

# La bibliothèque numérique européenne

## Une stratégie culturelle de la Toile

**F**ace aux défis lancés par les moteurs de recherche tels que Google, relayés par la presse sensibilisée à la suite du débat public ouvert par Jean-Noël Jeanneney<sup>1</sup>, le président Jacques Chirac lance l'idée d'une bibliothèque numérique européenne. Le ministre de la Culture demande dès le 30 août 2005 d'établir un état des lieux précis et les contours d'une contribution française à un projet de bibliothèque numérique européenne (BNUE). Cinq groupes de travail réunissent des acteurs du privé (représentants du monde de l'édition, des secteurs industriels de la numérisation et des moteurs de recherche) et du secteur public (Culture, Éducation nationale, Recherche, Industrie, Finances). Renaud Donnedieu de Vabres préside le comité de pilotage. Le président de la BnF en assure la vice-présidence.

### Valérie Tesnière

Bibliothèque nationale de France  
VALERIE.TESNIERE@bnf.fr

### Noémie Lesquins

Bibliothèque nationale de France  
NOEMIE.LESQUINS@bnf.fr

Parallèlement, pour informer de l'état de la réflexion en France, des contacts sont pris tant avec la Commission européenne, qui organise la consultation « i 2010 » sur ce sujet, qu'avec la présidence de l'Union européenne. La réunion des ministres de la Culture des 17 et 18 novembre 2005 confirme l'intérêt d'un certain nombre d'autres pays.

Un livre blanc a été remis le 11 janvier 2006. Il propose une analyse stratégique du positionnement de la BNUE, des plus-values attendues, des contenus, des acteurs concernés en priorité et des scénarios d'organisation et d'architecture possibles. Un plan d'actions prioritaires a été présenté en conseil des ministres. Ces deux textes, qui émanent du secrétariat général du comité de pilotage

de la BNUE, sont synthétisés dans l'article qui suit<sup>2</sup>.

### État des lieux: les évolutions récentes d'Internet

Dix ans après la mise en place des premiers réseaux, Internet est devenu en Europe le troisième média favori des jeunes Européens, derrière la télévision et la radio, et devant la presse<sup>3</sup>. En France, le taux de pénétration atteint presque la moitié de la population, et la baisse des tarifs du haut débit laisse présager une évolution exponentielle. Les principaux usages d'Internet sont la communication interpersonnelle, de l'*e-mail*

1. Jean-Noël Jeanneney, *Quand Google défie l'Europe: plaidoyer pour un sursaut*, Paris, Mille et une nuits, 2005.

2. Fait partie du secrétariat général du comité de pilotage de la BNUE Valérie Tesnière, Noémie Lesquins et Alexandre Moatti.

3. « Enquête de l'European interactive advertising association », *Le Monde*, 24 juin 2005.

Archiviste-paléographe, **Valérie Tesnière** dirige le Département de la coopération (en charge des pôles associés, de Gallica et du CCFr) à la Direction des services et des réseaux de la Bibliothèque nationale de France. Elle a été précédemment directrice du département Histoire et chef de projet Politique documentaire à l'EPBF (Établissement public Bibliothèque de France). Auteur de plusieurs articles dans le BBF, elle a collaboré à La Bibliothèque nationale de France: collections, services, publics (Éd. du Cercle de la librairie, 2001) et a publié Le Quadrige: un siècle d'édition universitaire, 1860-1968 (Presses universitaires de France, 2001).

Archiviste-paléographe et conservateur de bibliothèque, **Noémie Lesquins** a commencé à la BnF comme chargée de mission pour les pôles associés, en particulier les bibliothèques du dépôt légal imprimeur. Elle travaille actuellement au Département de la bibliothèque numérique sur les questions d'accès et d'interfaces. Auteur de l'entrée sur la BnF dans l'Encyclopedia of Library and Information Science (Marcel Dekker, 2003), elle est coauteur avec Caroline Wiegandt et Mireille Bousquet d'un article sur la politique de coopération nationale de la BnF (BBF, 2003, n° 2).

à la téléphonie en ligne, en passant par l'explosion récente des blogs. Favorisé par la sécurisation des sites, la généralisation du haut débit et la mise en place récente de solutions de micropaiement, le commerce en ligne se développe fortement tandis que s'affirme une attente de plus en plus forte de contenus et services gratuits en ligne et une pratique généralisée du téléchargement.

