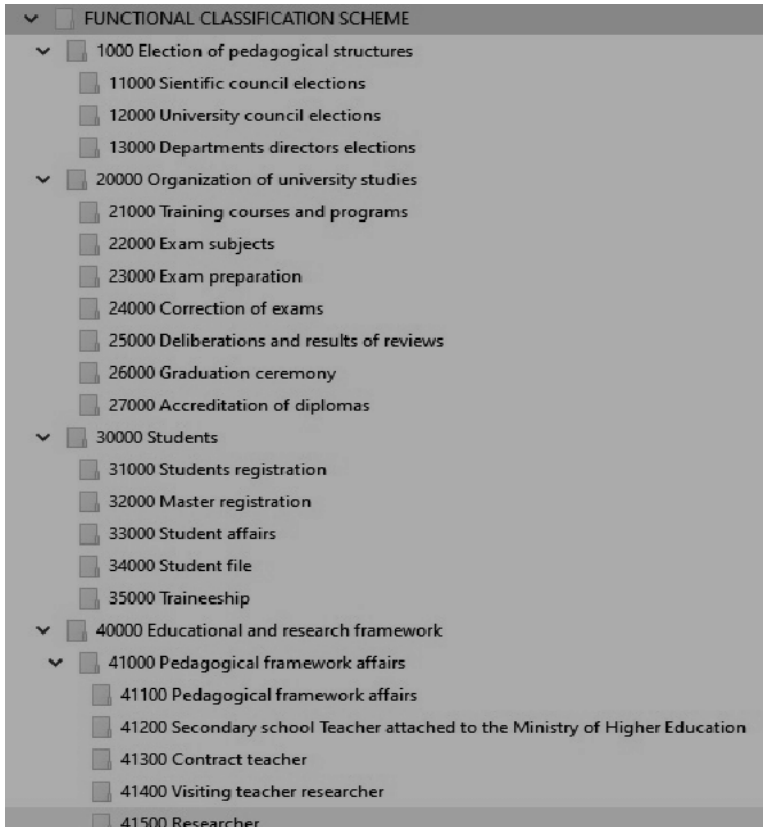


Figure 1.1. Example of part of a classification structure for higher education documents. For a color version of this figure, see www.iste.co.uk/mkadmi/archives.zip

1.1.3.4. Schedule of retention rules

A retention schedule, also known as a “sorting table” or “retention rules schedule”, is the cornerstone of an archives management system. It is a tool that indicates the retention period for each type of document, as defined by its owner

or required by law, with its final fate at the end of this period (disposal or transfer to historical archives). These retention periods are managed by so-called “retention rules” (see Figure 1.2). These are used to determine the minimum retention period for any document, even if nothing prevents the institution from archiving them longer, unless the document contains personal data.

10 ELECTIONS OF PEDAGOGICAL STRUCTURES				
11 TEACHER ELECTIONS FOR THE SCIENTIFIC COUNCIL				
Copy	Unit	Active	Semi-active	Inactive
Main [c]	General Secretariat	3 ¹	3 ¹	E ²
Secondary [c]	Scientific Council	X ³		E
Notes: (E = Elimination; C = Conservation) ¹ Conservation 3 years in active age and 3 years in semi-active. ² Elimination. Elimination concerns only paper documents, but the list of voters, the list of candidates, the list of elected officials and the minutes of the elections must be kept in digital form. ³ Retention 3 years after receipt of the document.				
10 ELECTIONS OF PEDAGOGICAL STRUCTURES				
12 STUDENTS ELECTIONS FOR THE SCIENTIFIC COUNCIL				
Copy	Unit	Active	Semi-active	Inactive
Main [c]	General Secretariat	3 ¹	3 ¹	E ²
Secondary [c]	Scientific Council	X ³		E
Notes: (E = Elimination; C = Conservation) ¹ Conservation 3 years in active age and 3 years in semi-active. ² Elimination. Elimination concerns only paper documents, but the list of candidates, the list of elected officials and the minutes of the elections must be kept in digital form. ³ Retention 3 years after receipt of the document.				
10 ELECTIONS OF PEDAGOGICAL STRUCTURES				
13 STUDENT ELECTIONS FOR THE UNIVERSITY COUNCIL				
Copy	Unit	Active	Semi-active	Inactive
Main [c]	General Secretariat	3 ¹	3 ¹	E ²
Secondary [c]	Scientific Council	X ³		E
Notes: (E = Elimination; C = Conservation) ¹ Conservation 3 years in active age and 3 years in semi-active. ² Elimination. Elimination concerns only paper documents, but the list of elected officials and the minutes of the elections must be kept in digital form. ³ Retention 3 years after receipt of the document.				

Figure 1.2. Example of retention rules. For a color version of this figure, see www.iste.co.uk/mkadmi/archives.zip

Together with a classification scheme, the retention schedule is the cornerstone of any document management system in a public or private organization. By specifying retention periods and media, retention policies ensure the effective management of the lifecycle of all types of records or series of records, including distinguishing between inactive records that should be retained over the long-term and those that should be disposed of [BAN 18].

1.1.4. *Digital archives*

Discussing digital archives (also called “electronic archives”) is in a way also discussing digital archiving, since archives have already been defined previously. It is rather the digital, with characteristics related to the immateriality of objects that raises many questions. We must go back to 1984 when the ICA in its dictionary of archival terminology defines digital archives as archives that are “... generally encoded, readable only by machines, recorded on media such as discs, drums or magnetic tapes, cards or punched tapes” [WAL 84, p. 103]. They are either natively digital documents or dematerialized paper documents.

