

Timbuktu Libraries Project



ARELMAT

Archivage Electronique des Manuscrits de Tombouctou

Project description ([en francais](#))

ELECTRONIC DOCUMENT MANAGEMENT (EDM)

(Archivage Electronique des Manuscrits de Tombouctou - ARELMAT)

1. Background

The most promising strategy for saving the intellectual and artistic content of endangered volumes is to transform or convert it to a different medium - by copying it to film or digital electronic form.

Joint Task Force on Text and Image, April 1992, Preserving the Illustrated Text.

Each year more documents maintained at IHERI-AB become brittle and unreadable and manuscripts are being handled without care or in the case of some private collections, sold to tourists. Through earlier projects, IHERI-AB has been furnished with some basic equipment. The photo and microfilm lab at IHERI-AB have several pieces of good quality equipment, but they are not being used, either because the technician does not know how to use them, or because they lack some spare part or supply items, or because they suffer from inappropriate technology syndrome.

The initiative for this component came from Mamadou Diallo Iam at the CNRST who was shocked at the state of affairs and proposed in 1993 to scan and/or photograph the manuscripts and develop an image database for consultation. Hard copies of the manuscripts could then be archived in a safe place.

This component, which has primarily been financed through a grant from the Ford Foundation, will handle the data technical portion of the project (including eventual microfiche options) for developing a comprehensive electronic document management system to make the collections known and facilitate research through organized access to electronic resources. Databases for pre-and post conservation images will be developed and managed within this unit as a support to the conservation team. A modest desktop publishing unit will be established within this component as support to the Research component.

A close collaboration with UNESCO's Information & Informatics Division has been established and the work prepared for UNESCO by the [National Library of Prague](#) will be considered for the selection of standards and digitisation techniques. Technical options in this relatively new field are changing too rapidly to detail in a proposal seeking long-term funding. Current options presented in a separate document entitled Technical Requirements

and Options will lay the groundwork for future choices.

2. Objectives

The overall objective of this component is to develop a comprehensive electronic document management system for the IHERI-AB collection in Timbuktu which would qualify IHERI-AB as a service provider for private collectors. Preserving images of texts which are important to scholarship through the use of electronic technology for image capture, storage, retrieval, and networked access is the primary goal of this component. In general, the aim is to create conditions whereby researchers who wish to consult these manuscripts could do so without physically handling the individual items, while at the same time the artistic quality of the calligraphy and decoration of the manuscripts could be preserved for future generations of Malians to take pride in. It is expected that the experience gained and skills developed will increase national capacity to undertake EDM in other libraries such as the national archives.

3. Project Activities

3.1 Development of Data bases

Building upon already existing catalogues and databases, a relational database will be developed (first through MS Access and then migrated to a more professional system) and connected through a local network to facilitate future access and consultation of information on the manuscripts as well as the actual images.

- 1) Manuscript Image (la banque de données des images des manuscrits - BDIM)
- 2) Catalogue (la banque de données des catalogues des manuscrits - BPCM)
- 3) IHERI-AB Library (la banque de données de la bibliothèque du IHERI-AB - BDBC)
- 4) The Physical Restoration of Manuscripts (la banque de données des restauration des manuscrits - BDRM)

1) The Manuscript Image Data Base will only include the actual image of the document, (including all details such as illuminations, calligraphy, footnotes, marginal comments, etc...) and transfer them into a server devoted to the task of storage. Mirroring techniques will be used to avoid any loss of data later on. Images will be accessible from the catalogue database through a relational link. Access to the image will be non-obligatory for use of the catalogue base, given the size of the image base. The images can be stored in the proprietary database format, or in Web format (preferably jpg, or whichever format is most space efficient), depending on the database's display ability. Selected items of this databank will be available on the internet. A minimum of information (number, title, physical size & capture resolution).

