

Michel Richard

Bibliothèque de France
Département Informatique et nouvelles techniques

LE PROGRAMME DE NUMÉRISATION DE LA BIBLIOTHÈQUE DE FRANCE

« Quiconque a connu les longues heures de vagabondage entre les livres, quand le moindre fragment, le moindre code, la moindre initiale semble ouvrir une voie nouvelle, aussitôt perdue dès que se présente une nouvelle rencontre, ou quiconque a éprouvé le caractère illusoire et labyrinthique de cette " loi du bon voisinage " sous l'autorité de laquelle Warburg avait placé sa bibliothèque, sait que l'étude non seulement ne peut pas avoir de fin, mais qu'elle ne désire pas en avoir ».

Giorgio Agamben, « Idée de l'étude », *Idée de la prose*, Christian Bourgois, 1988.

L'UTILISATION des nouvelles technologies, et plus particulièrement des techniques de numérisation, s'était imposée aux rédacteurs des *Propositions pour une grande bibliothèque*. Patrice Cahart et Michel Melot avaient alors indiqué la place que ces pratiques « *d'un type entièrement nouveau* » devaient occuper au sein du nouvel établissement. Il s'agissait, d'une part, de faciliter le transfert des livres sous une forme immatérielle du site de Richelieu à celui de Tolbiac, dans l'hypothèse de la « césure » des collections de la Bibliothèque nationale ; d'autre part, de mettre à la disposition des chercheurs des « postes de lecture assistée par ordinateur » (PLAO) permettant la consultation de plusieurs centaines de milliers de documents électroniques.

Michel RICHARD est chef de projet Numérisation

Ces éléments les plus novateurs du projet ont été repris par l'établissement constructeur qui les a développés selon deux axes majeurs :

- la constitution de plusieurs collections numérisées, comprenant des imprimés (objectif de 100 000 volumes pour l'ouverture de la bibliothèque), des œuvres sonores provenant de la Phonothèque nationale (10 000 heures) et des images fixes, illustrations, beaux livres et collections de photographies (600 000) ;
- parallèlement, le développement d'outils d'accès, de consultation et de travail, sur place et à distance : poste de simple consultation, PLAO, poste audiovisuel, poste multimédia expérimental.

La réalisation de ces ambitieux objectifs passe par un ensemble de dispositifs théoriques et pratiques qui permettent d'analyser les besoins puis de valider les choix à mettre en œuvre.

Sur le plan scientifique, la concertation avec les chercheurs, les universitaires, les bibliothécaires et tous les utilisateurs potentiels éclaire l'établissement sur les choix éditoriaux, les collections, les titres à numériser en priorité. Des chercheurs sont également mis à contribution pour l'évaluation des logiciels de lecture / écriture par l'expérimentation des stations de travail et la confrontation avec leur propre pratique.

Sur le plan informatique, le programme doit s'assurer des liens avec l'ensemble du système d'information de la bibliothèque, de la capacité de dialogue des nouveaux outils avec les applications internes qui lui sont immédiatement associées (catalogue, gestion des demandes, facturation) et avec les réseaux extérieurs, y compris avec les matériels dont les usagers disposeront chez eux à la fin du millénaire.

informatique. Ce sera bien sûr le cas des projets qui utiliseront les ressources de l'hypertexte ou de l'hypermédia (le projet PERSEUS appliqué aux textes grecs de l'Antiquité en est aujourd'hui la meilleure illustration). Pour la Bibliothèque de France, cela signifie clairement que ces divers

semble des bibliothèques disposant de fonds patrimoniaux ou spécialisés, et par celles qui souhaiteraient mettre à la disposition des chercheurs des stations de travail informatique sur les textes – seront en mesure de répondre aux besoins et objectifs suivants :

numérisation est une technique de « préservation ».

Enfin, une partie des ouvrages microfilmés à des fins de conservation pourront être eux-mêmes numérisés. A terme, la consultation électronique remplacera progressivement celle des microfilms.

– *Faciliter l'accès à l'information*

Grâce au support numérique, le document sera immédiatement accessible et consultable simultanément par plusieurs utilisateurs. Sa communication se fera sans recours au personnel dans les salles de lecture.

Le document pourra être communiqué électroniquement, ou reproduit sur papier, autant de fois qu'il sera nécessaire, soit pour enrichir les collections en libre accès, soit pour répondre à une demande du lecteur : c'est l'équivalent des travaux photographiques actuels ou des tirages sur papier à partir des microformes.

Enfin la recherche automatique du document permettra des travaux bibliographiques complexes dans des délais très courts grâce à la compatibilité du poste avec les autres systèmes documentaires de la bibliothèque.

Le document pourra être communiqué électroniquement, ou reproduit sur papier, autant de fois qu'il sera nécessaire

types d'édition devront être mis à la disposition de son public. Sur le poste de lecture assistée par ordinateur, on pourra donc disposer des publications électroniques existantes, de CD-ROM (qui figureront également dans les collections en libre accès) et de bases textuelles (selon des procédures d'accès éventuellement hétérogènes, interrogation à distance à partir du poste de lecture ou intégration dans la base de la bibliothèque). A cela s'ajouteront les reproductions numériques de livres provenant des fonds patrimoniaux des bibliothèques (principalement la Bibliothèque nationale) ou d'ouvrages encore disponibles, qui seront réalisées directement par l'établissement. Par ailleurs, la « représentation » du texte à l'écran, selon le type d'ouvrage et la technique de numérisation employée, pourra être soit en mode image (photographie de la page originale), soit en mode caractère (texte au kilomètre plus ou moins structuré).

