HELLENIC LITERARY AND HISTORICAL ARCHIVE (ELIA)
Digitizing our collections – from matter to pixels

Vassiliki Hatzigeorgiou, thilde Pyrli
Photographic department of ELIA

The Hellenic Literary and Historical Archive (ELIA), is a cultural non-profit organization situated in Athens (Greece). Its primary aims are the collection, preservation, classification and study of printed and generally archival material of the 19th and 20th centuries. Thus, twenty six years after its foundation in 1980, ELIA holds more than 800 hundred archives of historical and literary content, approximately 300.000 images, as well as the third largest library in Greece with volumes of the above mentioned periods. All collections are open to the public for consultation and research.

The digitization project – facts and aims

The digitization project is funded by the Information Society Operational Program, within the framework of Measure 1.3 "Culture", Invitation 65. ELIA was granted the sum of € 1.000.000 for the realization of its proposal: "Digitization, Development and Projection of the Cultural Collections of the Hellenic Literary and Historical Archive".
The project began on September 2004 and is expected to finish on January 2007. It involves the digitization of 84.500 items: 48.000 images and 10.000 post-cards (Photographic Archive), 12.300 items of the Theatrical and Music Collection (theatrical programs, cinema posters, music scores), 11.200 items (journals, newspapers, calendars) of the Newspaper Collection and 3.000 ephemera.

1. Project structure

The project consists of four sub-projects:
Sub-project 1. (Realized and supervised by ELIA)
- ?? Cataloguing of collections (84.500 items)
- ?? Publication of a multilingual database with the title "Photographers – Greek and Foreigners – who worked in Greece during the 19th and 20th centuries (1839-1960)
- ?? Development of a multilingual internet site (portal) with the title "Everyday life in Greece through the collections of ELIA (1830-1960)"

Sub-project 2. (Realized by contractor)
- ?? Supply of hardware(working stations for cataloguers) and software (database)
- ?? Digitization
- ?? Production of the digital matrix (CD) for the publication "Photographers – Greek and Foreigners – who worked in Greece during the 19th and 20th centuries (1839-1960)
- ?? Set-up and support of the portal (search engine) that will host ELIA's on-line database along with the 84.500 images (url:)

Sub-project 3. (Realized by contractor)
- ?? Translation of texts into English and Spanish

Sub-project 4.
- ?? Projection of the results (organization of conference)

2. Preparation: thorough inspection of one's collections

Overall inspection and command of one's collection is a key feature, a prerequisite for the analytical and thus successful planning of a digitization project. We need not stress the importance of the exact knowledge of one's holdings – or as close to exact as possible – since a detailed account of the number of items, the substrate types (negative, print) and the formats will be needed. In the case of the Photographic Archive, our chief archivist along with the head of the department went through the approximately 300.000 items of the collection for the sole purpose of answering to the *What*, *How Much* and *How* of the project.

What: substrate types and formats to be handled (glass plates, films, slides, print?, b/w or color formats)
How much: impractical as it may seem to put down the precise amount of items, it is absolutely necessary to give a pretty tight estimation. We have to bear in mind that correct planning of the cost, time and means required for the project depends heavily on this estimation. The more accurate it is, the smaller the loss of time and money.
How: means of packaging and ways of moving the items to be digitized. Our prints are filed in folders, in passepartouts, in bound albums. Films are in strips of two or three images, but the vast majority of them are cut in single items. The preservation status of the material must also be taken into account and problems such as curling, mirroring, cracks on glass plates need to be considered when estimating the time and cost of preparation for scanning.