The immaterial nature of these types of objects certainly brings considerable advantages, but it also implies a set of problems. This immateriality allows *a priori* the storage of large quantities of documents in reduced volumes, which enables access to these documents in a fast, instantaneous manner and for a greater number of users. It also makes it possible to enhance and enrich archive collections through statistical, linguistic and cultural studies. However, these intangible aspects pose a number of challenges, particularly in terms of the evidentiary value of documents, as well as the risks of loss, alteration and modification of documents, the technological obsolescence of storage formats and media, and so on. This requires the development and implementation of complex technical systems.

1.2. Electronic Records Management

1.2.1. *ERM: elements of definition*

The term “ERM” (Electronic Records Management) represents a set of techniques and tools that allow documents to be transformed into digital

formats, and filed, managed and stored on a computer platform as part of normal business operations. Also known as “DMS” (Document Management System) for the electronic management of existing information and documents, this concept aims to manage the complete cycle of documents and all types of information from their creation to the determining of their disposition (permanent retention or destruction).

The function of electronic document management is to manage, using computer applications, digital documents within an organization in the normal course of business. It is a process that helps to develop a learning culture in organizations through collaboration, capitalizing on information and exchanges. In fact, different categories of ERM could be deployed in different institutions according to need, company size and management objectives [PRA 04]. We can already distinguish different types:

- Administrative DMS: usually part of a global management application, this category allows users to quickly access various administrative documents that are often in digital image formats;

- Office DMS: in line with standard office automation applications (Word, Excel, etc.), this category is aimed at group work while integrating document enrichment functions through annotations, messaging functions, workflows and so on. Lotus and Microsoft tools are good examples of this type of application;

- COLD (Computer Output on Laser Disc) DMS: used for computer data, this application allows us to store and automatically index the documents generated by management applications. This solution is composed of several software modules allowing the formatting, extraction and indexing of data from spool files in direct connection with a central computer. It is possible to share files instantly between several users, while guaranteeing data protection through passwords. The information is usually presented on screen in the form of a listing;

- Documentary DMS: directly resulting from documentary research, this type of application is generally used in libraries, scientific documentation services and possibly for a press review. It uses more or less heterogeneous indexing and search methods (full-text, thesaurus, proximity operators, concept search, natural language, etc.);

- Technical DMS: directly linked to a profession, it is also called “business DMS”. It is an application for managing and manipulating the formats

specific to a profession, such as maps, plans, technical documents related to industrial projects, and so on;

– Image DMS: this type of application generally groups together all ERM applications and enables the management, as its name suggests, of images or scanned documents.

The service of the National Archives of Luxembourg [ANL 11] summarizes the main differences between an ERM system and an EAS (electronic archiving system) in Table 1.1.

	ERM system	EAS
Purpose and use	Share and disseminate digital information using common business applications across the enterprise	Preserve digital documents and data while maintaining their integrity, durability and readability over time
	Manage documents on a daily basis to ensure the conduct of business activities within the company	Ensure the security of documents with evidential value for the company
Technical form	Computer software installed on a server with a database for storing information and application tools to acquire, store and distribute information via API-type interfaces	Much more complex computer software with functions that ensure compliance with standards (especially those related to authenticity and durability, such as ISO 14641 and OAIS) The system also includes storage disk sets (bays) and tools for identification, security, timestamping, etc.
Features	Authorizes document modification and versioning	Prevents modification of documents
	Authorizes the destruction of documents by their authors	Prevents the destruction of documents by their authors except during the regulated application of the disposition by the administration of the EAS
	Can be based on a classification scheme and retention schedule. Under user control	Must be based on a classification scheme and a retention schedule that ensures that the final disposition of the documents is applied
	Does not meet the constraints of durability, legibility and integrity of documents over time	Compulsorily meets the constraints of durability, legibility and integrity of documents over time through a monitoring of formats and software and a policy of migration of documents and media

Table 1.1. *Differences between an electronic records management system and a digital archiving system [ANL 11]*

1.2.2. *ERM: implementation steps*

ERM is traditionally presented in four main stages, which are the acquisition, management, storage and distribution of documents. These stages can be complex, with other secondary or complementary stages depending on the size and type of project.

1.2.2.1. *Acquisition of documents*

The primary function of ERM is document acquisition. It results from a human or automatic process of dematerializing documents through “digitization” (by means of scanners), production and/or integration. This process aims to create, save, classify and index the digital document.

The creation of the documents comes from different sources. The digitization of existing paper documents through scanners is the most widespread and fastest process. The scanner, also known as “digitizer”, is a data-processing device that enables the transformation of a document or a real object into a digital image. The choice of the scanners and their characteristics depends on the formats and the states of the documents to be digitized, as well as the needs in a given digitization project. Several models of scanner exist. These include:

– *Flatbed scanner*: this is the best-known and most widely used type of scanner for scanning unbound printed documents and handwritten pages (see Figure 1.3). This type of scanner consists of placing the document to be scanned on a glass plate and a light source mounted on a movable arm then scans the entire document to transform it into a digital image.



Figure 1.3. *Flatbed scanner (Epson). For a color version of this figure, see www.iste.co.uk/mkadmi/archives.zip*

– *Vertical scanner*: this is a somewhat special version of the flatbed scanner (see Figure 1.4). The difference between the two is that, in this

vertical version, the light and the sensors are placed at the top in a support bar under which the bound document can be placed open and facing upward. This type of scanner is very convenient for objects with some relief and for fragile objects, such as bound books.