– *Diminuer la pression sur les collections*

La numérisation n'a pas encore fait ses preuves en matière de conservation. La politique de la Bibliothèque de France s'appuie donc, de manière classique, sur la restauration et le microfilmage.

En revanche, les ouvrages numérisés pourront être communiqués sous une forme de substitution : à l'écran ou sur papier. La pression sur cette partie de la collection diminuera d'autant : la

Pourquoi numériser ?

L'établissement constructeur s'est donné pour objectif d'offrir l'équivalent de 100 000 volumes papier à l'ouverture de la bibliothèque. Au-delà de la constitution de ce fonds initial, si l'on se place dans une perspective à long terme, on peut légitimement penser que les techniques de numérisation – expérimentées à grande échelle par la nouvelle bibliothèque, puis appliquées par l'en-

– *Transmettre à distance*

Que ce soit pour le prêt entre bibliothèques ou le réseau des postes de lecture des pôles associés, la transmission à distance des ouvrages par des réseaux numériques, en ligne ou en différé, s'imposera rapidement compte tenu des avantages décrits ci-dessus : rapidité, disponibilité, simultanéité. Cette transmission pourra s'effectuer par des techniques numériques *stricto sensu*, mais aussi par la fourniture de copies électroniques voire de tirages sur papier des documents transférés (notamment pour des raisons d'économie).

– *Favoriser la lecture et l'écriture assistées par ordinateur*

Par la constitution de fonds numérisés et la mise à disposition de postes de lecture assistée par ordinateur, le programme de numérisation a pour ambition de proposer au public de nouveaux outils de lecture et de travail : il constitue de ce fait la part la plus innovante du système d'information de la Bibliothèque de France.

La démarche de conception du poste est fondée sur le constat que la lecture « savante » reposant sur une appréhension de l'œuvre et du texte

en tant qu'objets intellectuels qui ne se ramènent pas à de simples gisements d'informations, mais qui se fréquentent de diverses manières. C'est une lecture d'appropriation, qui ne revient pas à la consultation classique d'une base de données textuelles. On peut en qualifier les principaux traits fonctionnels : c'est une lecture compétente, puisque l'utilisateur connaît son domaine de recherche, mais qui n'exclut nullement la flânerie. Elle s'inscrit dans le long terme, dans un parcours individualisé qui se traduit par la confection de dossiers personnels ; lecture d'observation scrutative qui segmente les textes, regroupe, synthétise et qui engendre enfin une importante activité de rédaction.

Le poste de lecture assistée par ordinateur

Les caractéristiques essentielles du dispositif technique reposent donc prioritairement sur cette analyse. Ce dispositif offrira les fonctions suivantes :

– mise à disposition des usagers d'un outil informatique de type ordinateur personnel relié au système d'information de la Bibliothèque de France par

un réseau local et constituant un nouvel instrument de lecture appelé « poste de lecture assistée par ordinateur » ;

– à partir de ce poste, accès au catalogue, survol rapide du fonds électronique de la Bibliothèque de France et possibilité de sélection des ouvrages et parcours des textes ;

– téléchargement dans le poste de travail des ouvrages du fonds identifiés comme pertinents pour le travail de recherche, et constitution, ainsi, d'un corpus accessible localement ;

– enrichissement éventuel de ce corpus par des fonds numériques personnels apportés par l'utilisateur ;

– possibilité d'effectuer sur un corpus individualisé de recherche des opérations de lecture, de constitution de bases de données, de comparaison, de structuration, d'annotation et d'analyses textuelles.

Il s'agit ainsi, pour les lecteurs, de retrouver et d'enrichir, dans un cadre informatique, les fonctions courantes du travail en bibliothèque, où lecture et écriture se combinent étroitement.

Il est également prévu, et il s'agit d'une disposition importante, que la bibliothèque personnelle de chaque utilisateur puisse être conservée par l'établissement, favorisant ainsi le travail de longue haleine en lui permettant de retrouver son environnement de travail (extraits du fonds numérique structurés, fichiers, etc.) par la simple réservation d'un poste de lecture. De même, la numérisation à la demande sera possible dans certaines conditions (état de conservation de l'ouvrage, autorisation des ayants droit, etc.).

Le PLAO sera doté d'une mémoire de masse de plusieurs centaines de méga-octets et pourra utiliser des mémoires externes tout aussi volumineuses. Il comportera en périphérie des moyens de restitution – une imprimante – permettant la sortie de documents. Il comportera également des moyens de saisie – un lecteur de support externe, un scanner avec un

logiciel de reconnaissance optique de caractères – permettant à l'utilisateur de compléter son corpus de travail par des documents qu'il apportera ou numérisera lui-même.

Il est prévu d'installer progressivement jusqu'à 260 stations de ce type dans les deux espaces de la bibliothèque.

Pour maîtriser les développements de la station de travail à réaliser, l'établissement a mis en œuvre une méthodologie spécifique, fondée sur un banc d'essai de longue durée et sur le prototypage par élaboration de maquettes faisant l'objet d'un appel d'offres.

