

## **COST EVALUATION OF DIGITAL IMAGES: THE SPANISH EXPERIENCE\*\***

**By Javier Docampo**

"The cost of digitization is the biggest single challenge to achieving our goal"

Suzanne Thorin, chief of Staff  
of the Library of Congress

**Abstract:** One of the main objectives of the Biblioteca Nacional (Spanish National Library) is to make its collections available in digital format. This to achieve the B. N. has initiated a most ambitious project called Memoria Hispanica. The first phase - taking place between 1996 and 1998 - will involve the digitization of about 50.000 documents of all kinds: books, manuscripts, graphic material etc. In this field the B. N. can already look back to a whole host of experience of which most important is its experience gained in conjunction with the Heraldic data base. This can serve as an excellent example for carrying out a cost and efficiency evaluation of the digital image services rendered to the library users. The article analyses the effect costs and prices for digital services have had on their use in order to draw valuable conclusions for further on-going projects.

### **1. Memoria Hispánica: the importance of digitization in the preservation and diffusion of Bibliographic Heritage.**

The growing development and progressive cheapening of the digitizing technologies have spread out all over the world the projects for the storage and diffusion of information in digital format. The Spanish National Library, as the main repository of information of the Spanish world, could not remain outside of these new challenges. In this context the Memoria Hispánica project is being developed. Its aim is the creation of the Biblioteca Nacional Digital (Digital National Library), in which every document of the more than eight million items housed in the Library would be accessible in digital format through telematic

---

\*\* Paper presented at the 62<sup>nd</sup> IFLA Conference in Beijing, 25 - 31 August 1996 in the Workshop of the Section Art Libraries: "Pay or Profit: Fee or free"

networks<sup>1</sup>. This project will allow a permanent solution to the dilemma that can be found all over the Library's history: how to join the preservation of the collections with access to them.

Logically this project only can be a long-term one. The first phase will develop between 1996 and 1998 and it will consist of the digitization of about fifty thousand items selected by three criteria: bibliographic importance, frequency of use and bad state of conservation. The cost calculation is 1.000.000.000 pts. (about 8.000.000 \$) from sources both public and private in a ratio yet to be determined<sup>2</sup>.

Much has been written about digitizing technologies, mostly about purely technical matters. But the problem of the costs of these new services and the influence of their prices on its use has been, in comparison, less discussed<sup>3</sup>.

## **2. First experiences: evaluation of the Heraldic Database.**

The first experiences in digitization in the Biblioteca Nacional took place in 1993. The most important one in image digitization was the implementation of the Sistema Integrado de Información Heráldica (Integrated System of Heraldic Information)<sup>4</sup>. Its aim was to satisfy one of the more usual questions of our readers, both researchers and ordinary readers: to know the origin of a surname and the connected heraldic device. For this purpose the library began digitizing the best known and important work about Heraldics in Spanish: the Diccionario heráldico y genealógico de apellidos españoles y americanos by Alberto and Arturo García Carraffa, published between 1952 and 1958 in eighty-six volumes.

---

<sup>1</sup> There are some similar projects in the world but the most important belong to the Library of Congress: American Memory Project and The National Digital Library Project. See about them: „The National Digital Library“, in: Information Retrieval & Library Automation, vol. 30, n. 5 (1994), pp. 1-3 and the vol. 53, n. 20 (1994) of Library of Congress Information Bulletin, issue devoted to the digitization projects of the Library of Congress.

<sup>2</sup> It should be noted that the first phase of The National Digital Library Project, estimated in five years, has been carried out with a budget of 10 million dollars

<sup>3</sup> See the bibliographic report by Javed Mostafa „Digital image representation and access“, in: Annual Review of Information Science and Technology, vol. 29, 1994, pp. 91-135. The best study about the costs of the new digital libraries appeared last year: Saffady, William „Digital library concepts and technologies for the management of library collections: an analysis of methods and costs“, in : Library Technology Reports, 1995, n. 3.

<sup>4</sup> More information about this project in: „El Sistema Integrado de Información Heráldica: digitalización de imágenes“, in: Biblioteca Nacional: revista de comunicación interna, 1994, n. 4, pp. 12-13 and Xavier Agenjo and Francisca Hernández: „La digitalización de materiales bibliotecarios en la Biblioteca Nacional“, in: Boletín de la Asociación Española de Archiveros, Bibliotecarios y Documentalistas, 1995, n° 3, p. 85.

After evaluating the offers of fourteen companies the Library chose as the best one that of Idea Informática. The first phase was the implementation of the digitized image database, the cost of which for the Library was 15.000.000 pts (130.000 \$). Afterwards the same firm created the text database at a cost of 5.000.000 pts (40.000 \$). Both amounts were from the Library's budget. The first experimental model was exhibited for the first time and with considerable success in the National Library pavillion in the Fair joined to the Barcelona IFLA Conference in August 1993. In June 1994 it was carried out in the Biblioteca Nacional. Since the beginning the database was considered as a primary service, so the users have no need for another card or requirement to subscribe to it.

