Digitize documents in the Archivo General de Indias: Technical advantage to preserve a historical legacy.

Antonio Sánchez de Mora, Ph. D.
Reference Service, Archivo General de Indias (MECD)

New Information and Communication Technologies (ICTs) offer many possibilities for archival practice.

- The automated management of information and the development of relational databases make easier the management of users and documentary funds.
- The rise of telecommunications has improved the dissemination and exchange of information, as well as the remote access to documents and their information.
- The digitization of documents has become an essential tool for the preservation and dissemination of documentary heritage. The availability of digital copies allows a restriction of queries of the original documents, without reducing the access to its information through descriptions and quality reproductions.

Archival science has adapted to everything that is offered by new technologies, provoking a revision of its theoretical principles and its practices.

- As a result, national and international standards have been developed, which have contributed to archival standardization and its participation in the design of new electronic documents and records. Among them, the digital copies and descriptions of historical documents.
- Archival criteria are present in the design of digital documents, claiming for the preservation of documentary information and its internal and external characteristics; the preservation of records’ metadata and the respect to the records’ context.

The Archivo General de Indias in the Age of Digital Technology

The Age of Digital Technology arrived in the Archivo General de Indias in the 80s.
ArchiDOC Project

As a result of the first advances in the computerization of archives, the Spanish Ministry of Culture considered the development of an innovative project that was intended to serve as a test for its later application in other state archives.

To sponsor this project, the Ministry of Culture, IBM Spain and the Ramón Areces Foundation signed a collaboration agreement in 1986, as part of the commemoration of the 500 years of the Discovery of America.

It involved the design and development of a computerized management system, capable of handling most of the archival functions. Specifically:

- The creation of an information and reference system that incorporated the arrangement of documentary resources and the existing instruments of description, improved through the services offered by the system itself.
- The indexing of information through a relational database.
- Creation of a system for digitizing and storing images of archival documents.
- The integration of previous services into a user management module, responsible for controlling the access to documents and services offered by the system.

El Archivo General de Indias was chosen to apply this project, because it preserves documents related to the discovery of America and the later Spanish presence there and in Asia during the 16th to 19th centuries. It is a documentary heritage unique in the world, significant for the history of many countries and peoples, and so is recognized by UNESCO.

First stage: Initial project, planning and implementation (1986 – 1992)

Archivists and computer technicians maintained a close relationship to understand and adapt to the new tool the actions carried out in the archive, in order to design the software.

- Since 1988, a local network for users’ management was established. Including the registry of users and their authorization for access to the Archive; description and research tools; requests for original documents; control of document movements; requests of information and reprography; and a module for statistics.
- The project included an analysis of the organization of documentary funds, descriptions available and existing research tools for the development of a Textual Database. The Management Database resided on the IBM AS/400 System.
- With all these information should be made a textual Database. Its structure was adapted to the hierarchical arrangement of the documentary funds, respecting its different levels and taking the signatures as lower levels of description. It was developed to the entire collection of the archive, not only to the digitized portion of it and included:
- Hierarchical arrangement of funds, sections and series, with their information.
- Converted information of previous catalogues and inventories, that were integrated in the hierarchical arrangement and adapted to the project standard.
- A new index, that related the textual information.
  - There should be different interfaces for access to the different databases.
  - The system should operate across a Local Area Network.

**Digitizing historical documents.**

After an analysis of the intended services and requirements were decided the resolution criteria for the images, and the hardware and software required for capture, storage and query were decided.

- Images should be of enough quality to facilitate reading and reproduction.
- Digitizing was done by scanning the original documents. As manuscripts, mostly in the 16th and 17th centuries, with problems of spots, faded inks, transparencies, etc., reproduction was chosen in 16 gray scales and 100 dpi (dots per inch).
- The image’s interface was completed with tools that allowed modifying the gray tones of the image, without altering the master copy.

After preliminary work, the software started to be construct, as well as the dump of information. A multidisciplinary team proceeded to gather available information. The archivists took part in the process of preparation of the information and later revision of the databases. IBM started the process of capturing images using 15 scanning stations. Xerox 7650 flatbed scanners were used, attached to IBM PS/2 computers, and the compressed images were stored on Panasonic 9347 optical disks that had a maximum capacity per disk of about 900 MB. There were scanned around 250,000 pages.