Les moteurs de recherche se sont imposés comme les nouveaux médiateurs, points de passage obligés pour localiser et sélectionner l'information dans une masse informe et multiforme à la fois, en forte croissance. Ainsi, pour l'Hexagone, 18,8 millions d'internautes, soit 9 sur 10, ont consulté au premier trimestre 2005 un moteur de recherche tel que Google, Voilà, Yahoo!, MSN<sup>4</sup>. Ces outils sont considérés désormais par les étudiants comme une meilleure source d'information que les médias traditionnels<sup>5</sup>. Les usages observés du web général vont du mode de recherche simple, intuitif et plein texte au téléchargement, en passant par un besoin de certification de la source

4. La Tribune, 14 novembre 2005.

5. Étude de Yahoo! Search Marketing, automne 2005.

et de mise à disposition du document dans sa forme intégrale pour un usage personnalisé.

Dans cet ensemble, le web documentaire, constitué de bases et de documents certifiés et organisés par une autorité intellectuelle (éditeur, bibliothèque, centre de documentation ou d'archives), fait l'objet d'un usage très spécifique<sup>6</sup>. Si ces usages tels qu'ils se présentent actuellement sont le reflet d'une pratique encore minoritaire dans la population, à l'exception des journaux en ligne, il faut retenir que les outils généraux du web et des moteurs de recherche comme Google formatent les attentes du public.

Les objectifs des trois grands moteurs de recherche Google, Yahoo!, MSN sont clairs: d'une part, fidéliser des utilisateurs toujours plus avides d'information et de convivialité virtuelle; d'autre part, augmenter les flux en direction de leurs sites où les espaces publicitaires constituent le plus gros du chiffre d'affaires (85 % du chiffre d'affaires de Yahoo! en 2005). La presse quotidienne en ligne est pour eux un enjeu majeur, parce que la maîtrise des contenus en ligne et l'intérêt des ressources publicitaires sont centraux dans leur stratégie<sup>7</sup>.

Le secteur des documents de type patrimonial est aussi stratégiquement sensible. Il encourage une politique de communication planétaire et fa-

6. Le public de la bibliothèque numérique de la Bibliothèque nationale de France, Gallica, rejoint celui des sites spécialisés de la Documentation française: issu des classes socioprofessionnelles supérieures, il est généralement doté d'un diplôme de troisième cycle, et est constitué principalement de chercheurs, étudiants, professeurs, amateurs, professions libérales, associations, eux-mêmes médiateurs de l'information.

7. Les travaux du Geste (Groupement des éditeurs de services en ligne) ont récemment attiré l'attention des pouvoirs publics sur les enjeux d'une stratégie nationale, sinon européenne pour l'avenir de la presse d'information. Google, dans son interface Google News, agrège les contenus de l'ensemble de la presse internationale sans souci de marquage de la provenance et remonte les sites non payants moins crédibles. Pour se démarquer, un concurrent comme Yahoo!, qui se revendique comme une société de média et non pas seulement comme un agrégateur de contenus, embauche actuellement les meilleurs journalistes économiques aux États-Unis.

vorise le développement de parts de marché sur d'autres secteurs plus rentables (commerce ou information en ligne). C'est donc dans une logique d'intégration verticale et de maîtrise des maillons de la chaîne des contenus que les moteurs s'intéressent à la diffusion et à la numérisation du patrimoine culturel.

Du point de vue de la propriété intellectuelle, les coups de boutoir de Google, au nom du *fair use*, mettent en avant une conception différente: ce n'est plus « *tout ce qui n'est pas explicitement autorisé par les ayants droit est interdit* », mais au contraire, « *tout ce qui n'est pas explicitement interdit est autorisé* ». La mise en ligne de tout ou partie d'œuvres sous droits par Google ou Amazon soulève en effet la question de la sécurisation des données numériques et fait planer la menace du piratage et du téléchargement illicite.

### Internet et les acteurs de la chaîne de l'écrit

Dans le contexte général du développement de la Toile, comment réagissent les différents acteurs de la chaîne de l'écrit?

### Journaux et revues

Le secteur de la presse quotidienne française a depuis longtemps investi Internet en proposant des produits en ligne, mais les retards par rapport aux Anglo-Saxons s'expliquent notamment par le fait qu'en France, la publicité sur Internet n'est pas encore une source majeure de revenus comme aux États-Unis.