The database stores the images of 14.000 heraldic devices and 25.000 text pages that describe the origins and history of the correspondent surnames. The search strategy is simple, because the only possible access is by surnames. From each one it is possible to obtain both the image and the text. The user can get as well a black and white hardcopy of the text and a colour one of the devices images.

The statistics in 1995 shows that the demand is more focused in the texts (2070) than in the images (333). The reasons may be the imperfect quality of the images. This difference is more pronounced if we take into account the number of hardcopies: 654 text pages and 78 devices images. In this case the reason is clearly the price. A text hardcopy costs 10 pts (0,08 \$) and a device image hardcopy costs 1.000 pts (8 \$), a very high price for a not hardcopy of insufficient quality.

The future development of this database has to focus on three aspects. First of all it is necessary to improve the flexibility of information retrieval. It would be very interesting to develop algorithms that get a more exact description of the heraldic devices, considering that Heraldics is a very accurate science in their definitions.<sup>5</sup>

It would be important also to enter more information trough the digitization of other reference works that fill the gaps of the García Carraffa work. Finally the library has to increase the number of users by distributing it on CD-ROM or by allowing remote access, via Internet or another network.

---

<sup>5</sup> The devices description system is very interesting, useful as well for other kinds of materials, proposed by Harold E. Thiele: "Heraldry and blazon: a graphic-based information language", in: Library trends, vol. 38, n. 4, 1990, pp. 717-36. A similar experience to ours is the HISTORIA: Heraldic Images STORing Applications project, carried out by the Westminster University and the Biblioteca Nazionale Marciana, that is developing a database with images from venetian heraldic manuscripts.

### 3. Ongoing projects: the digitization of Iconografía Hispana and the CD-ROM Obra gráfica de Goya.

The National Library and more specifically its Prints and Drawings Section has in this moment two projects of image digitization, different both in scope and goal, that show clearly two possible ways to finance this kind of project.

Before discussing the two projects we must gain a brief understanding of how Spanish library services are currently financed. In Spain, most libraries and a great number of documentation centers belong to public institutions of different rank (locals, regionals, state...). There are few private companies that offer library services (Savings Banks or some educational foundations are the most significant) nor are documentation centers usual within private firms.<sup>6</sup>

These circumstances led Spanish libraries to a specific practice in which safe financing, though always insufficient, and its consideration of public service made them give up external financial sources. This caused an underemployment of their resources and a fall in demand. The situation has changed in the last few years. The economic development and a greater interest in Spanish society for cultural affairs has produced growing private financing of library services.

Therefore we reach the two digitization projects that are going to attract our attention. Three years ago we announced the implementation of an optical disk containing the Spanish portraits collection of the library, known as Iconografía Hispana by the repertory in which they are described<sup>7</sup>. The project has been much delayed but now it is almost completed.

The project was charged to Instituto Histórico Tavera, a cultural institution belonging to the insurance company Mapfre. This institution has a specific section for digitization projects that has had some previous experiences. This institution is creating a CD-ROM that contains about 20.000 images and the related records that describe them. Their goals are both preservation, since they are to a great extent engraved portraits of the XV-XIX<sup>th</sup> centuries, and improvement of information retrieval.

---

<sup>6</sup> A review of the financing of Spanish libraries can be found in: Peón Pérez, Jaime Luis "Principios de carácter económico para la organización y planificación de bibliotecas y centros de documentación", in: Boletín de la Asociación Española de Archiveros, Bibliotecarios, Museólogos y Documentalistas, vol. XLI, n. 1 (1991), pp. 53-59.

<sup>7</sup> Docampo, Javier and Colodrón, Victoriano "Automatización de fondos de material gráfico: la experiencia de la Biblioteca Nacional" in: Bibliotecas de arte, arquitectura y diseño: perspectivas actuales. Actas del Congreso organizado por la Sección de Bibliotecas de Arte de la IFLA, el Grup de Bibliotecaris d'Art de Catalunya y el Museu Nacional d'Art de Catalunya, München: Saur, 1995, p. 94.

The printed texts have been converted into ANSI text and afterwards they have been processed using optical character recognition (OCR) software. The result needs later correction, including the manual introduction of the code of the location in the library. The final product will be a database with some basic retrieval fields: sitter name and profession, biographical dates, artist, description and source.

The images are scanned in grey scale and with a resolution of 600 dpi that gives sufficient quality for the goals of the project and a storage capacity of 4.000-6.000 images in each CD-ROM. It is possible to enhance this capacity by introducing compression algorithms, so that all the work fits into two or three disks. Finally the implementation of an user interface in Windows will complete the project.