Depuis le mois de mars 1990, un banc d'essai réunissant neuf « grands lecteurs » (philosophes, linguistes, traducteurs, historiens) a évalué les hypothèses sous-jacentes à la conception de la station en utilisant un environnement matériel et logiciel existant sur le marché et s'approchant des paramètres de la future station.

ger les premières recherches entamées avec le banc d'essai par un élargissement du « public » virtuellement concerné. Des laboratoires du CNRS et d'autres équipes ont donc été invités à s'impliquer dans l'évaluation. Leur apport permettra de tester, sur des corpus qui peuvent être très spécifiques, la capacité de la machine à répondre à tous les besoins, et de faire l'inventaire des potentialités génériques à développer.

Des conventions de partenariat ont été ou vont être signées avec l'Institut des textes et manuscrits modernes (ITEM), les Archives Husserl (Ecole normale supérieure), le Centre Louis Gernet (Ecole des hautes études en sciences sociales), l'unité de recherche associée, Histoire des théories linguistiques (Université Paris VII), le Groupe d'études quantitatives interdisciplinaires sur la langue du XVI^e siècle (Université Blaise Pascal, Clermont-Ferrand), l'Institut Mémoires de l'édition contemporaine (IMEC),

Pour la gestion de la communication, ce sont les documents les plus demandés qu'il faut privilégier en tenant compte toutefois de la nature du document et de sa pertinence pour le travail informatique (des documents très structurés comme les dictionnaires posent de gros problèmes d'utilisation s'ils ne sont pas accompagnés de logiciels de navigation sophistiqués).

L'archétype du livre à numériser serait donc représenté par une œuvre rare, susceptible d'être souvent demandée, et bien adaptée au travail de lecture approfondie.

En partant de ces critères généraux, et en prenant en compte le fait qu'il s'agissait bien de constituer une collection numérique a priori, l'Etablissement public de la Bibliothèque de France a également examiné systématiquement les logiques et les programmes d'autres institutions comparables : le projet Adonis (British Library et Communauté économique européenne), les bases de données de l'INIST³ (CNRS), l'expérience FOUORE⁴ des bibliothèques universitaires françaises, le projet « *American Memory* » de la Bibliothèque du Congrès, le programme de la Bibliothèque nationale d'agriculture aux Etats-Unis, l'*Oxford Text Archive*, le projet soutenu par Xerox concernant les livres de mathématiques, etc.

Tous ces projets, et d'autres, combinent, de diverses manières les critères évoqués plus haut : le projet de la Bibliothèque du Congrès vise la préservation et l'usage informatique, le projet FOUORE la transmission entre bibliothèques universitaires, l'INIST la transmission et la rationalisation de la communication, etc.

Des conclusions peuvent être tirées de ces différentes expériences :

– Le coût des opérations de transfert interdit le recours à cette technique pour un seul usage. On tente au contraire de diversifier les exploitations des fonds numériques et de les valoriser. C'est le cas même pour les

Il est prévu d'installer progressivement jusqu'à 260 stations de PLA0 dans les deux espaces de la bibliothèque

S'appuyant également sur les conclusions de l'expertise effectuée par Jacques Virbel, chercheur à l'Institut de recherche informatique de Toulouse, le cahier des charges de l'appel d'offres a conduit à sélectionner deux projets : l'un présenté par la société AIS, filiale de Berger-Levrault, l'autre par Cap Gémini Innovation. Les maquettes furent ensuite installées respectivement sur des stations SUN et NeXT.

Actuellement sur le site de l'établissement public, à Ivry, ces applications facilitent la poursuite du travail d'expertise dans deux directions. D'une part, sur le court terme, l'analyse des prototypes permet d'élaborer les spécifications du système informatique tant du point de vue des outils installés sur le poste que de la gestion des interfaces avec les autres systèmes. D'autre part, et à plus long terme, il est apparu indispensable de prolon-

l'Université de technologie de Compiègne et l'Institut de recherche en informatique de l'Université Paul Sabatier de Toulouse.

Conséquences sur le choix des textes

Les différents usages possibles d'un texte numérisé à l'intérieur d'une bibliothèque entraînent différentes logiques de sélection des documents. Elles ne sont pas contradictoires. Elles doivent être intégrées de manière à définir un programme d'acquisition.

Dans une logique de préservation ou de transmission à distance, c'est le critère de rareté (dans les bibliothèques) qui prévaut, mais, pour décider de numériser, il faut croiser ce critère avec celui de la demande potentielle.

3. INIST : Institut national de l'information scientifique et technique.

4. FOUORE : Fourniture de documents sur réseau électronique.

bases de données à orientation linguistique (*Thesaurus linguae graecae*) : bien qu'elles soient destinées à la recherche documentaire informatisée, leur édition sur CD-ROM leur fait jouer aussi le rôle de support de diffusion des textes.

– La numérisation est rarement retenue comme seule technique de transfert et de stockage avant une communication sur papier. Qu'il s'agisse de recherche documentaire ou de consultation, des moyens informatiques se trouvent associés aux utilisations des documents numérisés.

– L'idée de « numérisation à la demande » (reprise de la notion « d'édition à la demande » apparue au début des années 80) soulève de nombreuses objections, et en particulier : le rassemblement des documents numérisés selon cette seule logique constitue des ensembles hétérogènes, sans visibilité pour les utilisateurs. Au contraire, l'importance accordée à la cohérence des collections numériques est un trait saillant de tous les projets récents. Elle est d'ailleurs confortée par les pratiques de l'édition électronique.