3. Classification and cataloguing: it goes before scanning

It would be futile to digitize one's holdings before classifying and cataloguing the material, especially when large numbers are concerned. A huge amount of digital files named after some serial number with no reference at all to the scanned material will no doubt drive archivists in search of the prototype to desperation. A unique reference number for each item that would also be the name of the digital file (that is, its digital copy) constitutes a safe way to ensure the successful "connection" of database entry and digital image.
As regards the database, one should always keep in mind that the aim of any similar project is to share information with others, that is, publication via internet. Therefore, the database used must conform to certain widely acknowledged standards, such as Unimarc, ISAD-G (for cataloguing) and Dublin Core (for metadata exchange).
In the case of ELIA, our chief archivist and the head of each collection established the preferred cataloguing fields according to the above-mentioned standards. The first entries were made in a Microsoft Access® database, where a separate cataloguing form with appended thesauri was devised for each collection. Our current database software, LibSolution®, is compatible with Unimarc, ISAD-G, and ISO 2709 and is being used for both cataloguing purposes and the digitization process. We should like to note here that we consulted SEPIADES but did not use it since it does not support the Greek language. The selection of the material to be included in the digitization project and its classification took, for the photographic archive, almost one year. This preparation took place before the commencement of the actual project and it included first aid treatment to torn prints and cleaning of glass plates.

4. Digitization: modes, equipment, rates, results

The actual digitization process is taking place within the premises of ELIA, and more specifically, in the ground floor of our building.. The equipment used was brought by the company and will be removed at the completion of the project. It included three flatbed scanners that would cover all possible substrate types. Specifically:
One Microtek i900 scanner used for glass plates, one Microtek ArtixScan 120tf scanner used for films and one A2 sized Widetek scanner for large formats (prints).
Defining the digitization parameters was the first step of the process. We consulted photographers, specialists from the National Technical University of Athens as well as the standards proposed by the Information Society Operational Program and we concluded to produce three digital files for each item:

- a master file TIFF (4000 pixels in the largest dimension, 16 million colours)
- a projection file jpeg (150 dpi resolution, ~600 pixels in the largest dimension)
- a thumbnail jpeg (72 dpi resolution, ~150 pixels in the largest dimension)

As far as the negatives are concerned we decided two things: a) that their master files should be in the negative form and therefore we inverted only the jpeg files and b) that glass plates should be scanned in RGB mode but films in Grayscale mode. This latter distinction owes to the fact that glass plates are never really black and white and we wanted to keep that extra colour information. For films such need does not arise, thus for economy reasons (economy of space in the servers, time) we went ahead with Grayscale mode.
The material for scanning is prepared and moved to the “digitization lab” once a week and returns 10 days later. We send material (both electronically via the software and physically) every Friday and receive scanned material every Monday along with a series of the digital master files in DVDs. These latter serve a double end: on the one hand, they are used for the quality control of the digital files (checking the naming of the file 100%, and of digital image - sampling), and on the other hand as an extra means of security (cannot afford to keep master files only in one place).
At the end of each week, the digital files are moved at the storage units which are: two (2) database servers in a cluster connected to an array of discs with an overall capacity of 13TB and a backup unit. There is also one web server for the manipulation of the digital images in our portal.

Conclusions

The most important conclusion deriving from our experience is that no collection can be successfully digitized if it has not been previously classified and catalogued in an electronic database. Even in cases, the most common we feel, that this procedure has been completed for only a part of one’s collections, care should be taken that the material to be scanned will be given a priority over others. Bearing in mind that each item selected for digitization must have a unique reference number, we suggest that cataloguing takes place in item-level and not in group-level.

Selecting the items to be digitized is, without a doubt, the primary step to such a project and at the same time a greatly painstaking experience. In our case, we selected 48.000 items from a collection of 300.000, or the collection’s 16% (a figure shaped by our budget). So how does one pick this image and not another? Although this is an extremely complex matter and cannot be properly discussed within the limits of this presentation, we would like to note the person(s) deciding should have a

thorough knowledge of the collection in hand, or figuratively speaking, to know the collection by heart. And even more, to be able to foresee the needs of researchers. We feel that it would also help considerably if the institution housing the collection has a certain acquisition policy as well as an exhibition policy because these latter two define the function and content of one's holdings. In our case, we lacked both and had to invent or improvise. For the material selection, two persons worked for one year; the selected material (in archival boxes) was flagged with special adhesive signs that gave a visual impression of the selection.
In a nutshell, we could say that we gave priority to 19th century material and that we decided not to scan material created after 1985. As far as the substrate type is concerned we prioritized glass plates and (unstable) diacetates and nitrates as well as bound albums.

Last but not least, we discovered that in digitization it is not necessarily the machine that makes the difference but the person handling the machine. A state of the art scanner can produce mediocre results if the person handling it is not experienced enough both in the fields of photography and scanning.