The cost of the project will reach ca. 5.000.000-5.500.000 pts (about 40.000-45.000 \$) paid entirely by the Instituto Histórico Tavera. About 500 disks will be issued, 400 for the Instituto and 100 for the Library, which it is not permitted to sell but may offer it in the Internet. The Instituto is going to sell the disks at a price of about 30.000-40.000 pts (ca. 250-350 \$), a very competitive price in relation to the printed repertory, if we consider the improvement of information access and the fact that every image is reproduced.<sup>8</sup>

The second project on which we are working now is the first CD-ROM created from an exhibition held in the Library. Spain celebrates in 1996 the 250th anniversary of the birth of one of the most important Spanish painters: Francisco de Goya. In light of this the Biblioteca Nacional is holding an exhibition with part of its great collection of Goya's prints and drawings, one of the best in the world. As a consequence of cataloguing the works it was thought possible to produce of a CD-ROM. With this purpose the Biblioteca Nacional contacted the company Hobbypress, publisher of the magazine PCMANÍA, which is devoted to the PC world and includes in each issue one or some CD-ROMs about various topics. This firm has created the disk from the material given by the Library. The images have been scanned with a Scanview device (Scanmate 5000) running with software Colorquartet 3.3.1 and with 300 dpi resolution.

On the disk can be found about four hundred images, mostly Goya's prints, plus some images of drawings and pictures. Each of these images has a text record that describes it. The CD-ROM also contains fifty pages of text, music (five minutes)

---

<sup>8</sup> See the interesting analysis done by Jennifer Rowley and David Butcher: "A comparison of pricing strategies for bibliographical databases on CDROM and equivalent printed products", in: The Electronic Library, vol. 12, n. 3, june 1994.

and moving pictures about the exhibition, the Biblioteca Nacional and print techniques.

The master cost is about 2-3 million pesetas (16.500-20.000 \$), whereas the duplication cost is about 10-12 million pesetas (82.000-100.000 \$). The most interesting aspect of the project is the massive distribution of the product. The edition of the magazine is between 80.000 and 100.000 issues, with a price of 1.300 pts (about 10 \$), and two hundred thousand readers. The Biblioteca Nacional receives as well 300 disks and the ownership of the master, with the engagement that it will not exploit the product commercially. The result is a work less ambitious than the disks of Iconografía Hispana and focused on a different market, much broader but less specialized.

As we have seen the funding of both works has had a similar basis. The Library has neither the structure nor enough funds to afford these kinds of projects. We have therefore had to resort to the private sector, a cultural institution and a publishing firm, and they have funded entirely both projects. The Library gives the reproductions of their collections, the text records and, in the case of Goya's Graphic Work CD-ROM, the introductory texts to each section. Besides the external distribution through the sale of the CD-ROM the Library is preparing a new Study Room for the users of the Prints and Drawings Section, in which they will be able to view the originals, when their use is justified, and the reproductions both photographic (open cards, photographs) and digital. Every new service is going to be free and only the reproductions requested by the users will have a charge intended to cover the expenses but not to obtain a direct economic profit.

#### 4. Conclusions

The ambitious project Memoria Hispánica can only be completed by a process of successive specific programs coordinated with the same goal: the creation of the Digital National Library<sup>9</sup>. The funding of these projects in countries like Spain, in which most of the libraries and the documentation centers belong to public institutions, has to come from private hands, both cultural institutions without profit goal and private companies. The former source is appropriate to projects for the care and spread of the cultural heritage, and this heritage is often in libraries

---

<sup>9</sup> In the National Library the collections of the Prints and Drawings Section can be a privileged field for these plans. Future projects could be the next catalogue of German prints of the XV-XVI<sup>th</sup> centuries or a part of the vast photograph collection. About digitization of photograph collections see Benemann, William E. "Reference implications of digital technology in a library photograph collection" in: RSD: Reference Services Review, vol. 22, n. 4 (1994), p.. 45-50.

collections. Private firms have to answer the growing demand for digital products and therefore often need these historic collections for their activities.

If the prices of the new services are too high for the users because of the high costs for the Library it is likely that the use of these services will be low. On the other hand if we can encourage the companies to assume the whole cost in exchange for the possible profits of the venture, the resulting services will be cheap or even free to users, thus greatly increasing their use. The confluence of the budget possibilities and interests of the private firms with the needs of preservation and diffusion of the bibliographic heritage is the best way digitizing technologies can enter the libraries even in times of scarce budgets and without making more expensive the prices of the services.

Javier Docampo  
Biblioteca Nacional  
P. de Recoletos 20  
Madrid 28071  
Spain