



World Library and Information Congress: 70th IFLA General Conference and Council

22-27 August 2004
Buenos Aires, Argentina

Programme: <http://www.ifla.org/IV/ifla70/prog04.htm>

Code Number: 057-E
Meeting: 76. Newspapers
Simultaneous Interpretation: -

Project: Digitization of the 19th Century Latin American Press
(Digitization of the Great Colombia Press: 1820-1830)
Final report, May 30th, 2004.

Alfonso Quintero
ABINIA Executive Secretariat, Caracas

1. Background

At the beginning of the nineties, after researching collections in various Latin American national libraries, the conclusion was drawn that a great part of nineteenth century Latin American press ran the risk of suffering irreparable damage or even disappearance due to the instability of its medium and to the existing limited preservation capabilities. For this reason its inventory and preservation and the spreading of the valuable information these publications contain became a priority for all member countries of the Latin American National Libraries Association (ABINIA,) as well as for UNESCO, which included 'Memory of Latin America' as one of its World Memory Pilot Projects. This situation was address through the 'Nineteenth Century Press Rescue' project, set up by the National Library of Venezuela and later approved by ABINIA. Three phases were established for its development, the last of which gave rise to the present project.

Phases of the 'Nineteenth Century Press Rescue' project:

1. Creation of a catalogue collecting existing information on nineteenth century press in Latin American national libraries so as to create a data base for researchers to use.

2. Training on a regional level in microfilming in order to assure the reproduction of the nineteenth century press and the preservation of library collections in this form.
3. Indexing and digitization of the most representative newspapers and the editing of a CD-ROM in order to facilitate access to available information for researchers from other countries.

The first two phases were completed successfully. An inventory of nineteenth century press existing in Latin American libraries was made, in which the evaluation of the state of conservation of the newspaper collections were emphasized, as well as the computerized cataloguing processes and microfilming. As a result, a computerized collective catalogue of nineteenth century newspapers was developed and can be found on microfilm in the national libraries and other Latin American institutions; and, amongst other training activities, the Venezuela National Library set up in 1996 a microfilm workshop for national library staff in six Latin American countries.

With UNESCO contributions, the CD-ROM 'Nineteenth Century Latin American Press' was edited, which contains a data base with 8,990 bibliographic entries. This data base facilitates the retrieval of titles from any of the national libraries or other institutions within the fourteen participant countries. This same CD-ROM gives access to the first digitized images- corresponding to a period of one year's circulation- of the most important newspaper from each of the fourteen countries. This CD-ROM became the first product of the project.

2. Description:

2.1. Objective:

To digitize the most important nineteenth century Latin American newspapers so as to facilitate access for the research community to their content by means of a CD-ROM or on-line through the Internet.

Initially, (just as the Andrew W. Mellon Foundation agreed in 1997) it was decided to edit the 'Grandcolombian Press, 1820-1830' CD-ROM, which contains the images from one representative newspaper from each of the countries (Venezuela, Colombia, Ecuador, Panama, Peru and Bolivia) which participated in the political integration or confederation of the Latin American countries experience.

2.2.Expected results:

The edition of a CD-ROM which includes digital images of the newspapers, a revision of each title, exhaustive name, geographic and subject indexes to facilitate the retrieval of the information.

2.3. Institutional Coordinator:

Aurelio Álvarez Juan, Director of Library Technical Services, the National Library of Venezuela.

2.4. Technical Coordinator:

Pía Rodríguez, Director of the National Paper Conservation Center of the Venezuela National Library.

2.5.External Financing Institutions:

Andrew W. Mellon Foundation.

2.6.Approval Date:

December 1996.

2.7. Donation:

Andrew W. Mellon Foundation: Bs.(Bolivian currency) 95,150,000,00 (\$US 200,000) received on February 26th 1997 by the National Library of Venezuela.

3. Developments:

3.1.Acquisition of equipment and software:

Digitization Unit: In 1997, the Computing Division of the Venezuela National Library (SAIBIN) identified and selected the equipment and software required for digitization by direct microfilm scanning. The company 'RAM-ROM Computer' was chosen, representative for 'Houston Fearless' in Venezuela, to supply a scanner for 16 and 35mm microfilms, 'MEKEL M500 (SACN MASTER)', adapted to a computer with a 'SCANNER SEAPORT model 1D20'; 'Paint Pro Shop and Imaging' software for image manipulation and a 'CD-Recorder'. This equipment arrived in September 1997.

The setting up of the scanner took more than a year. The problems that this equipment had were not resolved by the company RAM-ROM, who implemented the recommendations supplied by 'Houston Fearless'. Thanks to the direct help from the 'MEKEL Engineering' technical representative and the expertise of our digitization operator, the adequate running of this equipment was achieved in January 1999. All the Venezuelan press digitized up until that date had to be repeated.

