



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>

juin 17, 2005

Code Number:

141-F

Meeting:

97 Newspapers

Les développements en matière de périodiques électroniques a la Bibliothèque Nationale de France: la numérisation de la presse quotidienne française du milieu du XIXème siècle a 1944

Pascal Sanz,

Directeur du Département droit, économie, politique,
Bibliothèque nationale de France
Paris, France

La Bibliothèque nationale de France (BnF) possède un ensemble de presse française et étrangère extrêmement important, comptant, fin 2004 quelque 60 000 titres morts ou vivants.. Mais c'est, bien entendu, tout d'abord à l'égard de sa collection de titres de presse français que la BnF estime avoir un devoir de sauvegarde et de mise en valeur. Il était donc tout à fait légitime autant que naturel que la bibliothèque choisisse, dans cet ensemble considérable, de numériser un corpus crucial pour la recherche dans pratiquement tous les domaines des sciences de l'homme et de la société : la presse quotidienne .

Pour présenter ce programme de numérisation, je commencerai par présenter rapidement les collections de presse française de la BnF : cela constituera la première partie de cet exposé . Puis, dans une deuxième partie, j'indiquerai quels sont les choix scientifiques et techniques du chantier de numérisation de la presse. Je conclurai en évoquant les perspectives d'évolution et d'élargissement de ce programme.

1. Les collections de presse française de la Bibliothèque nationale de France

Il existe différentes classifications de la presse française (nomenclature INSEE des produits presse, classification de l'Office de justification de la diffusion, etc.). S'inspirant de celles-ci on peut diviser la presse en quelques grands ensembles :

- la presse nationale, qui comprend :

- la presse politique et d'information, regroupant quotidiens (comme *Le Temps*, *Le Figaro* ou *l'Humanité*) et hebdomadaires
- la presse populaire fortement illustrée et comportant romans-feuilletons ou bandes dessinées, comme *Le Petit Journal* lancé au prix de 5 centimes par Moïse Millaud en 1863
- la presse d'échos et la presse satirique
- la presse régionale et locale : grands quotidiens régionaux à éditions locales (ex : *Ouest-France*), hebdomadaires locaux
- la presse d'outre-mer qui concerne les anciennes colonies et ex-territoires français (citons *La Dépêche algérienne* ou *L'Echo d'Oran*) ainsi que les actuels DOM-TOM
- la presse spécialisée comprenant notamment la presse féminine (ex : *le Petit courrier des dames*), la presse de loisirs, la presse économique, juridique, sportive (ex : *L'Auto*), etc.
- la presse de lecture dont font partie, entre autres, la presse enfantine, la presse des jeunes ou les illustrés pour la jeunesse (par exemple *La Semaine de Suzette*), la presse d'évasion, la presse religieuse, la presse associative
- la presse gratuite
- la presse de type publication officielle : journaux officiels, bulletins municipaux, etc.
- la presse alternative

La notion de presse recouvre donc une très grande variété de publications et on retrouve des fonds de presse dans tous les départements de collections de la BnF. C'est cependant au Département droit, économie, politique que se trouve concentrée la plus grande partie d'entre eux. En effet, au moment du transfert des collections de l'ancien bâtiment de la rue de Richelieu, il y a huit ans, ce département nouvellement créé s'est trouvé être le principal héritier du Département des périodiques, qui disparaissait en temps que tel. Au sein du Département droit, économie, politique, ce sont le Service de la presse et le Service des publications officielles qui détiennent la majorité des collections de presse.

Quelques chiffres pour donner une idée de l'ampleur de ces collections :

Entrées courantes en magasin (support papier)

Il s'agit de quotidiens ou de périodiques de grand format, d'information générale ou relevant des disciplines droit, économie, politique ainsi que des publications officielles périodiques.

Les différents modes d'entrée en magasin sont :

- le dépôt légal pour la très grande majorité des titres, y compris pour des titres étrangers
- les acquisitions onéreuses (abonnements)
- les dons
- les échanges

Le tout représente actuellement 832 titres dont 225 quotidiens. Plus de 7300 fascicules sont reçus en moyenne chaque mois. L'accroissement annuel s'élève à plus de 200 mètres linéaires (ml) auxquels s'ajoutent environ 40 ml de publications officielles périodiques constituant un sous-ensemble de la presse.

Fonds clos (support papier)

Il comprend deux ensembles.

Les périodiques (cotés Jo, Gr fol-Jo, JoB et JoA) allant jusqu'en 1990 inclus, le tout représentant plus de 100 000 cotes. Il s'agit de publications d'intérêt local, de revues techniques ou de magazines d'une extrême diversité.