Figure 1.4. Vertical scanner (Amazon). For a color version of this figure, see www.iste.co.uk/mkadmi/archives.zip

– *Scanner with feeder*: uses the same technology as flatbed scanners, but with a feeder instead of a glass pane (see Figure 1.5). This type of scanner is well suited for scanning separate pages of the same size and is strong enough to withstand the handling of the feeders.



Figure 1.5. Scanner with charger (Canon). For a color version of this figure, see www.iste.co.uk/mkadmi/archives.zip

– *Drum scanner*: this allows the best scanning among all types of scanners, with the highest resolution (see Figure 1.6). They are suitable for fragile documents and a certain skill is required in handling them. This type of scanner generally costs more than other types and is not suitable for all types of documents; they must be able to wrap around the cylinder and must be no thicker than 1 mm.



Figure 1.6. *Drum scanners (Fujifilm). For a color version of this figure, see www.iste.co.uk/mkadmi/archives.zip*

– *Book scanner*: this is a device used to scan bound documents, such as registers or books (see Figure 1.7). There are many different models of book scanners with varying capacities.



Figure 1.7. *Book scanner (Wikimedia France). For a color version of this figure, see www.iste.co.uk/mkadmi/archives.zip*

– *Scanner with digital camera*: for three-dimensional objects that are difficult to digitize flat, as well as for fragile documents.

– *Scanner for microfiche and microfilm*: this is used to digitize microfilm and microfiche.

All scanners work in almost the same way and in accordance with three principles: lighting, reflection and capture:

- the scanner uses a lamp to illuminate the page to be scanned;
- the mirror system leads the rays reflected on the page to a sensor bar called a CCD (Coupled Charged Devices) bar;
- the received light is transformed by the sensors into an electrical signal and is then processed by the electronic part of the scanner.

The choice of scanners is made according to several characteristics, of which the main ones are the resolution (measured in dpi [dots per inch]), speed (pages per minute), – the type of acquisition (black and white, grayscale, color)– and the format of the documents (A4, A3, A1, A0, etc.).

The creation of documents also comes from:

- digital document exchange: this is generally done through an EDI (Electronic Data Interchange) device by agreeing on a standardized data format (EDIFAC for Europe and ANSI X.12 for the United States);
- digital document production: in addition to the operations related to the management of existing documents, the DMS is also involved, in an equal way, in various operations of document production. Indeed, workflow tools allow the different agents to work on administrative procedures dealing with documents, scheduling, routing and job tracking. Groupware also offers communication, cooperation and document sharing functionalities.

1.2.2.2. *Document pre-processing*

After passing the documents through a scanner, the result is always a file in an image format. The nature of these images depends on the scanned original documents and on the subsequent processing. These images can be, according to requirements, in black and white (or converted to black and white), in dark or light gray or in color. Color images can be 8, 16, 24, 30 or 36 bits. Each time the resolution increases, the clarity and size of the image increases.

Several types of processing can be provided to be able to exploit the digitized documents:

- *Compression*: It consists of reducing the size of files, thus reducing the space used on archiving media and facilitating their circulation on networks. Several compression methods exist, depending on the scanning method and the nature of the original documents:

- CCITT² G3/G4 compression, also known as “G4” or “modified reading”, is a lossless image compression method used in Group 4 facsimile machines, as defined in the ITU-T T.6³ fax standard. It is only used for bitonal (black and white) images. Group 4 compression is available in many proprietary image file formats, as well as in standard formats such as TIFF (Tagged Image File Format), CALS (Computer-aided Acquisition and Logistics Support), CIT (Combined interrogator transponder, Intergraph Raster Type 24) and PDF (Portable Document Format),

- JBIG⁴ (Joint Bi-level Image Group) compression: this is a two-level compression of an image, in which a single bit is used to express the color value of each pixel. This standard can also be used to code grayscale images and color images with a limited number of bits per pixel. JBIG is designed for images sent using facsimile coding and offers significantly higher compression than Group 3 and 4 facsimile coding,

- the JPEG⁵ algorithm (Joint Picture Expert Group) is used to reduce the size of color images. This format of graphic file allows very important compression rates, but with a weak resolution that influences the quality of the image: the compression entails a loss of information;

2 CCITT (Consultative Committee for International Telephony and Telegraphy): this is a standardization body that has developed a series of communication protocols for the facsimile transmission of black and white images over telephone lines and data networks. These protocols are formally referred to as CCITT T.4 and T.6 standards, but are more commonly referred to as CCITT Group 3 and Group 4 compression, respectively.

3 ITU-T T.6 is a recommendation of the International Telecommunication Union that defines the fax coding schemes and fax coding control functions for Group 4 facsimile machines.

4 JBIG (Joint Bi-level Image Expert Group) is a group of experts that develops standards for bi-level image coding. Officially, JBIG is the ISO/IEC JTC1 SC29 working group, which is also responsible for the JPEG standard for a suite of compression algorithms used to produce image files supported by web browsers and typically used for complex images such as photographs.

5 JPEG (Joint Photographic Experts Group) is a standard that defines the recording format and decoding algorithm for a compressed digital representation of a still image.