2) The Catalogue Data Base is the overall manuscript catalogue which will include items listed in the Format of Descriptive Catalogue Entries developed by the research team. The documents will be catalogued in a fairly simple database application building upon and improving the already existing catalogues including as entries title, subject. Cataloguing in itself requires rigorous specialists in the field of Arabic documentation and interpretations of, for example, authorship should be carefully supervised by experienced researchers. The database application will be developed in-house on the basis of an off-the-shelf database management program which fulfills the following requirements:

- Relational, linking the four databases described above;
- Ability to properly handle Arabic script, for display, indexing etc.;
- Ability to store images or link externally to image files, locally or on external drives, discs or CD-ROMs;
- Ability to communicate with Web server software.

The assumption is that the database will run under Arabic Windows or Windows 2000 Professional, MultiLanguage Version. The catalogue database will have a mixture of Arabic and Latin text fields, so that the relevant fields can be searched in the form required. The catalogue will be based on the format developed for the provisional Furqan catalogue (see general introduction), with minor amendments as required, in particular the addition of Latin text fields. The data from that catalogue may then be typed or imported into the new catalogue.

To maintain confidentiality of some manuscripts photographed from private collections, only a selection of data contained in the BDCM will be accessible to the public on line.

3) The IHERI-AB Library Data Base is the general IHERI-AB's modern library catalogue which provides archival information on each book in the IHERI-AB library; information which is presently recorded manually. The BDBC will also contain bibliographical data for research and academic purposes. Data will be freely available to all, in Arabic, English and French. Comments and scientific findings on the manuscripts could also be accessed through this database.

4) The Physical Restoration of Manuscripts Data Base is a survey of restoration processes and will document the process of restoring each document. In order to maintain the restoration history of the manuscripts, each document will be photographed first before any physical restoration of the manuscript and secondly following physical restoration. These images will be linked record to record with the catalogue and it will be possible to search individual digitally photographed pre-restoration images of the manuscripts stored in this database. The database will provide information of particular interest to the physical conservation team and other conservation specialist concerning paper & ink types, date of physical document, size, condition & restoration notes. Some of this information can be imported from the catalogue database. The BDRM will be of no interest to non-specialists, so its access will be restricted to the Conservation team dealing responsible for the physical restoration and conservation of the manuscripts. Pre and post restoration images and data will however be available on line.

3.2 Image Capture

The objective is to convert the manuscripts to a high quality electronic media as digital image (i.e. no text recognition). Each page from the manuscripts already collected by IHERI-AB (ca. 18,000) will be photographed before and after restoration and dynamically added to the computer-based image bases. The images from the pre-restoration stage will be stored for documentation of the conservation process. Post-restoration images will be the basis for exploitation of the documents for academic purposes.

Manuscripts will be scanned at high resolution, photographed at high resolution with a digital camera or photographed using classical silver film photo apparatus. The film will then be developed and scanned using film scanners. The precise number of pages to be photographed is not yet known (each manuscript containing between 1 and 500 pages); however, it is estimated that the entire IHERI-AB collection contains a total of about 700,000 images. Pre and post restoration images will be stored in an appropriate back-up storage systems with a total capacity of about 1000 GB. Since manuscripts differ in size, quality, age, and historical & esthetic value, the documents will be grouped according to type document.

It is estimated that the IHERI-AB collection contains less than 6% of the manuscripts from the Timbuktu region, the rest are in private hands. The digitization team will approach the private owners and request that their collection be photographed in exchange of being trained by the conservation team in conservation and restoration techniques. Ideally, all pages of all manuscripts in the entire region would be digitized. This would take a decade

and would depend on the private collectors' willingness to have their collections photographed. A selection would be made during in-depth surveys to determine priorities based on the following criteria, which are again based on IHERI-AB's prospecting policy:

- local production
- high esthetic value (illuminations) including unique calligraphy
- rarity of the manuscript
- trade & legal documents
- age of the document
- pre & post images for restoration data base (determined through restoration survey)

3.3 Storage & Back-up

The resulting post-restoration image files will be stored in TIFF format and a copy of the images compressed for consultation purposes. The size of each image file of the photographed page will vary depending on precise selection of software, image type, and resolution. The images should be readily available both for producing new copies and for access on demand potentially over the internet. Appropriate storage equipment will be selected from current market options.