Il est frappant de constater que ces conclusions vont toutes dans le même sens : il ne s'agit plus de considérer la numérisation comme une technique au bénéfice d'un service particulier de la bibliothèque, mais de constituer des collections cohérentes en elles-mêmes et répondant à la vocation particulière de la bibliothèque concernée, puis de valoriser ces collections pour développer l'ensemble des missions et des usages de la bibliothèque.

Le fonds électronique à la BDF

Ces analyses ont conduit à une première approximation du fonds électronique et de sa place dans la Bibliothèque de France.

1. Que ce soit pour faciliter la transmission à distance ou pour mettre à la disposition du public des PLAO, le fonds électronique doit répondre à des critères de cohérence intellectuelle permettant d'offrir la plus grande visibilité.

En ce sens la réponse à la demande de numérisation ponctuelle d'un ouvrage de la part d'un usager devra principalement être appréciée en fonction de la pertinence d'intégrer ou non l'ouvrage dans l'ensemble de la collection.

qu'en soient le statut documentaire et le support matériel.

Il faut souligner que l'accès au document numérisé ne doit pas se substituer à la possibilité de consulter l'ouvrage original. Il semble donc logique, d'un point de vue intellectuel

L'accès au document numérisé ne doit pas se substituer à la possibilité de consulter l'ouvrage original

2. Le fonds électronique est intégré aux fonds de la bibliothèque. Son utilisation est intégrée à l'ensemble des services de la bibliothèque. C'est un des éléments constitutifs de la bibliothèque. Ce n'est pas toute la bibliothèque, ni même le clone électronique de la Bibliothèque nationale. Ce n'est pas une bibliothèque particulière composée uniquement d'« introuvables », une réserve dans la Réserve, ni un ensemble de documents rassemblés au fil de l'eau, en fonction de demandes particulières.

C'est un sous-ensemble des collections de la nouvelle bibliothèque, « neutre » par rapport aux autres unités ou ensembles documentaires, mais représentatif de la mission et du patrimoine détenu par la Bibliothèque de France.

Des différentes utilisations évoquées, aucune ne doit être privilégiée de manière systématique, pour des raisons de cohérence et de rationalité budgétaire.

En revanche, l'accès à distance faisant partie de la « charte » de la Bibliothèque de France et s'appuyant nécessairement, pour ce qui concerne les documents, sur la numérisation, il faut que, dès la mise en fonctionnement de ce service, les lecteurs des autres bibliothèques soient intéressés par les documents transférés.

3. Le fonds électronique n'est pas un substitut aux livres et autres documents de la bibliothèque.

Il est conçu comme un instrument de travail complétant en amont ou en aval l'ensemble des livres mis à disposition par la bibliothèque quels

et financier, de privilégier, pour la constitution du fonds initial, une démarche de programmation qui partirait : 1. d'une présomption générale d'usage, 2. d'un présupposé intellectuel lié à une utilisation des fonds patrimoniaux, 3. du préjugé de l'intérêt le plus collégial possible des livres rassemblés.

La collection numérique doit constituer un fonds de référence commun à tous les lecteurs dans les disciplines « d'excellence » de la bibliothèque. Elle doit être visible pour être immédiatement et fréquemment utilisée. Sans prétendre à l'exhaustivité et à l'encyclopédisme, elle doit offrir des corpus systématiques.

L'avis de la « communauté intellectuelle »

Les objectifs généraux ayant été définis, les contraintes techniques évaluées et les critères bibliothéconomiques précisés, il restait à concrétiser le contenu de la collection, à établir l'inventaire des corpus à rassembler en priorité, à dresser la liste des ouvrages à emprunter, acheter, reproduire, pour les numériser.

À la suite des résultats du groupe de travail consacré à la numérisation qui s'était réuni dès 1989, de nombreuses consultations, donnant parfois lieu à des débats, ont contribué à éclairer l'établissement public sur les choix à entreprendre. Au mois de novembre 1991, une séance du Conseil scientifique de l'EPBF examine les grandes

orientations scientifiques et technologiques. Les premières conclusions prévoient qu'une commission rassemblant des chercheurs de toutes disciplines, des éditeurs et de « grands lecteurs » s'associera au travail de l'équipe responsable du programme scientifique et précisera l'ampleur et la nature des corpus souhaités. Simultanément, des expertises sont sollicitées auprès d'institutions spécialisées et de personnalités diverses : universitaires, auteurs, éditeurs, directeurs des grandes bibliothèques de recherche, etc. La coopération avec la Bibliothèque nationale se met en place.

Parallèlement s'établit une mise en commun des travaux menés par les différentes commissions qui participent à la politique documentaire pour les acquisitions destinées au libre accès.

Enfin, au début du mois de novembre 1992, les experts réunis par André Miquel examinent le programme de numérisation des textes. La commission valide la démarche générale de l'établissement public et confirme la nécessité pour la future bibliothèque de disposer des technologies les plus avancées dès son ouverture, compte tenu de l'évolution des pratiques telle que l'on peut l'observer à l'étranger. Les experts insistent également sur l'intérêt de constituer un corpus important de textes destinés en priorité au traitement informatique, c'est-à-dire d'œuvres immédiatement accessibles en mode caractère.