Quelques exemples :

- la presse régionale ou locale (incluant les anciens territoires français comme l'Algérie). Ainsi, nous avons aussi bien *L'Abeille des Vosges* (Jo-11021) que le *Progrès de Sétif* (Jo-88403) ;
- la presse professionnelle : *L'Industriel forain* (Jo-8942) ou le *Bulletin des fruits et primeurs* (Jo-31150) ;
- les bulletins syndicaux ou d'associations comme *L'Auto-tram de Paris* (Jo-30242) ou *La France mutilée* (Jo-25331) ;
- les revues pour la jeunesse (*Les Belles images* Jo-55958) .

Les périodiques grand folio et/ou quotidiens précédemment conservés au Département des périodiques. On trouve ainsi dans ce fonds la collection de *l'International Herald Tribune* (Pb-1751 bis) ou du quotidien *Les Echos* (Lc2-6878).

Dans ce fonds clos, les titres qui ont été microfilmés ne sont plus communicables sous leur forme papier et, dans la très grande majorité des cas, les collections elles-mêmes ne sont plus conservées dans les magasins du site François Mitterrand mais ont été envoyées au Centre technique de Bussy-Saint-Georges.

Fonds de microfilms de presse

Le fonds de microfilms de presse précédemment conservé au Département des périodiques a été déménagé, sauf exception, dans le Département Droit, économie, politique : (cotés Micr D ou Mfilm + la cote papier d'origine). Ce fonds représente environ 2200 cotes et comprenait plus de 75 000 bobines lors de son déménagement. On y trouve les grands titres de presse nationale ancienne et courante (tous les grands quotidiens des 19^{ème} et 20^{ème} siècle) et une large sélection de titres régionaux morts ou vivants ou de presse des anciens territoires français.

Pour les nouveaux titres de microfilms l'accroissement annuel (Micr D et Mfilm Jo) s'élève à environ 5500 bobines comprenant,

- le microfilmage courant de plusieurs titres de presse quotidienne nationale (Le Monde, Le Figaro, La Croix, L'Humanité ...) et de presse quotidienne régionale (PQR). Précisons que pour les titres de PQR à éditions multiples, la BnF achète ou fait réaliser le microfilm de toutes les éditions alors que pour le papier elle ne conserve que l'édition principale et que depuis 2004 la conservation de la PQR est partagée avec les bibliothèques municipales en charge du dépôt légal imprimeur ;
- le microfilmage à titre rétrospectif. Pour les collections en très mauvais état, on est obligé de restaurer partiellement ou intégralement les fascicules par thermocollage avant de pouvoir effectuer la prise de vues.

En 2004, c'est ainsi 229 titres qui ont été microfilmés : pratiquement 400 000 images.

2. Le programme 2005-2009 de numérisation de la presse quotidienne nationale rétrospective .

Dans cet ensemble, la presse quotidienne à grand tirage, dès le milieu du XIXème siècle, a drainé, à côté des articles des journalistes, les contributions des hommes politiques, des écrivains, des artistes, des savants, qui se sont emparés de ce média à la diffusion sans précédent pour publier essais, critiques, tribunes et romans. La qualité et la diversité de ces textes font de la presse du XIXème et de la première moitié du XXème siècle une source incomparable pour l'étude de la vie politique, sociale, scientifique, littéraire et artistique.

Et pourtant, les contraintes matérielles de conservation et de consultation du papier ou du microfilm ont entraîné une sous-exploitation de la presse comme objet d'histoire et plus largement comme objet de recherche. C'est pourquoi la numérisation des corpus de journaux et magazines apporte des solutions, jusqu'à présent inégalées, d'exploitation de ces corpus documentaires.

Ce programme a été défini et décidé l'an dernier (2004). Il s'agit d'un plan pour cinq ans, 2005-2009, qui a effectivement commencé début 2005.

Le choix du corpus

La BnF a décidé de numériser vingt et un journaux français de diffusion nationale et les suppléments de six d'entre eux, de leurs origines respectives à 1944. En voici la liste :

- La Croix
- Le Temps
- Le Figaro (+ supplément littéraire)
- L'Humanité
- Le Petit parisien (+ supplément hebdomadaire)
- Le Matin
- Le Siècle
- La Presse
- Le Petit journal (+ supplément hebdomadaire)
- L'Action française
- L'Univers
- Le Gaulois (+ supplément hebdomadaire)
- Le Rappel
- La Lanterne (+ supplément hebdomadaire)
- Gil Blas (+ supplément hebdomadaire)
- La Justice
- L'Intransigeant
- L'Aurore
- Le Constitutionnel
- L'Echo de Paris
- Le Journal des débats

Certains d'entre ces quotidiens, nés entre le milieu et la fin du XIXème siècle ont cessé de paraître bien avant 1944. D'autres ont disparu pendant la seconde guerre mondiale. Quelques-uns, qui s'étaient maintenus pendant la guerre ont, précisément, disparu en 1944. Quelques-uns, enfin, sont des titres toujours vivants et c'est par eux, d'ailleurs que nous commençons la numérisation : La Croix (1883 -), L'Humanité (1904 -), Le Figaro (1826 -). Ces trois titres ainsi que Le Temps (1861 – 1942) devraient être entièrement numérisés d'ici à fin 2006.