Workstations: In 1999 two additional workstations were acquired in order to populate the data base, control the quality of existing data and to clean the images; a 'JAZZ (2GB)' unit, to back up the images; and 'SCAV 2.5' software, for the creation of the data base and the production of the CD-ROM and 'Adobe Photoshop'.

In October 2002 a decision was made to buy an additional computer to make easier the processes of entering these records, image manipulation and quality control in order to meet the Andrew W. Mellon Foundation requisite of completing the project by the end of April 2003. Nevertheless, this purchase was delayed until September 2003 due to the commercial suppliers giving very short term estimates making it difficult to follow through the acquisition processes. And although at that time the official exchange rate was 1,600 Bs(Venezuelan currency) to the U.S.\$, importers complained that that was not competitive enough for them and that they had to pay for transactions at the unofficial and fluctuating rate existing in the market which was much higher than the official one.

3.2. Training and hiring of human resources

Training. In 1997, staff from the field of information science were trained by a 'Houston Fearless' systems engineer. The tasks relating to the digitization process (configuration, testing) were completed in advance by an operator from the National Center for Paper Conservation, Micrograph Division, who, together with six other civil servants, was trained in using the equipment by computing division staff (SAIBIN).

In September and November 1999, The Director of Library Technical Services attended two digitization meetings in the United States which allowed the revision of the structural organization and the activities carried out within the framework of the project: the first from 21st to 23rd September, *Preservation Options in the Digital World: microfilming and scanning*, presented by the Northeast Document Conservation Center in Nebraska and the second, from 2nd to 3rd November, *Digital Images for Libraries and Archives*, given by the University of Cornell, New York.

At the same time, from the 5th to 12th December 1999, the person in charge of the digitization process from the National Center for Paper Conservation carried out a study at the National Center for Editing Compact Discs (CENEDIC) at the University of Colima, Mexico, to observe the complete process: from digitization to the edition of a CD-ROM.

Hiring of human resources. In 1999 two historians were hired in order to compile the subject, name and geographic indexes. This job involved reviewing page by page each of the selected titles. Between June and November 1999 a transcriber was hired using project resources and who later became part of the permanent staff of Library's Digitization Unit until June 2001. In the year 2000 we could count on a graphic designer who dealt with the images in order to maximize their quality.

Due to the high turnover of staff (historians who worked independently and workers who were looking for permanent posts,) in the year 2000 three new civil servants for the National Library were authorized to work on the final stages of the projects activities: a historian for the indexing process of the remaining titles and two civil servants for the creating and quality control of the indexes.

3.3. Digitization and Indexing

At the suggestion of each country's National Library, the most representative Grandcolombian newspapers were selected and due to the limitations imposed by the CD-ROM capacity only five titles we chosen (see table no. 1) represented in 3,542 images.

Comprehensive indexing

The first decision, recommended by Mr. Hans Rutimann, adviser for the Mellon Foundation, was not to apply Optical Character Recognition (OCR) because none of the experimented software gave good results, due to the fact that the types used in Latin America at that time were very irregular. For this reason a decision was made to comprehensively index all the information contained in the five selected newspapers in order to achieve a retrieval of pertinent and complete information by means of name, geographic and subject indexes. This indexing process was carried out by historians specialized in that period and involved a meticulous reading of each page in order to extract the necessary information to complete 26 different fields in each record. This fact, together with the lack of sufficient historians specialized in this period, and their reduced availability to work continuously on the project -- most of the time there were only two historians and for long periods only one-- meant that this work was not completed until January 2003.

The indexing data entry into the data base was completed in March 2003 and the quality control of each record was finished in May. The quality control of the whole data base concluded in August 2003, due to the loss of information at the moment when records created at different workstations were merged in a single data base.

During the quality control process of scanned images the non-existence of some and the repetition of others was discovered. The subsequent substitution or inclusion of images was done after the repairing of the scanner in April 2003 (it had been damaged since April 2002 due to various technical problems.)

By-product:

Without doubt, the indexing process was much slower than the digitization. Therefore, the microfilmed press continued to be scanned and at present we have 63 titles, 12,335 images, corresponding to the period 1820-1830. For this simultaneous task the indexing process was simplified — by country, title, number and page- in order to facilitate retrieval of these titles. To this date the indexing of these titles has been completed. Also, the process of loading the records onto a server in order to allow access on line has been started.

Table No.1 Digitization of the Grandcolombian press (1820-1830).