La collection

Aujourd'hui, les domaines ayant fait l'objet d'une exploration approfondie sont les suivants : littérature et linguistique françaises, sciences du langage, philosophie, histoire des sciences et sciences exactes, histoire du Moyen Âge et histoire contemporaine, Antiquité (histoire, littérature, philosophie).

Au cœur de la collection on trouvera, pour toutes les disciplines, les « classiques », les « textes de référence » ou les textes fondateurs de la discipline. La plus grande partie sont des textes rares, mais, pour des raisons évidentes, les auteurs très connus,

dont les ouvrages sont encore bien diffusés, figurent aussi dans le projet de collection : pour le XVI^e siècle littéraire, par exemple, Montaigne, mais aussi Charondas, Pasquier, Paré, Postel... ; et pour l'histoire des sciences : Galien, Tycho Brahé, Gesner, Newton, Fermat...

Autour des textes de référence : des corpus. Il est envisagé de compléter cette collection de textes de référence, en y adjoignant, pour la littérature et l'histoire, deux types de documents. En littérature : une sélection de textes critiques permettant de constituer des petits corpus par auteur, avec une continuité significative entre l'œuvre, ses diverses éditions, avec l'apparat critique et les commentaires. Seules les véritables œuvres critiques ayant fait autorité ou faisant autorité par les changements qu'elles introduisent dans la réception de l'auteur étudié seront retenues. On y associera ensuite d'autres types de documents, bulletins de sociétés

Le noyau de la collection électronique répondra dans un premier temps à une typologie dont on peut donner les exemples suivants :

– *Le corpus littéraire*

Il s'étend de l'Antiquité gréco-latine au XX^e siècle mais il sera essentiellement consacré, à l'exception des auteurs de l'Antiquité, aux écrivains français ou de langue française : il comprend la totalité des œuvres et des corpus de disciplines connexes (poétique et rhétorique, histoire de la langue et de la grammaire, histoire du livre, de la lecture, de l'édition, des bibliothèques, des revues, de la presse, etc.).

Les collections de revues de création et de critique qui ont tenu une place très importante depuis le siècle dernier sont intégralement reconstituées et versées au fonds, à côté des grandes revues incontournables comme la *Nouvelle revue française*, le *Mercure de France* ou la *Revue d'histoire littéraire de la France*,

Pour répondre à une demande unanime des experts consultés, un fonds important de revues sera intégré à la collection

savantes et, si possible, manuscrits. En histoire, il est possible de reprendre des fonds disponibles sous forme de microformes et qui ont l'intérêt de constituer des « archives collectives de la recherche » : par exemple, l'ensemble réalisé par Pergamon à partir essentiellement des collections de la Bibliothèque nationale, *Archives de la Révolution française*.

Enfin, pour répondre à une demande unanime des experts consultés, un fonds important de revues sera intégré à la collection. Celles-ci constituent en effet des outils de référence pour toutes les disciplines, et parfois l'essentiel des documents souhaités pour certains champs (ethnologie, linguistique). De plus, certaines collections de revues peuvent être très difficiles d'accès (revues de littérature et d'art du début du XX^e siècle, par exemple).

Romania, Romantisme, Studi Francesi. On y trouve également la collection complète des revues symbolistes, dadaïstes, surréalistes, et *Commerce, Le Grand jeu, Littérature, Le Navire d'argent, Mesures, La Revue blanche, La Revue européenne, Les Soirées de Paris*, dont certaines ont connu une existence éphémère.

Les glossaires, dictionnaires anciens, les traités de rhétorique, les ouvrages de lexicologie, les grammaires font l'objet d'une recherche systématique. A ce titre, le travail très utile réalisé par Bernard Quémada pour les *Archives de la linguistique française* (*Grammaire de Port-Royal*, de Beauzée, *Dictionnaire de Bayle*, de Bescherelle, œuvres de du Marsais, d'Estienne, de Fromant, de Girard, *Traité de Meigret*, textes de Vaugelas, etc.) sera complété par des travaux plus contemporains.

Le principe général retenu pour le choix des éditions consiste à faciliter le travail génétique ou comparatif. Il est donc souhaitable de prendre des éditions homogènes, notamment les œuvres complètes les plus proches de l'édition princeps, la publication la plus récente faisant référence, et quelques éditions intermédiaires lorsque le travail d'établissement du texte ou l'apport critique présentent une originalité incontestable.

Pour les XVI^e et XVII^e siècles, on retiendra par exemple les œuvres de d'Aubigné, Baïf, Belleau, Bèze, Boaistuau, Bourdigné, du Bartas, du Bellay, Charron, Dolet, Flore, Forcadet, Jodelle, Joubert, Larivey, Marot, Marguerite de Navarre, Matthieu, Montaigne, Pasquier, Peletier du Mans, Rabelais, etc.

– *Le corpus philosophique*

Pour la philosophie, la démarche privilégiée, dans un premier temps, le choix des œuvres, pour constituer un corpus de référence d'environ 1 000 textes (couverture linguistique française et si possible bilingue). Le même souci de faciliter l'analyse comparative des textes sera appliqué, la question des traductions justifiant la mise à disposition de plusieurs versions de référence. Il y aura ensuite un élargissement du corpus par cercles successifs intégrant les auteurs par ordre chronologique : Abélard, Albert le Grand, Anselme, Aristote, Averroès, Avicenne, Basile le Grand, Bodin, Bruno, Buridan, Calvin, Cues, Descartes, Erasme, Ficin, Helvétius, Luther, Machiavel, Malebranche, Occam, Pascal, Platon, etc.