Outre la césure historique que représente l'année 1944, le choix de cette date limite a permis de trouver très rapidement et facilement un accord avec les directions des journaux vivants, qui peuvent aspirer très légitimement à numériser et mettre en ligne commercialement leurs collections des soixante dernières années. Il est convenu avec eux que cette limite -1944 - constituait une barrière mobile, qui pourra s'élever chaque année d'un an, c'est à dire que la BnF pourra, si elle le souhaite, fournir ensuite progressivement l'accès en ligne aux années suivantes.

Pour le moment, des origines jusqu'en 1944 donc, le total des 27 titres représente 3,2 millions de pages.

Les modes et formats de numérisation

Il a été décidé de numériser, dans une première étape, ce très important corpus en mode image et ce pour deux raisons :

- parce que cette technique, très bien maîtrisée, est beaucoup moins coûteuse que la numérisation en mode texte,
- parce que les chercheurs qui ont recours à la presse rétrospective ont besoin (même s'ils peuvent avoir aussi d'autres besoins) de consulter les articles dans leur forme originale, avec leur contexte, le mise en page, le rapport aux illustrations, etc.

Les pages des journaux sont numérisées à partir des originaux et conservées en format TIFF non compressé, jugé le plus fiable pour la conservation pérenne, tandis que la diffusion se fera au format JPEG, sur le site web de la BnF, dans sa partie bibliothèque numérique, « Gallica ».

La numérisation sera effectuée par trois chaînes de numérisation :

- 2 chaînes internes à la BnF, dont une sur son site de Sablé, déjà installée et qui a commencé à produire et l'autre au Centre technique de Bussy-Saint-Georges, en cours d'installation,
- une chaîne confiée, après appel d'offre, à un prestataire externe, mais qui sera installée également dans les murs de la bibliothèque, à Bussy-Saint-Georges.

Mais nous savons bien que, si l'accès à l'image en fac-similé de chaque page est indispensable à la plupart des chercheurs, ceux-ci souhaitent également disposer d'autres modalités de consultation : toutes les possibilités de la recherche plein texte, les facilités de la navigation hypertextuelle, etc.

Afin d'offrir ces possibilités, la BnF se prépare donc, dans une deuxième étape, à convertir le produit de la numérisation originale en mode image en un corpus numérique accessible aussi en mode texte, après utilisation d'un logiciel de reconnaissance optique de caractères (OCR).

Les modes et fonctionnalités d'accès et de navigation

Le corpus de presse numérisée sera accessible soit à travers le catalogue Bn-Opale Plus, qui signale clairement quels sont les documents faisant l'objet d'une version électronique, soit en interrogeant directement la base Gallica.

Les fonctionnalités de consultation des documents ont d'abord fait l'objet d'un recueil d'informations et d'expériences auprès d'un certain nombre de grandes bibliothèques ayant déjà engagé un programme de numérisation et de mise en ligne d'un corpus de presse (Australie, Autriche, Etats-Unis, Norvège, Nouvelle-Zélande, Royaume-Uni). Les informations ainsi recueillies ont ensuite été étudiées par un groupe de travail comprenant les différents départements concernés au sein de la BnF. De plus une enquête par entretiens auprès d'un échantillon de chercheurs utilisateurs de la presse dans les salles de lecture de l'établissement a été menée, afin de disposer d'une connaissance plus fine de leurs propres souhaits en la matière. A l'heure actuelle, les fonctionnalités suivantes ont été retenues :

- recherche par date : accès par un calendrier (année/mois/jour) ; fonction « kiosque » (exemple : tous les journaux d'un jour donné) ;
- recherche plein texte avec possibilité de segmentation (notamment, isolement d'un article donné dans une page en plein texte) ;
- au sein d'un même fascicule : consultation du chemin de fer, pour connaître la structure du numéro ; affichage de la pagination, pour se repérer dans cette structure ; zoom sur un extrait, pour en faciliter la lecture ;
- au sein d'un titre : navigation d'un fascicule à l'autre.

Le financement

Le coût de la numérisation en mode image pour l'ensemble du corpus a été évalué à 3,5 millions d'euros. Ce coût recouvre notamment :

- les salaires et charges des agents de la BnF affectés à cette tâche (vérification, complètement, préparation des collections ; fonctionnement des deux chaînes internes),
- le coût du stockage des données,
- les prestations confiées à la sous-traitance externe (troisième chaîne de numérisation ; certains travaux relatifs au mode d'accès au corpus numérisé).