– *Optical Character Recognition (OCR)*: The purpose of *OCR* is to convert text in image format into a computer-readable text format by translating the groups of dots in a scanned image into characters with the associated formatting. It is carried out by dedicated systems called “OCR”. The challenge today is to find the most efficient OCR among several tools of this type and the best suited to its application. Among the criteria for the choice of the tool, we often evoke the criterion of effectiveness, which is related to a high recognition rate. The objective to be reached is a rate of 100%. However, the recognition rate does not depend solely on the recognition engine, but also on several other measures to be taken into consideration, such as the material preparation of the paper document upstream and the performance of the OCR engine in the parameters used to adapt to the type of content, taking into account, inter alia, the language, quality and layout of the document.

OCR can be applied within an ERM system in two ways:

1) Application on whole pages in text in order to index them in full text using spell checkers.

2) Application on some areas within the pages (such as titles) in order to use them as an index. Different technologies have existed for a long time and are based on OCR techniques to extract information from these digitized documents and enrich their metadata (category, author, title, date, etc.):

– *Automatic Document Recognition (ADR)*, which consists of distinguishing one type of document from another, according to a few pre-defined parameters. This will make it possible to sort images electronically;

– *Automatic document reading*: this technology uses artificial intelligence technologies to perform linguistic checks on recognized words and interpret them using text-mining functions, for the purpose of pre-analysis and/or thematic classification of the scanned documents.

In addition, this OCR technology is always limited and depends on the quality of the text to be scanned (if it is distorted, faded, stained, folded, contains handwritten annotations, etc.) and on the quality of the scan itself. It often generates several interpretation errors that require human intervention to be corrected, otherwise raw OCR makes it impossible for the text to be read and indexed by search engines. This is why this work is generally outsourced to service providers who use low-cost labor or Internet users (in the absence of financial means). The latter alternative, which is increasingly

used by library and archive services, is called crowdsourcing. Several OCR projects have been developed through this alternative with regard to the correction of digitized newspaper texts for the National Library of Australia, the correction of OCR through gamification for the National Library of Finland and the involuntary correction of OCR via reCAPTCHA for the Google Books service, among other projects [AND 17].

1.2.2.3. *Document indexing*

After having acquired the document through scanning, exchange and/or production, and in order to find it and facilitate its use, it is necessary to describe its content. This second stage of electronic document management is the most important one as regards being able to keep the document and use it later. This operation can be done by type (with a formal description, author, title, date, etc.), by concepts or keywords selected in a free way, or based on a thesaurus in order to harmonize practices. In web documents in HTML format, the description is created through META tags that allow the creator of these documents to define the relevant keywords representative of the content, the subject, the author and so on. There are many metadata⁶-related standards today, such as DC (Dublin Core), RDF (Resource Description Framework), EAD (Encoded Archival Description), EAC (Encoded Archival Context) and LOM (Learning Object Metadata) [MKA 08]. The objective is to make this metadata usable by a large number of search tools.

1.2.2.4. *Storage of documents*

1.2.2.4.1. *Storage media*

Storage, or what is sometimes called archiving (in the primary sense of the term), supports the conservation of documents over time. In order to implement an effective storage solution, it is first necessary to establish a needs analysis related, in particular, to the volume of data, their importance, the frequency of their consultation, the degree of confidentiality, the degree of importance of security, the length of time they are kept and the interest of putting them online, among other factors.

To facilitate the different needs of this conservation function, an ERM system uses several storage media, according to the following criteria:

⁶ Metadata are the set of identification and description elements of electronic resources. These elements are increasingly modelled and standardized to unify and generalize their use.

- criteria relating to the document: types of documents, frequency of consultation, interest in having it online and retention periods;
- criteria relating to the medium: document access time, storage capacity, cost, rewritability or non-rewritability and secure access.

There are several storage media that can be classified into generations:

- First generation media are considered to be analog media and have not been used since the late 1990s. This refers to the perforated card and perforated tape system, which originated in the 18th century. Their storage capacity is very small and is measured in a few tens of bytes.

- Second generation media are magnetic media and have a digital recording mode, except for magnetic tape, which has both analog and digital recording modes. They include magnetic tape, cassette, hard disks, cartridges and diskettes. These media have, however, been able to withstand technological developments over a long period of time [FLE 17].

- Third generation storage media are considered to be recordable digital optical media. We are talking about CDs (Compact Disk), DVDs (Digital Video Disk) and Blu-ray disks, also known as BDs⁷. In today's market, we are talking about several new optical media such as glass discs, M-Disc (the main characteristic being that the burning layer is made of synthetic diamond) and nanoform (a disk that has a very high resistance to damage).

- Fourth generation storage media are considered removable storage media (flash memory). This refers to USB (Universal Serial Bus), SD (Secure Digital), microSD, SSD (*Solid-State Drive*) and so on. The advantages of this type of media lie in its high speed access and storage capacities, which are close to those of a "classic" hard disk, the case of SSDs.

- Storage media under development are revolutionary storage media that should offer high storage capacity and high access speed, such as DNA (deoxyribonucleic acid), which is the hereditary genetic material of all living species on the Earth. According to the researchers, we estimate that the information stored on this support could be preserved over hundreds of thousands of years [CHU 12, COL 14]. Another equally revolutionary medium is quartz, which has proven its solid thermal resistance.

⁷ This is a digital disk format patented and marketed since 2006 by Sony (Japanese industrialist) allowing videograms to be stored and remastered in high definition.

1.2.2.4.2. Strengths and weaknesses of storage media

By definition, a document storage medium must perform at least four functions:

- data integrity (all data without any modification);
- data accessibility (data must be accessible according to a classification scheme);
- data intelligibility (reading comprehension);
- data security (protection of information);
- data permanence (guaranteed conservation over time).