Les grandes revues généralistes seront également rassemblées autour des auteurs : *Allgemeine Zeitschrift für Philosophie*, *American Quarterly*, *Analecta Husserliana*, *Archives de philosophie*, *Critique*, *Filosofia*, *Kant Studien*, *Kierkegaardiana*, *Nietzsche Studien*, *Revue d'histoire et de philosophie religieuse*, *Revue de métaphysique et de morale*, *Philosophical Studies*, *Philosophie*, *Etudes philosophiques*, *Revue de synthèse*, *Revue philosophique de la France et de l'étranger*, *Russell*, *Journal of the Bertrand Russell Archives*, *Schopenhauer*

Jahrbuch, *Speculum*, *Studia Leibniziana*, *Philosophical Review*, etc.

– *Le corpus scientifique*

La réalisation d'un corpus numérisé en histoire des sciences se situe dans la continuité des principes généraux de politique documentaire de la bibliothèque. Le programme se définit ainsi : ouvrages scientifiques classiques, l'ensemble des textes de théoriciens scientifiques français et étrangers, en s'en tenant aux ouvrages parus avant la Seconde Guerre mondiale. La priorité sera donnée aux éditions des œuvres complètes, lorsqu'elles existent. Environ 400 titres ont été repérés jusqu'à présent :

– Antiquité : Anaximandre, Archimède, Euclide, Galien, Hippocrate, Pline l'Ancien, Ptolémée, Théophraste, etc.

– Moyen Âge / Renaissance : Al-Fakhri, Agricola, Belon, Benedetti, Brahe, Cardano, de Bouelles, Clavius, Copernic, Desargues, Fine, Gassendi, La Roche, Lefèvre d'Étaples, Maurolico, Paracelse, Ptolémée, Stevin, Tartaglia, Vinci, Zabarella, etc.

– XVII^e siècle : Aldrovandi, Auzout, Barrow, Bernoulli, Collins, Dampier, Fermat, Galilée, Hooke, Huyghens, Kepler, Leeuwenhoek, Leibniz,

Mariotte, Mersenne, Newton, Plumier, Roberval, Viète, Wallis, Wilkins, etc.

– XVIII^e siècle : d'Alembert, Berkeley, Birch, Buffon, Castel, Cavendish, Condorcet, Diderot, Euler, Franklin, Gravesande, Haller, Jussieu, Lacaille,

Lalande, Lambert, Lavoisier, Maclaurin, Maupertuis, Needham, Parmentier, Priestley, Réaumur, Riccati, Richter, Scheele, Sigaud de Lafond, Spallanzani, Varignon, Zanotti, etc.

– XIX^e siècle : Abel, Ampère, Arago,

Bernard, Berthelot, Berthollet, Boltzmann, Carnot, Cuvier, Darwin, Frege, Galois, Gauss, Gerhardt, Haüy, Henry, Humboldt, Lagrange, Lamarck, Laplace, Lucas, Montucla, Pélouze, Siegel, Verdet, Volta, Weierstrass, Würtz, Young, etc.

– XX^e siècle: Borel, Castan, Heisenberg, Born, Coste, Julia, Néel, Neumann, Schrödinger, Volterra, Kastler, Perrin, Yukawa, etc.

Pour la littérature secondaire, les biographies scientifiques, les études récentes sur un ou des aspects de l'oeuvre des auteurs précédemment sélectionnés, ainsi que les réflexions – historique et philosophique – de ces mêmes auteurs sur leur discipline ou sur les sciences et l'activité scientifique dans leur globalité : *Aperçu historique sur l'origine et le développement des mathématiques* de Michel Chasles, *Le Système du monde* de Pierre Duhem, *La Science et l'hypothèse* de Henri Poincaré.

Le fonds sera spécialisé en sciences dites « pures » : mathématiques, physique, chimie et sciences naturelles.

Enfin, les instruments de recherche seront, dans la mesure du possible, versés au fonds : *Dictionary of Scientific Biography* de Gillispie, *Biographisch-literarisches Handwörterbuch zur Geschichte des exakten Wissenschaften* de Poggenorff, *Catalogue of Scientific Papers* de la Royal Society.

– XIX^e-XX^e siècles : de Broglie, Cartan, Couturat, Curie, Daumas, Duhem, Einstein, Hadamard, Humbert, Jordan, Julia, Klein, Koyré, Lebesgue, Lévy, Mandelbrojt, Peirce, Picard, Poincaré, etc.

Dictionary of Scientific Biography de Gillispie, *Biographisch-literarisches Handwörterbuch zur Geschichte des exakten Wissenschaften* de Poggenorff, *Catalogue of Scientific Papers* de la Royal Society.

Seront considérés également comme usuels les grandes séries transdisciplinaires comme les *Comptes rendus* et *Mémoires de l'Académie des sciences*, *Acta Eruditorum Lipsiensiae*, *Philosophical Transactions of the Royal Society*...