Par ailleurs, si la presse n'a pas complètement rentabilisé ses modèles économiques dans la diffusion en ligne, elle a su régler en large partie la difficulté technique majeure sur laquelle butent encore beaucoup d'éditeurs de livres, avec notamment Internet Plus, système sûr, qui favorise le micropaiement, et elle a négocié avec les fournisseurs d'accès Internet de façon avantageuse.

La mise en ligne de l'ensemble de leurs archives imprimées n'est enfin pas prioritaire pour les journaux, qui observent une concentration des demandes d'usagers sur la dernière année parue. L'importance du plan de numérisation de la presse rétrospective nationale, lancé par la BnF en 2005, ouvre de ce point de vue des perspectives de collaboration avec les organes de presse. Quatre premières conventions ont ainsi été signées avec *Le Figaro*, *La Croix*, *L'Humanité* et *Ouest-France* pour aboutir à une mise en ligne de ces journaux, de leur fondation jusqu'à une date définie en concertation (antérieure à 1940). La jonction entre ce qu'ils mettent en ligne sur leurs propres sites et ce qui sera proposé en ligne dans Gallica à compter de 2007 constitue l'étape suivante.

En France, tandis que les éditeurs spécialisés (sciences, droit...) ont déjà une politique affirmée de diffusion en ligne de leurs contenus et que les éditeurs d'ouvrages de référence (dictionnaires) et en partie de sciences humaines s'y mettent grâce à la synergie créée dans les grands groupes, les éditeurs de littérature générale considèrent avec circonspection l'enjeu de l'édition en ligne. Le modèle économique n'est en effet pas encore trouvé et les bibliothèques de textes en ligne telles que Questia aux États-Unis, pourtant fortes d'un marché universitaire et scolaire anglophone autrement supérieur au marché francophone par exemple, ou, à plus petite échelle, en France, les éditions hypertextuelles de Champion, conçues comme de la philologie numérique, s'en tiennent à un public académique circonscrit.

Internet est perçu comme un champ exploratoire de nouvelles pratiques impliquant des produits encore à inventer, mais surtout comme un outil complémentaire, pour la promotion en ligne des catalogues. Sur un plan juridique, la nécessaire renégociation des droits de diffusion en ligne est l'un des freins identifiés. D'autre part, les questions soulevées par les œuvres orphelines (dont les ayants droit sont

inconnus) et des œuvres sous droits, non disponibles dans le commerce (rassemblées sous le terme de zone grise par François Stasse dans son rapport)<sup>8</sup>, ont suscité, du moins en France, des inquiétudes quant à d'éventuels manquements au respect des prérogatives des ayants droit..

Du côté des revues, le passage à l'édition en ligne est sinon une réalité pour les sciences exactes, du moins une mutation très avancée et inéluctable pour les sciences humaines et sociales. En France, différentes initiatives ont vu le jour récemment : trois portails sont offerts pour les revues de sciences humaines, en sus de l'initiative plus ancienne de « revues.org<sup>9</sup> » : le projet Adonis<sup>10</sup> porté par le CNRS, le programme Persée<sup>11</sup> du ministère de l'Éducation nationale et l'initiative d'un groupe d'éditeurs spécialisés, Cairn<sup>12</sup>.

Ce préalable de la protection des ayants droit fait encore obstacle aux avancées collectives et aux réponses concertées face aux enjeux des moteurs. En France, le secteur de la presse se distingue à cet égard de celui du livre. En Europe, l'initiative la plus intéressante est celle du Börsenverein. L'équivalent du Cercle de la librairie a lancé un programme visant à assurer l'indépendance des éditeurs allemands, c'est-à-dire prévoyant la mise en place d'une plate-forme locale de gestion des données (*Volltextsuche online*), mais également la multi-accessibilité des données *via* le plus grand nombre de moteurs possibles (Google, Yahoo!, MSN...), tirant ainsi profit des services offerts par ces sociétés sans pour autant perdre la main sur les contenus. Il est à noter que ce projet associe les libraires, conformément à une tradition forte outre-Rhin et prévoit un partenariat avec la Deutsche Bibliothek.

8. François Stasse, « Rapport au ministre de la Culture et de la Communication sur l'accès aux œuvres numériques conservées par les bibliothèques publiques », avril 2005.

9. <http://www.revues.org/>

10. <http://info.cens.cnrs.fr/article8.html>

11. <http://www.persee.fr/>

12. <http://www.cairn.info>

## Bibliothèques

Du côté des bibliothèques, l'irruption du web a d'abord focalisé les bibliothèques nationales ayant des responsabilités en matière de dépôt légal sur la problématique de sa collecte, de sa communication et de sa conservation pérenne. En France, la loi sur les droits d'auteur et les droits voisins dans la société de l'information constitue un pas décisif dans la prise en compte de cette mutation<sup>13</sup>.