Country	Title	Digitized Images	Indexed Images	Records Inserted
Bolivia	El Cóndor de Bolivia (1825-1828)	646	100%	100%
Colombia-Panamá	Gazeta de Colombia (1821-1826)	1,208	100%	100%
Ecuador	El Colombiano de Guayas (1828-1830) El Colombiano (1830)	278	100%	100%
		138	100%	100%
Perú	La Abeja Republicana (1822-1823)	753	100%	100%
Venezuela	Correo del Orinoco (1818-1822)	519	100%	100%
	TOTAL	3,542	100%	100%

3.4 Creation of the Grandcolombian Press (1820-1830) CD-ROM

In June 1999, the company Semi Crom, C.A., together with the Directors of Information Technology (SAIBIN) from the National Library, made a `demo´ with the digitized images from *El venezolano*, so as to evaluate the SCAV software.

Since the year 2000, various prototypes made by the company Archicentro,S.A., who developed SCAV software, were tested. The prototype includes the information from the data base corresponding to the Grandcolombian newspapers.

The last prototype was received in September 2003, without including the final corrections in the data base.

4. NEW DEVELOPMENTS:

At the time of the arrival of the new CD prototype (September 2003), the institution coordinator, as well as the project's technical coordinator and the person responsible for digitization at the National Library, for different reasons, had ceased working at the Venezuela National Library. Apart from this the National Center for Paper Conservation was beginning its move to a new site, within the new National Library Headquarters. Both of these situations brought with them new delays in the project. On the other hand, ABINIA's Executive Secretariat, complying with their designation as an entity sponsoring this project and conscious of the need to achieve its definitive completion, assumed a much tighter and more direct supervision of the project. With this in mind, we hired the new Director of the National Center for Paper Conservation at the National Library at the beginning of 2004 and we met with her and the technical staff who had worked on the project. We were able to establish, through common agreement, a work time schedule, which defined the pending activities along with their respective people in charge and finalization dates, up until the end of the project; that is to say, the edition of 1,000 copies of the CD-ROM containing the five previously selected and indexed titles of the Grandcolombian press-1820-1830 and its free of charge distribution to prime academic libraries in the United States, Latin American National Libraries and other libraries, museums and archives in Latin America and the Caribbean.

Technical Developments.- As a result of moving the scanner it was thrown off balance and a specialist had to be hired to get it working again. This was achieved by the middle of February 2004 with the support of ABINIA. At the same time the technical operator from the digitization team was able to continue his work. However, the most significant was to get a very competent professional in technical processes from the National Library to be in charge of revising the record indexing and, in most cases, to practically make new indexes.

Operational Developments.- Working with the company responsible for producing the CD-ROM we were able to maintain similar conditions to those established in 2003, for the technical revision of the prototype, edition and production of 1,000 copies of the CD.

Administrative Developments.- Support from new authorities at the Venezuela National Library and, in particular, from the Director of the National Center for Paper Conservation, as well as the careful following-up of pending developments by ABINIA's Executive Secretariat.

5. LIMITATIONS

5.1. Research Material and Contents:

Scarce availability of microforms in the Gandcolombian Countries.- From the beginning, with the help of Venezuelan historians and librarians from the national libraries of each country, the most important newspapers from the participant countries in the project were identified. However, few newspapers were already microfilmed. At the Venezuela National Library the microfilming was done from facsimile editions or duplicated rolls of microfilm positives that participating libraries sent.

5.2. Human Resources:

Handling new technologies.- This project was, the first of its kind, developed in Venezuela and became a learning process, as it developed, for all those involved. In fact, it served as a trial and

error process for the RAM-ROM Computer company and for the computer science science and conservation technicians from the National Library.

Highly specialized historians.- It became very difficult to find qualified staff to index 19th century newspapers. This type of information requires historians with a wide capacity to research historical figures from the time and achieve their correct identification. Generally speaking, these experts – so as to obtain enough income -- work autonomously for many different institutions at the same time and constantly look out for new projects. For this reason they do not undertake anything for a long period of time. At the best of times we could count on two historians working simultaneously on the project, but for the majority of the time there was only one.

5.3. Techniques:

As it happens with all new techniques, at first we are faced with the obvious problem of selecting and training adequate staff. Moreover, the microfilm scanner failed to work properly for long periods of time, the network configuration did not work (September 1997-January 1999, April 2002-April 2003), and both the supplier in Venezuela and the dealer took a long time to identify the causes and provide the technical support required.

5.4. Management:

At the beginning, the lack of experience in handling the new technology, the lack of knowledge concerning conditions in other National Libraries and the naïve academic aspiration to carry out exhaustive indexing (26 fields in each record) did not reflect reality and were the reason for the delay in carrying out the activities within the predicted terms and also in establishing the time scale for different activities.

5.5. Administrative:

The program was considered an external project (ABINIA), which made it difficult to obtain support from all of the technical and administrative departments of the National Library.

Executive Secretariat, ABINIA, Caracas, June 2004.