In addition to securing the individual image files, the entire data base system will require a reliable back-up system in case of breakdown. Back-up duplications of the image files will be copied and stored in a system of mirrored RAID hard disk storage for the short term protected in .special fire and water resistant cabinets at IHERI-AB (with an extra copy in Bamako.)

One or more sets of microfilm copies of the images can eventually also be produced from the image files.

During the registration, cataloguing and preservation process, image files and catalogues will be stored on removable drives or RAID system..

3.4. Image Retrieval and Consultation Facilities

From the consultation room at IHERI-AB, a client should be able to easily download an entire manuscript for reading or other forms of exploitation (in anticipation of an intranet based solution, ZIP drives will be used for in-house consultation. The system should allow a client first to consult a single page before downloading the entire manuscript. Note that this process refers to images, not recognized text. Electronical text recognition will not be attempted given the significant variation in the calligraphy of the individual authors of the manuscripts. Electronic images will in some cases be accompanied by new editions of the Arabic text, transcripts or translations in book form.

In addition to local consultation, access through internet or intranet might be established with a secure management system for maintenance and access.

For later consultation, three options should be available: permanent storage on a dedicated computer, distribution to intermittent users (sales) or data/image base distributed (cost-free) over the internet (See Technical Requirements and Options annexed). A website for the Timbuktu manuscripts will be developed potentially providing the following services either free of charge or at cost:

- 1) access to IHERI-AB library catalogue
- 2) access to catalog of manuscripts from the region
- 3) full-text publications of articles translations, transcriptions, and eventually the journal Sankore (see Research Component)
- 4) direct order of CDs with photographic images of manuscripts and possibly direct

download upon request

3.5 Dissemination

The project team will document experiences from work undertaken within this component and produce a bulletin on project activities, findings and outcomes for those interested in developing similar ventures.

5. Resources

Human Resources

The core of this Working Group is presently engaged by the CNRST, which has experience in desktop publishing, databank development and digitizing. This working group will perform manuscript digitisation, transfer and storage of images, database development, implementation of a LAN and its link to Internet through a Web site, training of IHERI-AB staff in computing, desktop publishing, and database management essential to the project operations.

The core group consists of the following members:

Modibo Haïdara Group Leader, CNRST

Abdelaziz Abid UNESCO

Knut Vikør, Catalog database development Univ.of Bergen, Norway

Cheickna Kondé, Desktop Publishing and Database, CNRST

Albrecht Hofheinz, University of Oslo

Alphadoulou Abdoulaye (Alpha Dicko) IHERI-AB, Timbuktu

Darahamane Salaha IHERI-AB, Timbuktu

Consultants in EDM can be very expensive and several low-cost options should be investigated. One option would be to attract a Malian EDM expert working at a university in the US to be part of the program currently installed in Mali to draw upon the expertise of expatriate Malians. Malians obtain release-time with salary from their University to give courses at the University of Mali. Expertise could be obtained through Malinet.

A data technician will provide support to researchers for lay-out and graphics presentation for submission of articles to journals (including Sankore). Mock-up and printing for Sankore and other electronically generated publications will be undertaken either at IHERI-AB, CNRST, at a local service provider or at a service provider in Bamako.

6. Risks & Problem areas

6.1 Technical obsolescence

As stated by the Task Force on Archiving of Digital Information for The Commission on Preservation and Access and The Research Libraries Group:

Continued access indefinitely into the future of records stored in digital electronic form cannot under present circumstances be guaranteed within acceptable limits. Although loss of data associated with deterioration of storage media is an important consideration, the main issue is that software and hardware technology becomes rapidly obsolescent. Storage media become obsolete as do devices capable of reading such media; and old formats and standards give way to newer formats and standards.

For the Electronic Document Management component, obsolescence of and damage to electronic storage facilities in the long run is a high risk requiring migration to new systems. This would be a natural responsibility for UNESCO.

6.2 Consultation of Manuscripts

During the pilot project, information was obtained on the level and reliability of interface technology available through service providers in Timbuktu. It became evident that internet facilities are too weak to encompass downloading of manuscripts through internet. In consultation with UNESCO, it was decided that satellite facilities were needed to run the project website from Timbuktu. Malians will receive extensive training in website design and management.