C'est un fait que nombre de bibliothèques nationales n'ont pas pu ni su créer une offre numérique suffisante à partir de leurs collections au cours de la dernière décennie. Cela vaut également pour de nombreuses bibliothèques académiques ou patrimoniales, où le bilan est inégal.

Ont une avance quantitative la Library of Congress et la BnF Gallica constitue même une exception remarquable avec 80 000 documents en ligne, libres de droits, et autant d'images, très consultés. La Belgique vient d'adopter un plan de financement pour combler son passif. L'Allemagne est dans une situation particulière puisque les fonds patrimoniaux sont répartis sur le territoire (Munich, Göttingen ont engagé des programmes), tandis que la Deutsche Bibliothek est centrée sur la collecte et le signalement de l'équivalent du dépôt légal. L'Italie est dans une situation voisine. L'Espagne a démarré un projet, avec la Biblioteca nacional de España, comme les Pays-Bas, l'Autriche et la Hongrie. La British Library vient de signer avec Microsoft un contrat qui prévoit la numérisation dans des délais brefs de 100 000 documents de toute nature libres de droits (mécénat de 2,5 millions de dollars). Le consortium de bibliothèques JISC (Joint Information Systems Committee), qui soutient les programmes de numérisation en Grande-Bretagne, vient aussi d'engager un plan de

13. Cette loi, dite loi Dadvsi, a été votée par l'Assemblée nationale le 21 mars 2006, et doit être discutée au Sénat (Ndlr).

numérisation pour un montant de 16 millions de livres.

Aux États-Unis, le consortium OCA (Open Content Alliance) rassemble des entreprises de l'Internet et de l'informatique (Yahoo!, Adobe, Hewlett Packard, Internet Archive), des bibliothèques, des centres d'archives et des éditeurs, dans le but de numériser et mettre en ligne un entrepôt de données numériques de tout type accessibles universellement, multilingues, pérennes, composés de textes numérisés et de contenus multimédias de toute origine. MSN (Microsoft) a rejoint ce consortium.

Au-delà des annonces, qu'est-ce qui sous-tend cette organisation qui essaie d'avoir des filiales en Europe? Le maître mot est l'*Open content and access*. Tout ce qui permet d'ajouter des moyens d'accès supplémentaires est permis et même encouragé. Il n'y a pas d'essai d'organisation de l'information: il faut être partout, au besoin en multipliant les doublons. Chaque partenaire met ce qu'il veut<sup>14</sup>. OCA se situe donc dans une logique d'entrepôt numérique. Même si ce consortium affirme, à la différence de Google, son respect des ayants droit, les éditeurs sont actuellement leur point faible, car l'accès, tel qu'il y est conçu, est trop ouvert pour l'édition commerciale, qui ne s'y ralliera que si le contrôle s'organise.

Rappelons que, pour le moment, MSN et Yahoo!, dans leur stratégie avec les bibliothèques, ne prennent en considération que les fonds patrimoniaux libres de droits des bibliothèques.

### Les principales initiatives de numérisation culturelle

En France, le premier Plan national de numérisation, piloté par la Mission recherche et technologies au ministère de la Culture, s'est efforcé d'imposer, moyennant incitation finan-

cière, des standards techniques aux musées, archives et bibliothèques. À côté de la Bibliothèque nationale de France (Gallica) et de quelques bibliothèques en nombre limité, les archives ont été particulièrement

**La stratégie, pour la BNUE, consiste à développer au mieux les apports, à savoir les marques, les contenus et le savoir-faire de chacun des partenaires principalement concernés, en ne s'interdisant pas des alliances opportunistes avec les moteurs sur des opérations ciblées**

actives en matière de numérisation et de mise en ligne de fonds patrimoniaux, grâce, entre autres, au soutien des collectivités locales (cadastre, état civil), effort relayé au niveau central. Mais l'articulation d'une numérisation à des fins de sauvegarde du patrimoine et d'une numérisation à but de diffusion culturelle n'a pas été toujours clairement définie: pour des raisons techniques et juridiques, le patrimoine culturel en ligne n'est pas toujours à la hauteur des ambitions politiques.