Preparatory work was conducted by archivists, who guided the technical team and prepared documents prior to scanning: Separating the manuscripts in the bundle as necessary; ordering them; creating an index document for subsequent entry into the Database, including the call number; and creating a sequential index that links the Database to the pages in the bundle / installation unit.

The following general criteria were adopted to undertake the digitization project:

- Conditions of manuscripts’ preservation.
- Degree of funds’ arrangement and description, because only were selected complete series.
- Historical relevance.
- Volume of queries.
A “scanning guide” or “control sheet” ensured that the right number of pages were scanned and in the right order, and in addition a quality control was the responsibility of two extra stations devoted to post-scanning verification.

It was designed an interface to enable researchers to display selected documents and to scroll or otherwise navigate through a bundle of documents, stored on an optical disk, that was previously mounted on the image server.

The scanning control information was used for referencing purposes and to provide researchers with a set of computational tools for enhancing sections of the displayed images in real-time tools. These enhancement tools used filtering techniques for increasing contrast, and removing stains and ink bleed through.

Second stage: Public dissemination and service opening (1992-1999)

In 1992 ArchiDOC system was presented to the public and started to be used in a Local Area Network. 9 million pages were already digitized, a 10 % of manuscripts, but a 30 % of queries.

There were about 60 workstations available, 20 of them located in the research room to consult databases and separate monitors for image views. The workstations consisted of IBM PS/2 Model 70's with two monitors attached: an IBM 8514 scrolling monitor for text display, and an IBM 8508 for grey-level display (1,200 x 1,600 pixels, 16 grey levels) of the scanned images.

The service of the optical disks that stored the images required their manual load by operators. One of the biggest advantages was that several researchers could access to the same document through the screens of their respective Computers. However, it was not decided to disseminate images and information on Internet.

Technological advances in the 90s allowed changes in the initial design: Storage and view of images in CD-ROM, software’s update and simplification of the search and unification of interfaces in a single monitor, that offered the query of images and textual information.

Third stage: Technological renewal. The end of ArchiDOC (2000 – 2007)

New technological advances influenced the renewal of software in 2000, when the OS/2 operating system, designed by IBM, was replaced by Windows 2000, by Microsoft. It was also changed the capture format of digital images, using the BMP and TIFF formats for the master copy and the JPEG for the query copy.

The need for greater storage capacity and the new performance of the magnetic disks caused the abandonment of the CD-ROM and the automation of the requests and the use of magnetic disks.

It was developed an extension of the textual database, with indexes of subjects, institutions, names and geographical references. The search for free text was also incorporated.
Finally, colour images of maps and plans were incorporated, something decided in the first project, but that could not be developed until 2000.

**Jump into Internet. The AER Project and the PARES**


In 2003 the Ministry of Culture decided to take a qualitative leap in the dissemination of documentary heritage. In addition, the Spanish archival authorities joined the reforms promoted by the International Council of Archives, aimed at standardization and adaptation to new technologies.

In 2003, the experience developed in the Archivo General de Indias, already partially introduced in other centers, was adapted to be consulted on the Internet, through the project *Archivos Españoles en Red* (Spanish Archives on the Net). The objectives of this project were to provide free access to Spanish national archives from anywhere in the World, and to modify the dissemination criteria, because it is thought to all the people and not only academics.

All text and image databases were integrated into software that, via the web portal, allowed remote access and, since 2005, printing and downloading of images and information.

*The web portal of Spanish Archives: Portal de los Archivos Españoles (2007 – 2017)*

PARES was the logical evolution of AER, with new applications as a result of technological innovations.

It constitutes a third stage in the adaptation of the state archives to the possibilities offered by the new technologies, after ArchiDOC and AER.

