



World Library and Information Congress: 71th IFLA General Conference and Council

"Libraries - A voyage of discovery"

August 14th - 18th 2005, Oslo, Norway

Conference Programme: <http://www.ifla.org/IV/ifla71/Programme.htm>

juillet 6, 2005

Code Number:

141-E

Meeting:

97 Newspapers

Development of electronic periodicals at the Bibliothèque nationale de France : digitisation of French daily newspapers from Mid 19th Century to 1944

Pascal Sanz

Director of the department Droit, Economie, Politique,
Bibliothèque nationale de France
Paris, France

The Bibliothèque nationale de France (BnF) keeps a vast collection of French and foreign newspapers : end 2004 it includes some 60 000 retrospective or running titles. But it is of course the collection of French newspapers that the BnF considers its duty to preserve and to develop. Therefore it seemed quite natural as well as justified that the library should digitise a crucial selection of this important newspapers collection in order to give better access for researches in almost all kinds of social and human sciences : daily newspapers.

To introduce the digitisation programme I want to give a short presentation of the BnF's holdings of French newspapers : it will be the first part of my paper. In the second part I shall point out technical and scientific choices of the Newspaper Digitisation Programme. My conclusion will deal with future prospects of developing and enlarging this programme.

1. The French Newspaper Collection at the Bibliothèque nationale de France

There are different ways of classifying French newspapers (the INSEE newsprint list, the classification of the Bureau for Justification of Newsprint Circulation (OJD), etc.). Having this in mind one may divide the newsprint into some great categories :

- National newspapers comprise :

- political and general information newspapers including dailies (such as *Le Temps*, *Le Figaro* or *L'Humanité*) and weeklies
- illustrated people magazines with serial stories and strip cartoons such as *Le Petit Journal* launched by Moïse Millaud in 1863 at an issue price of 5 centimes
- gossip papers and satirical papers
- Regional and local newspapers : big regional dailies with several local editions (ex. : *Ouest France*), local weeklies
- Overseas newspapers from the former French colonies and territories (for instance *La Dépêche algérienne* or *L'Echo d'Oran*) as well as newspapers from the present DOM-TOM (French overseas administrative departments)
- Specialised newsprint including in particular women's magazines (ex : *Le Petit courrier des dames*), leisure magazines, newsprint on finances, law, sport, etc. (ex. : *L'Auto*)
- Reading newsprint including, among others, children's magazines, youngsters' illustrated periodicals (ex. : *La Semaine de Suzette*), newspapers for the leisure market, religious newsprint, associations' periodicals
- Free newspapers
- Government publications : official gazettes, local council bulletins, etc.
- Alternative press

The word press covers indeed a great variety of publications and newspapers holdings are to be found in all BnF's collection departments. However, the department Droit, Economie, Politique keeps the majority of the library's newsprint holdings. In fact, when the library moved from the Richelieu site to the new building 8 years ago, this recently created department became the principal heir of the Periodicals department which disappeared in so far. Within the department Droit, Economie, Politique the Newsprint office (Service de la Presse) and the Government Publications office (Service des publications officielles) are in charge of most of the newspaper holdings.

Just a few totals to give you an idea of the extent of the holdings :

Current entries into the stacks (in paper format)

Daily newspapers and large size periodicals of general information contents or concerning law, finances, politics, as well as periodical government publications.

Different entry modes to the stacks :

- Most titles by legal deposit, including also some foreign titles
- Expensive acquisitions (subscriptions)
- Gifts
- Exchanges

At present the total number is 832 titles of which 225 daily newspapers. On average more than 7300 issues arrive every month. The yearly increase of the holdings covers 200 linear meters (ml) plus 40 linear meters of periodical government publications forming a subset to the newsprint collection.

The Retrospective Collection (in paper format)

It includes two groups.

Periodicals (allotting the call numbers Jo, Gr fol-Jo, JoB and JoA) up to end 1990 including more than a 100 000 titles : local publications, technical journals, extremely varied magazines.

Some samples :

- Regional and local newspapers (including former French territories such as Algeria).
Thus we have as well *L'Abeille des Vosges* (Jo-11021) as *Le Progrès de Sétif* (Jo-88403) ;
- Professional newsprint : *L'Industriel forain* (Jo-8942) or the *Bulletin des fruits et primeurs* (Jo-31150) ;
- Trade-union or association bulletins such as *L'Auto-tram de Paris* (Jo-30242) or *La France mutilée* (Jo-25331) ;
- Children's magazines (ex.: *Les Belles images* Jo-55958).

Large size periodicals and/or daily newspapers kept by the former Periodicals department. In this collection you find the holdings of the *International Herald Tribune* (Pb-1751 bis) or of the daily *Les Echos* (Lc2-6878).

Microfilmed periodicals belonging to this closed collection are not available in paper format as the majority of the holdings are no more stored on the François Mitterrand site but forwarded to the library's Technical Centre in Bussy-Saint-Georges for preservation.