1.2.2.5. *Dissemination of documents*

The last step in electronic document management is to distribute the documents. This is seen as the ultimate goal of this technology. Any document that is digitized, indexed and stored is intended to be found in searches for several reasons, including reading it, processing it again and transmitting it to others. Documents could be disseminated in different ways using different technologies depending on the need. It could occur through display on screen or printing or communication through networks such as email or workflow⁸. Searching for these documents occurs through a controlled query language or in natural language using a combination of keywords.

This diffusion operation can be automated through a workflow system that allows modeling of all the steps and actors involved in the ERM process. This workflow system optimizes the management, processing and communication of documents within the company. It also saves time for all employees by automating all repetitive and tedious tasks. Each employee is thus able to handle a large number of requests and concentrate on activities with high added value.

1.3. Records management

If ERM involves all types of documents and is essentially concerned with the ease of daily records management, Records Management (RM) is

⁸ A workflow can be defined simply as an industrial, administrative or other process through which a job is passed from start to finish.

essentially concerned with the management of archival documents (records) that provide evidence of a company's activities.

The concept of records management has its origins in the late 1940s in the United States based on the theory of archiving for the benefit of producers (Schellenberg in the United States and Pérotin in France). In the mid-1990s, an Australian RM standard was very successful and later became ISO 15489 in 2001, being revised in 2016. This standard defines RM as “a field of organization and management concerned with the effective and systematic control of the creation, receipt, retention, use and ultimate disposition of records, including methods of securing and preserving evidence and information related to the form of records”.

Applying the ISO 15489 standard essentially consists of:

- identifying documents that are binding for the company (records), whatever their format and medium;
- managing the documents throughout their lifecycle, in order to protect them and make them accessible over time. This management includes the organization of documents, their indexing, their conservation and the procedures for transmitting them.

In other words, the purpose of the RM is to guarantee the existence, accessibility, authenticity, reliability, integrity and usability of all documents.

The 15489 standard “complements the ISO 9001 (2000) quality management systems and ISO 14001 (2000) environmental management standards. Indeed, while these standards contain requirements for the production of documents known as ‘quality’ documents, to attest and testify to the activities of an organization, they do not contain any precision on what specifically constitutes a ‘quality document’. These clarifications are contained in ISO 15489, which aims to implement standards of quality and excellence in the management of routine and intermediate records in organizations of all categories” [COU 06].

1.3.1. Structure of standard 15489

Standard 15489 (2001 version) is presented in two parts, the first of which presents the guiding principles (concepts and principles) – ISO 15489-1, with the second presenting the guidelines (practical guide).

ISO 15489-1:2001, revised in 2016, determines the guiding principles and concepts on which the implementation of methods for the creation, capture and management of business records is based. These concepts and principles relate to business records with their metadata and document systems, their control, creation and management processes, as well as policies on accountability, monitoring, training and so on [ISO 16].

This first part of ISO 15489-1:2016 applies to all business records regardless of their structure, date, format and the technological environment in which they are created.

The second part of ISO 15489-2:2011⁹ represents a practical implementation guide for the first part ISO 15489-1 for records managers and archivists in general. This guide “provides a methodology that will facilitate the implementation of ISO 15489-1 in all organizations that need to organize and manage their archival records. It provides a general overview of the processes and considerations for organizations wishing to comply with ISO 15489-1” [ISO 11].

1.3.2. Content of the standard

As noted above, ISO 15489 is primarily intended to:

- produce and preserve documents while maintaining their authenticity, reliability, integrity and usability in operations in an effective and efficient manner;
- create a relevant record of the different activities of the company;
- protect and defend the company’s rights.

By “authenticity”, we mean the possibility of proving at any time that a record “is what it purports to be”, that it was created or received by the person claiming to be its creator or receiver, and at the time it purports to have been created or received. In order to ensure this characteristic of authenticity, a procedural manual must be in place to control the creation and/or receipt of records, their transmission, preservation methods and ultimate disposition.

⁹ This second part has not been revised and has already been annulled (according to its status on the ISO website).

Reliability refers to the exact representation in the document of the activities and operations described in its contents. In order to guarantee this characteristic, the document must be created at the time of the activity by the people who carried out the activity, and must contain all the information that makes it possible to carry out the same activity again and obtain the same result.

Integrity, on the other hand, is about protecting the document from any possible modification or alteration without documented authorization. In order to do this, a system must be in place to identify and keep a record of every action that affects the document, including consultation, modification and deletion.

Usability consists of maintaining the usability of the document at all times, and therefore it must be kept in a format and in a medium that guarantees its reusability for as long as possible. A policy of migrating to another medium must therefore be established periodically.

Finally, *metadata* aims to describe the context of the operation, the dependency relationships between documents and document systems, links between the social and legal contexts, and the relationships with the actors involved in the creation, management and use of business documents. To play their role fully, certain metadata must be assigned at the time of the capture or creation of the business record.

According to ISO 15489, in addition to document requirements, records management systems must:

- *comply* with the various requirements of the organization's functions and with the various laws, standards and good practices in the field of activity. This conformity must be maintained throughout the life of the system;
- be *reliable* by containing documents and metadata that properly retain the production context. Systems must be scalable to meet the needs of the organization without affecting the integrity of the documents and their value, even after transfer and/or migration operations;
- be *extended* to integrate all the activities of an organization, even in stages, given the very high cost of implementing such a system;

- have *integrity* by ensuring controls on user identities, access rights, security and the various manipulations of documents;
- be of a *systematic nature* by ensuring that the program is linked to all its processes and that all methods of document creation and storage are systematized [COU 06].