- It is an online database, formed by standardized multilevel descriptions that can incorporate digital images.
- It does not require pre-registration. It is done through the usual browsers, and does not need any additional software or application.
- The search can be done in a simple or advanced way, and incorporates hierarchical access to documentary funds through a *dynamic inventory*.
- It has a *user's agenda*, a research support tool that allows you to save searches and send images.
- It serves as a diffusion framework for other archival projects promoted by the Ministry of E, C and S.
- It has gradually implemented its applications, and provides on-line archival descriptions and user's management.
Impact of AER y PARES in the Archivo General de Indias

The international projection of the Archivo General de Indias its presence on the Internet through those software has increased the dissemination of documentary heritage, with an average of 1,500 annual queries per correspondence and millions of searches on PARES.

In spite of the slow and progressive increase of digital images on PARES, much work still remains to be done.

The diversity of users requires an adaptation of archival descriptions to a foreign public, that may be unfamiliar with the historical research. In fact, the textual information designed for the internal queries, which was incorporated into the ArchiDOC software in 1992, does not fully fit to the current archival regulations, nor do it always solve the doubts of an inexpert user of 2017. Descriptive projects have been developed to review and standardize information, in order to overcome those questions.

Far from reduce the queries, the Archivo General de Indias increases its users annually: Since 2000, an average of 850 new users per year. In-person research has not been reduced. Since 2000, an average of 800 in-person researchers come to the Archive, 500 of them as new user, and carry out around 9,000 work sessions per year.

The same may be said about the request for copies, that reaches around 400,000 pages per year.

Current Situation

Description’s projects associated with standardized digitization

Since 2007, PARES includes describing, indexing and associating digital images with informative records through a specific interface.

Archival standardization influenced the work of description that has been associated with digitization projects since 1990. In this sense, several projects have been implemented in the Archivo General de Indias, which have applied the norms established by the Ministry of Education, Culture and Sport:

- Standardization of the processes of description, revision and quality control of archival information (2010).
- Standards for the development of authorities and access points (2012).

Thousands of descriptive entries and associated images have been created or revised. Some progress has been made also in content indexing and the creation of standardized access points.
Today the Archive has 30% of its documents described and 20% of them with digital image, that sums a total amount of 7,500,000 images on PARES. There is an average of 50,000 to 70,000 new images per year since 2000.

The current parameters for the digitization of images are as follows:

- Master copy: 300 dpi; colour 24 bites; TIFF format. Hardware: Copibook i2S Digibook.
- PARES copy: 125 dpi; grey scale 8 bites; JPEG format / PDF for print.

All digitized document is available for research on PARES, while the original manuscript is removed from in-person query. Any reproduction is managed from the digital copy or downloaded from PARES, that it is free.

**Between austerity and chimeras: Optimization of resources and opportunities**

The interest of academic and cultural institutions for the reproduction of specific documentary series may result in mutual collaboration.

- There are cases in which the interest in obtaining copies from full series advises the development of a parallel description project, that will make the digitization process profitable.

- Some events occasionally causes the increase of queries and the requests for copies of some documents. In such cases, projects are designed to describe them, digitize them and load them on PARES.

**Non-standardized query copies. Does ends justify the means?**

Since June 2011 the Archive adapted the reprographic service to the realization of black and white digital copies. This measure answered to several arguments: Request from users, existence of technical possibilities, the microfilm copy was becoming obsolete, lower cost of mailing and environmental and economic criteria.

In the last years an average of 400,000 copies per year are request by users.

Digitization of single and not-described documents does not allow the subsequent use of the copy and, consequently, the query of the original is not avoid, nor its new reproduction after a new request.

Partial solutions to avoid this problem are:

- Non-standard copies of complete installation units are preserved.
- Since 2012, ca. 6 million pages of documents from the Archivo General de Indias, from 6,192 microfilm rolls, are available at the research room. These images are not standardized and described.
Conclusion

Digitizing historical records must be understood as a way of preserving original documents, as well as a way of disseminating them. In this question, archival criteria must be considered, in order to identify and contextualize each document, preserve it to future generations and ensure access to its information and image. In this process, standards are a necessary tool, but not the goal.

Bibliography


- Simó Rodríguez, Mª Isabel (2012): “Archivos hispanoamericanos y digitalización”. Instituto Cervantes. [Enlace web].