– Ethnologie et anthropologie

Pour l'ethnologie et l'anthropologie, il s'agit de rassembler dans un premier temps tous les textes des auteurs importants, généralistes ou non, quelle que soit l'aire géographique concernée, et de constituer progressivement une collection de revues internationales, nationales et régionales, afin de parvenir à la couverture la plus systématique possible (*Africa*, *American Anthropologist*, *American Ethnologist*, *Annual Review of Anthropology*, *Anthropological Quarterly*, *Anthropologie et sociétés*, *Anthropos*, *Cahiers de littérature orale*, *Current Anthropology*, *Droit et cultures*, *l'Ethnographie*, *Ethnologie française*, *Ethnology*, *Etudes corses*, *Etudes rurales*, *Gradhiva*, *l'Homme*, *Journal de la Société des américanistes*, *Journal de la Société des océanistes*, *Journal des africanistes*, *Journal des anthropologues*, *Man*, *Métis*, *le Monde alpin et rhodanien*, *Oceania*, etc.).

– Autres corpus

Enfin, en histoire, science politique, économie, droit, les nombreux contacts et les séances de travail organisées avec les chercheurs ont permis de mieux cerner des besoins très spécifiques, notamment en matière d'outils bibliographiques pour les historiens. Pour ce qui concerne le corpus en cours de constitution, il va s'agir dans un premier temps de rassembler tous les auteurs considérés aujourd'hui comme des classiques dont l'oeuvre intéresse un public beaucoup plus large que celui de leur strict champ disciplinaire : Adorno, Aron, Benjamin, Bentham, Blanc, Blanqui, Bloch, Bossuet, Buonarrotti, Chartier, Chastellain, Clausewitz, Comynnes, Comte, Condorcet, Croce, Eginhard, Febvre, Froissart, Fustel de Coulanges, Gibbon, Grégoire de Tours, Guizot, Hérodote, Hotman, Jaurès, Joinville, Labrousse, Lavisce,

Mabillon, Mably, Maistre, de la Marche, Michelet, Mignet, Mommsen, Pareto, Poincaré, Polybe, la Popelinière, Quesnay, Quinet, Reclus, Ranke, Renan, Renouvier, Saint-Simon, Seignobos, Sieyès, Sorel, Taine, Thiers, Thierry, Thucydide, Tocqueville, Volney, Weber...

Rassembler les documents

Il reste donc à réunir les documents repérés dans des ensembles pertinents, en exploitant toutes les ressources existantes : livres disponibles – neuf et ancien –, catalogues de microfiches et microfilms de fournisseurs français et étrangers, prêts des bibliothèques, dons, etc.

Des marchés ont été conclus avec deux libraires (Decitre à Lyon et la librairie des PUF⁵ à Paris) pour l'acquisition d'environ 35 000 livres dans les deux années à venir.

La coordination de ce programme avec le plan de reproduction photographique que l'EPBF mène avec la Bibliothèque nationale va permettre de croiser des préoccupations communes. Les besoins de la numérisation fondés sur des bases intellectuelles pourront compléter et enrichir des opérations réalisées selon des critères de stricte sauvegarde des documents.

Le partenariat avec d'autres institutions spécialisées devrait permettre de rassembler rapidement des collections de référence. La coopération avec les Archives de l'Académie des sciences, service qui développe également un projet de numérisation, a facilité l'accès à des ouvrages difficilement communicables, les *Comptes-rendus des séances* ou disponibles en peu d'exemplaires, les *Mémoires des savants*. Des échanges sont également en cours avec la bibliothèque de l'École polytechnique.

De même, des conventions avec les organismes disposant de bases textuelles informatiques auront pour but d'abonder le fonds initial de la bibliothèque. L'accord de partenariat avec

5. PUF : Presses universitaires de France.

l'INALF⁶ en est le meilleur exemple : il aboutira à la mise à disposition de l'utilisateur du PLAO, non seulement de toutes les œuvres rassemblées dans Frantext, mais également du *Trésor de la langue française*, l'équivalent du *British Oxford Dictionary*, dont deux versions circulent aujourd'hui sous forme de CD-ROM. L'existence d'un tel corpus, alimenté éventuellement par d'autres corpus linguistiques émanant de laboratoires associés, constituera à terme un fonds unique au monde sur la langue française. Leur disponibilité encouragera des enrichissements successifs et autorisera des échanges avec le réseau des institutions concernées. Enfin, l'une des ultimes ressources mais non des moindres dont pourra disposer le fonds électronique de la bibliothèque viendra des prêts ou dépôts des bandes de photocomposition des éditeurs. Les efforts de standardisation des langages de structuration réalisés par les professionnels du livre laissent espérer que toutes les difficultés rencontrées jusqu'à présent pour la récupération de l'information saisie dès sa création sous forme numérique seront rapidement aplanies – sur le plan technique.

Relations avec les éditeurs et les auteurs

Le programme de numérisation – dans la mesure où cette technique entraîne de fait une reproduction des œuvres – implique l'agrément des ayants droit ainsi que la mise au point de procédures de protection et de rémunération des auteurs et des éditeurs. Toutefois, les dispositions à prendre ne doivent pas remettre en cause les principes garantissant aux chercheurs l'accès aux documents.

Les négociations menées avec le syndicat national de l'édition et les sociétés représentant les auteurs (SGDL, SCAM, SADC, SPADEM, ADAGP, SDI)⁷ et la SCELFF⁸ ont abouti, au mois de novembre 1991, à la signature de deux protocoles définissant les modalités du partenariat jusqu'à l'ouverture de la nouvelle

bibliothèque. Il y est affirmé l'accord de principe donné par ces organismes représentatifs à l'EPBF pour la numérisation des livres dans le respect des règles de la propriété littéraire. Toutefois il ressort de ces accords que la négociation doit être concrétisée par la signature de conventions spécifiques avec chaque maison d'édition concernée et chacune des sociétés collectives de gestion des droits d'auteur, et pour chaque titre qui sera numérisé.