1.3.3. Design and implementation of an RM project according to the standard

According to ISO 15489, any project to design and implement a records management program goes through eight stages (see Figure 1.8).

1) *Preliminary survey*: the preliminary survey is the first step in designing a records management program and is the foundation for the entire intervention. It concerns all of the company's activities and practices, as well as its structure, culture and legal, regulatory and economic environment. This survey allows us to recognize all types of documents resulting from the activities of the company, and to detect the strengths and weaknesses in the management of these documents.

2) *Analysis of activities*: the second step consists of conducting an analysis of the company's activities during the preliminary survey. This analysis can lead to an activity classification scheme, a process map and work methods. These documents can be the basis for the development of archival tools, such as the retention period repository, the thesaurus or the list of authorities.

3) *Identification of archival requirements*: the third step follows on from the first two steps and concerns the identification of archival requirements. This identification consists of defining the needs of the company in terms of capture (creation and/or acquisition), management, organization, conservation and dissemination of documents. This definition of needs would make it possible to know the types of documents that the organization would be able to use to ensure a certain efficiency in the realization of its various activities and to illustrate them while guaranteeing its rights. This is in fact the main objective of standard 15489, which is to establish a records management program capable of creating, managing and preserving reliable, honest, authentic and usable records, as mentioned above [COU 06].

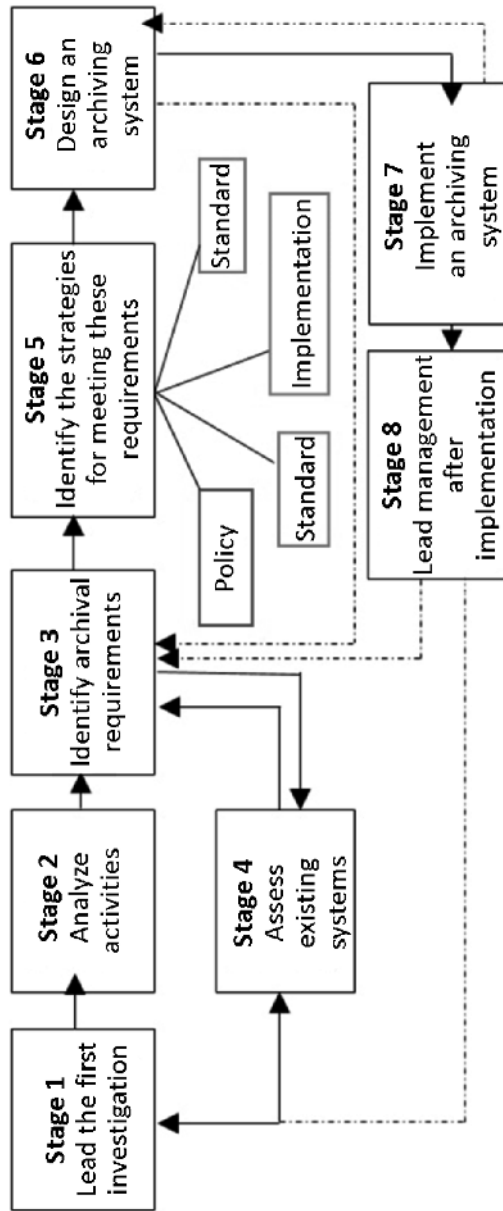


Figure 1.8. RM project design and implementation (ISO standard 15489) [COU 06].
For a color version of this figure, see www.iste.co.uk/mkadm/archives.zip

4) *Evaluation of existing systems*: this evaluation consists of deciding which of the methods and systems used could be kept, either in their current state or with adaptations. This decision must be taken after a serious analysis of the strengths and weaknesses of these existing systems, particularly in terms of efficiency and safety. It must always be kept in mind that a records management system must be part of the general information system.

5) *Identification of strategies for system development and implementation*: this fifth step consists of identifying strategies to meet archival requirements, thus developing and implementing the new records management system. These strategies consist of defining procedures (creation and codification of documents, metadata, recording, access rights and security, management, borrowing of documents, retention periods, transfer, disposal, control, etc.), policies (objectives and broad outlines of the system), retention rules (retention period and final disposition of all types of documents) and access.

6) *System design*: the sixth step represents the first operation in the development of the records management system, the design of the archiving system adapted to the company and the drafting of specifications. A specification is a document that formalizes the descriptive elements, processes and features selected for the system in question. It is the result of an internal reflection within an organization between different experts, document managers, computer scientists and others, after having made an analysis of the existing system, a study of needs, an opportunity study and a risk study.

7) *System implementation*: this step consists of implementing the system and training employees on its use. This training can be held at two levels: general training using the technical guides and more detailed training on the functionalities of the system that can be provided by the quality advisor, with the participation of the archivist.

8) *Follow-up and control after implementation*: this step consists of evaluating the records management system, once it has been implemented and is functional. This evaluation is carried out through internal and external audits regarding compliance with standards (notably ISO 15489). In order to ensure that the system can evolve, revisions of certain parts of the system, as well as the quality documents, should be considered every five years [COU 06].

1.3.4. *MoReq: the added value of RM*

As noted, ISO 15489 is perhaps the most influential standard for records management. It proposes that an organization use a RMS (Records Management System) to implement various management processes ranging from identifying the information that needs to be managed as records to documenting the process of managing those records, including creating, recording, classifying, storing, processing, controlling access, tracking records and disposing of records. It defines such a system as “an information system that captures, manages and provides access to records over time” (ISO 15489-1:2001, section 3.17). MoReq (Model Requirements for the Management of Electronic Records) is a practical specification for defining a records system expressed as a modular set of requirements. First published in 2001, MoReq has been continuously developed by the DLM Forum¹⁰ and the European Commission to create a series of specifications: MoReq1 in 2001, MoReq2 in 2008 and MoReq10 in 2011.