The Newspaper Microfilm Collection

If not otherwise stated the entire Newspaper Microfilm Collection, formerly kept in the Periodicals department, moved into the department Droit, Economie, Politique (under the call number Micr D or Mfilm + the call number of the original holding in paper format). This collection counts up to some 2200 titles (or call numbers). At the time it was moved to the new site, it included round 75 000 reels. It comprises the great national retrospective or current newspaper titles (all the dailies from the 19th and 20th centuries) and a large choice of past or running regional newspaper titles or newspapers from the former French territories.

The yearly increase of new microfilmed titles counts some 5500 reels including

- Current microfilming of several national dailies (*Le Monde, Le Figaro, La Croix, L'Humanité...*) and regional dailies (PQR = Presse régionale quotidienne).
To be precise, with regard to regional newspapers with multieditions the BnF purchases or makes the microfilm of all the editions but in paper format the Library only keeps the main edition. Since 2004 the preservation of these

regional dailies is shared with the public libraries in charge of the Printers' legal deposit ;

- Retrospective microfilming. As to very brittle holdings they must be restored partially or totally by thermosizing before filming.

In 2004, 229 newspapers holdings were microfilmed that way, which means some 400 000 frames.

2. The 2005-2009 Digitisation Programme of Retrospective National Daily Newspapers

Within this group, since mid-19th century, the highly circulated daily newspapers have attracted not only journalists' articles but also contributions from politicians, writers, artists, researchers. They all got hold of this media whose vast circulation was unknown until then, in order to publish essays, reviews, debates and novels. The quality and the variety of these contributions explain why newspapers from the 19th century to mid-20th century form an unrivalled supply to the study of the political, social, scientific, literary and artistic life.

However, material pressure due to preservation and access to the paper format or to the microfilm has resulted in an underdevelopment of the newspapers as a purpose for historical studies or, more generally, as a research aim. The digitisation of collections of newspapers and magazines therefore contributes to exploit such information collections in an still unequalled way.

This programme was defined and launched last year (2004). It is a five-year programme (2005-2009) which finally started in the beginning of 2005.

Selecting of the collection

The BnF decided to digitise twenty-one French nationally circulated newspapers as well as the supplements to six of those papers, from the first published issue up to 1944. These are :

- La Croix
- Le Temps
- Le Figaro (+ literary supplements)
- L'Humanité
- Le Petit parisien (+ weekly supplement)
- Le Matin
- Le Siècle
- La Presse
- Le Petit journal (+ weekly supplement)
- L'Action française
- L'Univers
- Le Gaulois (+ weekly supplement)
- Le Rappel
- La Lanterne (+ weekly supplement)
- Gil Blas (+ weekly supplement)
- La Justice
- L'Intransigeant

- L'Aurore
- Le Constitutionnel
- L'Echo de Paris
- Le Journal des débats

Some of these dailies which started mid-19th century stopped publishing long before 1944. Other titles disappeared during the Second World War. A few titles which continued to be published during the war stopped precisely in 1944. And some others are still running. As a matter of fact the digitisation is starting with these newspapers: *La Croix* (1883 -), *L'Humanité* (1904 -), *Le Figaro* (1826 -). Those titles as well as *Le Temps* (1861-1942) should be totally digitised before the end of 2006.

Besides the historical caesura of 1944 the choice of this deadline has allowed us to make an agreement quickly and easily with the administrations of the still running newspapers which rightfully could aim at digitising and putting online commercially their collections of the last sixty years. It also was agreed that the limit of 1944 should form a mobile barrier which may go up a year each year. This means that the BnF gradually will be able to give access on line of the following years.

For the time being, 27 newspaper titles (from the first published issue up to 1944) represent 3.2 million pages.

Digitisation methods and formats

As a first step of the Programme it was decided to use image digitising of this vast collection. Two reasons were pointed out :

- The image mode is very well controlled today and less expensive than the text mode
- Research workers using retrospective newspapers need to see (even if they have also other needs) the articles in their original aspect and background, their page-setting and their reference to illustrations, etc.

Pages are digitised from the original newspaper and preserved in non compressed TIFF format which is considered to be the most reliable for long-term preservation while the circulation will be in JPEG format on the BnF's website in its digital library *Gallica*.

The digitisation will be carried out by three digitisation lines :

- 2 lines within the BnF : one at its Centre in Sablé which is already set up and on work, the other one at the library's technical Centre in Bussy-Saint-Georges which is being set up now ;
- after inviting bids a third line committed to a provider off site who will also work within the library on the Bussy-Saint-Georges site.

If access to the facsimile image of each page is absolutely necessary to most research workers, we also know that they want to dispose of other access methods : all the possibilities of full text search, the facilities of hypertext navigation, etc.

In order to offer such possibilities the second stage of BnF's programme is to prepare the conversion of the original digitisation product in image mode to a digital collection which will

be accessible in text mode after having been treated by Optical Character Recognition (OCR) software.

Access and navigation methods and functions

The collection of the digitised newspapers will be accessible either through the BN-Opale Plus catalogue which clearly points out the documents available in an electronic version, or by searching directly in the *Gallica* database.