C'est au sein de comités de liaison que sont évoquées toutes les questions à traiter pendant la phase expérimentale, c'est-à-dire jusqu'à l'ouverture de la bibliothèque. Les négociations portent sur la numérisation elle-même, assimilée à une reproduction, sur la consultation qui donnera éventuellement lieu à rémunération – au titre du droit de représentation – et sur la reproduction, pour laquelle existe la jurisprudence en matière de photocopie.

Les sociétés représentant les auteurs ont contribué à faciliter le travail de réflexion conduit avec l'établissement sur des pratiques innovantes pour lesquelles il n'existe pas d'expérience dans les bibliothèques françaises. De leur côté, la majeure partie des maisons d'édition directement concernées (éditeurs d'érudition, de littérature et de sciences humaines) a accepté de participer à l'expérience dans le cadre contractuel qui lui était proposé.

Le chantier

L'objectif fixé pour l'inauguration est donc de 100 000 volumes, qui proviendront en partie de livres directement numérisés par l'établissement constructeur et, pour une moindre part, de fonds numériques hérités d'organismes extérieurs.

6. INALF : Institut national de la langue française.

7. SGDL : Société des gens de lettres. SCAM : société civile des auteurs multimédias. SADC : Société des auteurs et compositeurs dramatiques.

SPADEM : Société des auteurs des arts visuels. ADAGP : Société des auteurs dans les arts graphiques et plastiques.

SDI : Société de l'image.

8. SCELFF : Société civile de l'édition littéraire française.

Pour la production des livres électroniques, il est fait appel au concours de sous-traitants. Au cours de l'année 1991, des tests ont été réalisés permettant d'apprécier la faisabilité de la numérisation à grande échelle et de juger de la qualité de la production à partir de microformes. 3 000 micro-fiches appartenant à la Bibliothèque nationale et 4 000 ouvrages achetés en librairie ont ainsi été numérisés. Il a été décidé de produire 40 000 titres à partir de volumes papier et l'équivalent de 60 000 titres à partir de microformes. Un appel d'offres international ouvert en 1992 a permis de sélectionner deux sociétés qui utiliseront les deux filières de numérisation. L'équipe responsable de la réalisation de ce programme comprend quinze personnes à plein temps rattachées au département de l'informatique et des nouvelles techniques, installées sur le site d'Ivry. Elle a en charge l'ensemble des tâches nécessaires à la constitution de la collection (travail bibliographique, contacts avec les universités et l'édition, relations avec la recherche, gestion du fonds électronique, commandes aux libraires, négociations avec les ayants droit, etc.) et au suivi de production (choix des prestataires, tests, contrôle qualité, gestion des stocks, normalisation, etc.). Elle assure également le pilotage d'études complémentaires portant sur la structuration du fonds et le balisage des documents, sur la récupération des bandes de photocomposition, etc. L'idée, raisonnable, d'offrir aux chercheurs, à l'ouverture de la nouvelle bibliothèque, un fonds qui leur permettrait une fréquentation fructueuse d'un nombre de textes forcément limité ne donnera pas satisfaction à tous ceux qui rêvaient d'un déménagement purement immatériel des collections de la Bibliothèque nationale. L'ambition qui consiste à réaliser aujourd'hui le « cœur » de la collection suggère bien cependant que ce projet ne prendra tout son sens et ne connaîtra sa pleine rentabilité que s'il est ensuite progressivement et rapidement élargi, enrichi en fonction des besoins qu'il ne manquera pas de susciter et des virtualités qui resteront à développer.

Février 1993

BIBLIOGRAPHIE

1. Cahart, Patrice, Melot, Michel, *Propositions pour une grande bibliothèque*, Paris, La Documentation française, 1990.
2. Bolter, Jay David, *Writing Space, The Computer, Hypertext, and The History of Writing*, University of North Carolina, 1991.
3. Chahuneau, François, Lécluse, Christophe, Stiegler, Bernard, Virbel, Jacques, « Prototyping the Ultimate Tool for Scholarly Qualitative Research on Texts », *Screening Words : Users Interfaces for Text, 8th Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary and Text Research*, Waterloo, Oct. 1992.
4. Chartier, Roger, *L'Ordre des livres*, Paris, Alinéa, 1991.
5. Giffard, Alain, « L'invention d'un médium ? » (Culture et éducation), *La France à l'exposition universelle*, Paris, Flammarion, 1992.
6. Laufer, Roger, Scavetta, Domenico, *Texte, hypertexte, hypermédia*, Paris, PUF, 1992, Coll. Que sais-je ?
7. Neuman, Michaël, « The Very Pulse of the Machine : Three Trends Toward Improvement in Electronic Versions of Humanities Texts », *Computers and the Humanities 25*, Kluwer Academic Publishers (Netherlands), 1991.
8. Stiegler, Bernard, « Machines à lire », *Autrement (La Bibliothèque)*, 1991.
9. Virbel, Jacques, *Etude technique d'assistance à l'élaboration du projet de la station de travail de la Bibliothèque de France. Rapport d'étude*, BDF, juin 1991.
10. Vuillemin, Alain, *Informatique et littérature (1950-1990)*, Paris, Champion-Slatkine, 1990.