MoReq2010 is the latest simplified version. It aims to provide more flexibility in its specification and more openness to reach both the public and private sectors. It goes beyond the general description proposed by ISO 15489 to add a little more detail and specificity in the case-handling processes. One of its strengths, in addition to modularity, lies in the potential for interoperability that it guarantees between different archiving systems (see Figure 1.9). This interoperability allows all organizations to resolve the problems of migrating records from one system to another, as required by technology developments every 3–5 years [DLM 10].

MoReq2010's service-based architecture allows each organization to incorporate into its records management policies those parts of the specification that are relevant and applicable to the type of organization and to comply with the regulatory requirements of its industry. Not specifying any particular solution, MoReq2010 describes the essential elements that an archiving system should have to ensure effective records management, open

¹⁰ The DLM Forum is a European community of members of public archives and parties interested in the management of archives, documents and information throughout the European Union.

and immediate accessibility, retention of documents for as long as they are needed, and proper and rigorous disposal¹¹.

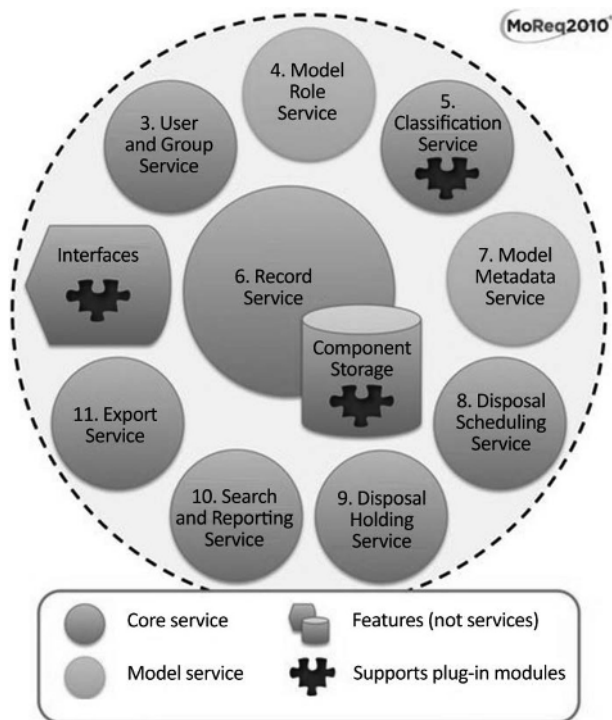


Figure 1.9. *MoReq2010 service-based architecture [DLM 10]. For a color version of this figure, see www.iste.co.uk/mkadmi/archives.zip*

1.4. EDRMS: merging ERM and RM

EDRMS (Electronic Documents and Records Management System) is a software application that manages a range of digital information, including word processing documents, spreadsheets, emails, images and scanned documents. As a result, an EDRMS can combine the functionalities of electronic document management and records management. Indeed, the main objective of electronic document management is the access and use of digital documents created from short-term office applications. Records

¹¹ We will return to these processes and elements of a record management system according to MoReq in section 2.4 of Chapter 2.

management, on the other hand, focuses on the management of information over time for evidentiary purposes, compliance with legislative and governance requirements and ensuring continued authenticity. EDRMS, on the other hand, can combine different functionalities of the first two systems, according to [NAA 11], and is able to:

- create and collaborate on new documents;
- capture records in a central repository for easy search and retrieval;
- integrate common office applications, such as email;
- make links between records;
- describe the records in a controlled way to facilitate searches;
- capture and maintain context-related metadata;
- assign unique identifiers to records during capture;
- protect the integrity of information and ensure its security;
- retrieve information using keywords, creators’ names and registration numbers;
- prevent recovery of “read-only” copies and limit deletion permissions;
- share documents and information to enable collaborative work;
- manage documents in a workflow process;
- control document versions and access authorizations;
- apply records retention rules and capture destruction metadata.

1.5. ECM: the overall data management strategy

ECM (Enterprise Content Management) is a term that has often been confused and used in the same way as ERM, IDMS (Integrated Document Management Systems), RM or even EDRMS. We have previously presented the main characteristics of each of these terms and we have shown that EDRMS has merged the two ERM and RM technologies. Since then, several important evolutions and additions have been made, both to the content to be managed by the latter technology, consisting, in particular, of integrating emails and different workflows, and also, and above all, to the management environment that has become the Web (see Figure 1.10).