The search functions of the documents were firstly dealt with in a compilation of informations and experiences from some big libraries having already carried out a digitisation programme on line of a newspapers collection (Australia, Austria, USA, Norway, New-Zealand, United Kingdom). All this information was then studied by a working group including members of the different relevant departments of the BnF. Moreover, a survey was set up consisting in interviews with a range of research workers using newspapers in the reading rooms in order to get a real idea of what they need in this field. At present the following functions have been hold up :

- search by date : access through a calendar (year/month/day) ; « kiosk » function (ex. : all the newspapers of a given day) ;
- full text search with possible segmentation (especially isolating of a given article on a full text page) ;
- within the same issue : logical consultation of the layout (like using a railway plan) is needed to know about the structure of an issue ; advertising of pagination so as to find one's way in this structure ; zooming of an extract to facilitate reading ;
- the possibility of navigating from one issue to another within the same title.

Funding

The estimated amount of the costs of digitisation in image mode of the complete collection is 3.5 million Euros. It includes particularly :

- salaries and national insurance contributions to members of BnF's staff in charge of this work (control, completing and preparing of the holdings ; working of the 2 lines within the house)
- costs of data storage
- providing committed to the sub-contracting off site (third digitisation line ; special work on the access method to the digitised collection)

The amount of 3.5 million Euros is entirely supported by BnF's proper budget which especially is supplied with grants from the Ministry of Culture and Communication.

In order to start the second stage of the work - passing over to text mode - the BnF is looking actively for supplementary funding, in particular through sponsoring and partnership. The first financial support to the BnF for this part of the 5 year digitisation programme has come from the Senate which decided to make a grant of 150 000 Euros per year during three years (2005-2007), a total of 450 000 Euros.

3. Future prospects of developing and enlarging of the Programme

From now on several tracks of enriching and enlarging of this newspaper digitisation programme are already explored or future plans are taking shape. The working tracks often

implicate to co-operate with other libraries or, as far as running titles are concerned, with the newspapers themselves.

Completing of the initial collection and its enriching through historical factors

When some issues are lacking in BnF's holdings or if the physical condition of the documents make it difficult to carry out proper digitising the BnF tries to find these issues in other libraries or, in some cases, it may be necessary to apply to the newspaper administrations whenever the newspapers are still running.

Beyond this completing, researches could be undertaken so as to find and digitise former censored articles of the relevant newspapers from brush proofs (last proofs before final printing of the newspaper) generally kept in some newsprint archives.

Enlarging by Including Regional Daily Newspapers

An agreement has already been made with the administrative staff of the regional daily *Ouest France* (the largest circulated French newspaper) so that the BnF may digitise and put online the former title of this daily, *Ouest-Eclair* (1899-1934, main edition).

The opening of BnF's digitisation programme to regional daily newspapers (PQR) will truly not remain at that stage. Co-operative prospects with libraries in some regions and, once more, with the newspaper administrations themselves are in preparation. For instance, in the region Champagne-Ardenne, the digitisation of the regional daily *L'Union* as well as of the other titles of its group (*L'Est-Eclair*, *l'Ardennais*) is studied at the moment jointly by the administrative staff of the newspaper, by the three main public libraries in the region (Châlons-en-Champagne, Reims, Troyes) and by the BnF. This co-operative effort may also concern a certain number of retrospective local newspapers, especially in the department Aube.

Enlarging to specialised collection of rare documents

Besides specialised journals which are already digitised in the Gallica database (ex. *Le Journal des économistes*, an important monthly economic journal from mid-19th century to 1940), collections of very specialised newspapers which are rare and of great historical value, will also be digitised and put online. As an example the BnF and the Bibliothèque de Documentation Internationale Contemporaine (BDIC) have undertaken a shared digitisation programme of their respective and complementary holdings of the « trench newspapers », that means newspapers written and carried out by the soldiers themselves in the trenches in Eastern France during the First World War.

Addition of tools to facilitate researches

When the time has come, we shall study the possibility of digitising existing tools (such as tables, indexes, newspaper yearbooks) in order to link them to the collection of digitised newspapers. For instance, we think of the tables of the daily *Le Temps* or of the subject file of the newspaper *Le Matin* which is examined by a research worker right now.

In this field the carrying out of a major project would represent a capital tool for the newspaper history and for the history in general : the digitisation of the Bibliography of the

French political et general information newspapers (Bibliographie de la presse française politique et d'information générale or BIPFPIG) which identify and locate such newspapers in libraries and archives in France, department by department.

The digitisation Programme of retrospective newspapers now undertaken by the BnF will bring up its first results in the beginning of 2006. As we have already noted, the programme can only be imagined in co-operation with certain newspapers and certain libraries in France.

It is clear that it falls particularly well in with an international co-operation. At least two axes of partnership and opening are taking shape from now on :

- the French speaking dimension, in particular with the future prospects of connecting BnF's Programme and the initiative of the National Library of Quebec (whose director is a member of BnF's Scientific Council) concerning digitisation of newspapers from both countries ;
- the European dimension, with the project aiming the creation of a European digital library supported by the European Union and several member countries, in which, according to us, newspapers should hold an essential position.