Ce coût de 3,5 millions d'euros est financé entièrement sur le budget propre de la BnF, lui-même principalement alimenté par les subventions annuelles que lui attribue le Ministère de la culture et de la communication.

Pour engager la deuxième phase de travaux - le passage au mode texte - , la BnF recherche activement des financements complémentaires, notamment via le mécénat et le partenariat. Le premier soutien financier ainsi obtenu par la BnF lui est accordé par le Sénat, qui a décidé d'apporter au plan quinquennal de numérisation de la presse une dotation de 150 000 euros par an pendant trois ans (2005-2007), soit 450 000 euros au total.

3. Les perspectives d'évolution et d'élargissement de ce programme.

Plusieurs pistes d'enrichissement et d'élargissement de ce plan de numérisation de la presse sont d'ores et déjà explorées ou se dessinent en perspective. Souvent, ces pistes de travail impliquent une coopération avec d'autres bibliothèques ou, pour les titres vivants, avec les journaux eux-mêmes .

La recherche de l'exhaustivité du corpus initial, son enrichissement d'éléments historiques.

Lorsque certains fascicules manquent à la BnF ou se présentent dans état physique permettant difficilement la numérisation dans de bonnes conditions, la BnF recherche ces fascicules dans d'autres bibliothèques ou, dans certains cas, auprès des journaux eux-mêmes.

Au delà de ce travail de complètement, un travail de recherche pourrait être mené pour retrouver et numériser, à partir des morasses, c'est à dire des dernières épreuves avant impression du journal conservées dans certaine archives de presse, les articles censurés, à certaines époques, pour les titres concernés.

L'extension à la presse quotidienne régionale

Dès à présent, un accord a été conclu avec la direction du Quotidien régional Ouest-France (le plus gros tirage de la presse française) pour que la BnF numérise et mette en ligne l'ancêtre de ce titre, Ouest-Eclair (1899 – 1934, édition principale).

Vraisemblablement, l'ouverture du programme de numérisation de la presse de la BnF à la PQR n'en restera pas là. Des perspectives de collaboration avec des bibliothèques de certaines régions et, encore une fois, avec certains journaux, sont à l'étude. Par exemple, en Champagne-Ardenne, la numérisation du journal régional L'Union et des autres titres de son groupe (L'Est-Eclair, L'Ardennais) est actuellement étudiée conjointement par la direction de ce journal, les trois principales bibliothèques municipales de la région (Châlons-en-Champagne, Reims, Troyes) et la BnF. Cette coopération pourrait aussi porter sur un certains nombres de titres rétrospectifs locaux, notamment dans le département de l'Aube.

L'extension à des corpus spécialisés de documents rares

A côté de titres de revues spécialisées qui se trouvent déjà numérisées dans la base Gallica (exemple : Le Journal des économistes, grand mensuel d'économie du milieu du XIXème siècle à 1940) des corpus de presse très spécifiques, rares et d'une haute valeur historique seront également numérisés et offerts à la consultation en ligne. Par exemple, la BnF et la Bibliothèque de documentation internationale contemporaine (BDIC) ont entrepris la numérisation concertée de leurs collections respectives et complémentaires de « Journaux de tranchées », c'est à dire les journaux rédigés et réalisés par les soldats eux mêmes dans les tranchées de l'Est de la France, pendant la Première Guerre mondiale.

L'adjonction d'outils facilitant la recherche

Il s'agira, le moment venu, d'examiner la possibilité de numériser des outils existants (tables, index, annuaires de la presse) et de les mettre en relation avec le corpus de presse numérisée. Nous pensons, par exemple, aux Tables du Temps ou au fichier par sujets du journal Le Matin, sur lequel travaille actuellement un chercheur.

En la matière, la réalisation d'un projet majeur constituerait un instrument capital pour l'histoire de la presse et l'histoire tout court : la numérisation de la Bibliographie de la presse française politique et d'information générale (BIPFPIG), qui identifie et localise dans les bibliothèques et archives, département par département, la presse de cette nature.

Le chantier de numérisation de la presse rétrospective, désormais engagé par le Bibliothèque nationale de France et dont les premiers résultats seront visibles dès le début 2006, ne se conçoit déjà, comme on l'a vu, que dans une coopération avec certains journaux et certaines bibliothèques en France.

Il est clair qu'il se prêterait particulièrement bien à une coopération internationale. Au moins deux axes de partenariat et d'ouverture se dessinent dès à présent :

- la dimension francophone, avec notamment la perspective de la mise en relation du programme de la BnF et de l'initiative de la Bibliothèque nationale du Québec (dont la directrice siège au Conseil scientifique de la BnF) en matière de numérisation de la presse des deux pays ;
- la dimension européenne, avec le projet de création d'une bibliothèque numérique européenne, soutenu par l'Union et plusieurs de ses pays membres et dans laquelle la presse devrait, selon nous, tenir une place capitale.