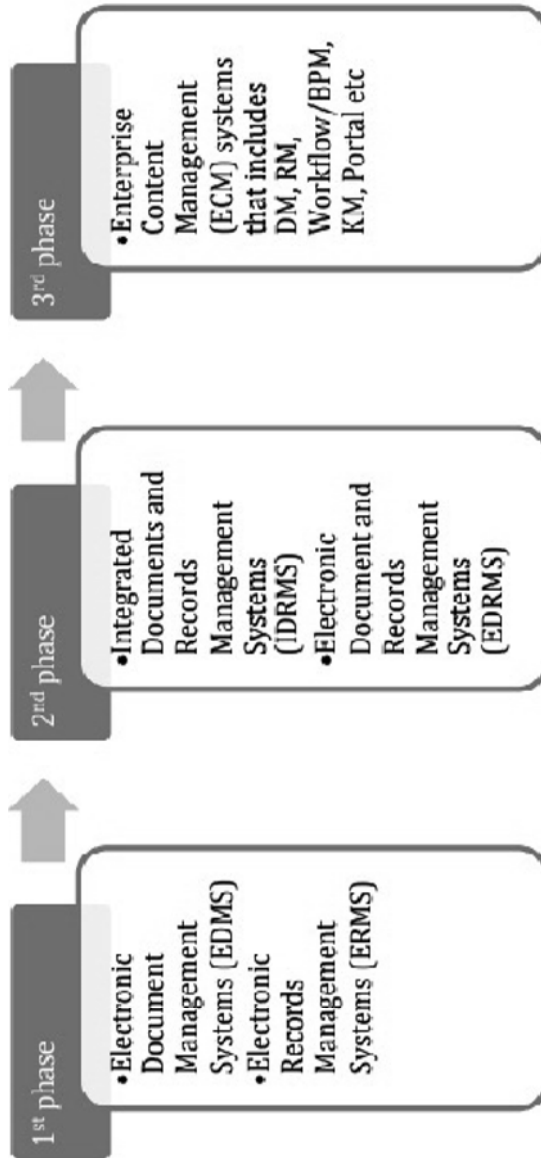


Figure 1.10. Evolution of different concepts leading to ECM [KAT 12]. For a color version of this figure, see www.iste.co.uk/mkadm/archives.zip

We can therefore summarize, according to [KAT 12], the process of the evolution of concepts in three phases:

- the first phase is represented by the ERM and RM concepts;
- the second phase is marked by the IDRMS and EDRMS concepts;
- the third phase is marked by the term ECM, which represents a type of system that includes ERM, RM, workflow, BPM (Business Process Management), portals, KM (Knowledge Management) systems and so on.

In this sense, ECM includes all the tools, methods and strategies for managing content regardless of its nature, whether structured or unstructured, and its type [AII 10]. It provides collaborative tools to facilitate teamwork and to improve and refine work processes. By targeting organizational processes, these methods and strategies are illustrated in several modules (see Figure 1.11).

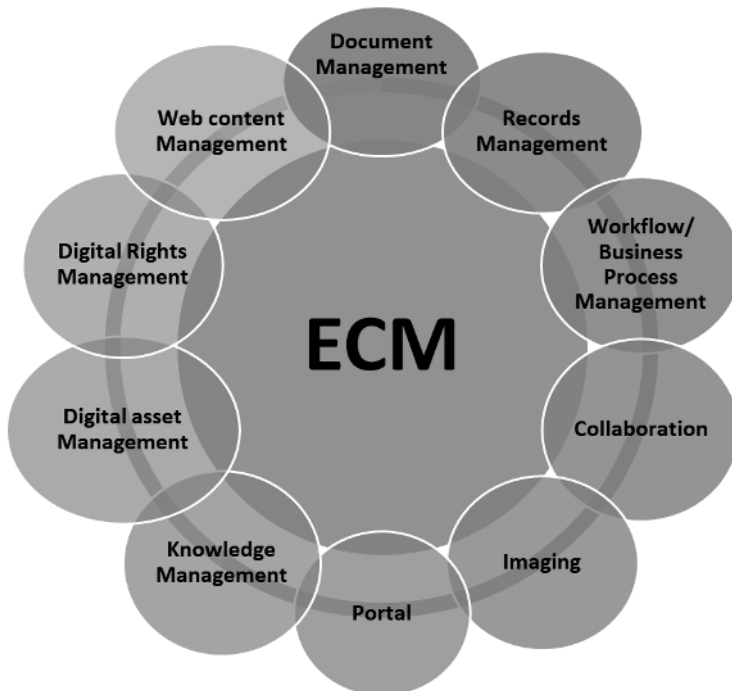


Figure 1.11. *The modules of a typical ECM application [KAT 12]. For a color version of this figure, see www.iste.co.uk/mkadmi/archives.zip*

Enterprise content management is therefore a working platform that includes a number of modules that can be expanded according to the needs, size and objectives of the company and that are defined in the document management policy. It can include, among others, several fundamental modules, which are records management, workflow, business process management, collaboration, knowledge management, digital asset management, digital rights management and web content management [KAT 12].

The advantages of this technology are that, compared to the other technologies mentioned above, it provides a complete mastery of the different business lines and functions of the company, and consequently a control over the different documents and incoming and outgoing flows. With collaborative tools, ECM ensures more collective intelligence. Nevertheless, to manage digital archives, it must be accompanied by a policy of awareness of different users in order to develop a culture of sharing and openness on new methods of processing, management, access and dissemination of archives, as well as new practices and the roles of archivists¹².

1.6. Conclusion

We have tried, in this introductory chapter, to give an idea of the terminology related to the field of archiving, while presenting some elements of definitions of archives, the challenges and objectives of their management, and the traditional tools on which any archiving policy is based. We also treated the transition of these archives into digital forms through the technology of electronic document management and records management. We have shown, on the one hand, the difference between these two technologies through the object of their management, namely, all types of documents for electronic document management and documents validated for records management, and, on the other hand, their link through EDRMS technology, as well as the evolution toward content management systems (ECM). We will try in Chapter 2 to develop this transition to digital further through the methods and strategies of digital archiving, while presenting the normative landscape governing this area.

¹² We will detail these new methods in Chapter 3, which is dedicated to archives in the age of digital humanities.