Ayant le souci de faire connaître les fonds numérisés grâce à un outil en ligne, le ministère de la Culture a développé Michael (Multilingual Inventory of Cultural Heritage in Europe)<sup>15</sup> en partenariat avec d'autres pays européens. C'est un répertoire européen de l'ensemble des fonds

numérisés relevant du domaine culturel patrimonial, à vocation multilingue. De ce fait, il n'est pas centré sur l'écrit. Il recense les projets et réalisations d'institutions culturelles telles que les musées et les archives. Il n'y a pas d'accès prévu aux textes ni aux corpus numérisés dans leur intégralité. Cette fonctionnalité peut être éventuellement développée dans une étape ultérieure.

En Europe, le programme The European Library (TEL)<sup>16</sup>, soutenu par la Conference of European National Libraries (CENL) est parti de l'idée d'une réunion de catalogues. C'est dans un second temps qu'il a repris le concept de mise en commun de fonds numérisés des bibliothèques nationales imaginé dans le cadre du projet Bibliotheca universalis. L'objectif est de donner accès en ligne au patrimoine écrit numérisé par 45 grandes bibliothèques européennes. Le service a été lancé en mars 2005 et donne actuellement accès aux catalogues et bibliothèques numérisées de 12 bibliothèques nationales. TEL reste encore souvent méconnu des internautes. Il offre l'avantage important d'une expérience de projet commun à plusieurs bibliothèques nationales européennes et pourrait jouer aussi un rôle dans la réalisation d'une BNUE selon des modalités qui restent à explorer.

### Une stratégie pour la BNUE

L'attente est forte quant à ce que pourrait être aujourd'hui une véritable stratégie culturelle du web dont la bibliothèque numérique européenne serait une des fondations. Il importe de bien cerner la différence et l'intérêt d'une telle initiative par rapport aux entrepôts numériques constitués et indexés par les puissants moteurs de recherche outre-Atlantique.

Sur un plan technique, les grands moteurs de recherche ont acquis une avance en termes de puissance d'in-

14. Il a été annoncé un objectif de 80 % des livres publiés entre 1923 et 1964 et qui sont libres de droits.

15. <http://www.michael-culture.org>

16. <http://www.theeuropeanlibrary.org>

terrogation et de volumes indexés, qu'il serait vain de chercher à concurrencer à l'identique. Le but n'est pas de faire autant, mais différemment. Car il existe un potentiel industriel en France et en Europe, avec des résultats probants, qui peut être encouragé à condition de cibler le domaine d'intervention aux réelles plus-values d'une BNUE.

Concernant les moteurs, la stratégie, pour la BNUE, consiste à développer au mieux les apports, à savoir les marques, les contenus et le savoir-faire de chacun des partenaires principalement concernés, en ne s'interdisant pas des alliances opportunistes avec les moteurs sur des opérations ciblées. Les « marques » sont présentes chez les éditeurs (prestige et certification de provenance ; droits), mais aussi dans les bibliothèques, qui sont garantes de la légitimité des sources utilisées. Les « contenus » sont majoritairement dans les bibliothèques (présence physique du support papier), dans certains cas chez les éditeurs (collections papier souvent partielles mais présence de plus en plus des documents natifs numériques). Le « savoir-faire » est inégalement réparti. Il faut mettre en synergie le meilleur de chacun des partenaires :

- compétences en matière de normalisation, expérience et expertise dans le domaine de la numérisation, responsabilités durables en matière d'archivage pérenne des données numériques pour les institutions publiques comme les grandes bibliothèques ;
- maîtrise de certains modèles économiques pour le secteur de la presse en ligne, d'une partie de l'édition spécialisée et des sociétés de recouvrement des droits ; compétences techniques en matière d'édition électronique pour le secteur de l'édition de référence ;
- plus-values sur les moteurs de recherche de nouvelle génération et techniques d'ingénierie pour les éventuels partenaires industriels.

Le président de la Commission européenne, José Manuel Barroso,

soutenu par un certain nombre de pays, est favorable au nouveau concept de bibliothèque numérique européenne, projet politique dont il reste à préciser les contours au-delà de l'accord de principe sur la nécessité d'une plus grande mise à disposition sur la Toile du patrimoine culturel européen. La Commission européenne a rappelé, par la voix de Vivian Reding, qu'elle continuerait à encourager le développement technique d'outils ou de formats utiles à l'ensemble des pays européens. Dans le nouveau contexte de bibliothèque numérique européenne, elle a lancé un appel à idées, dont les réponses ont été rendues fin janvier 2006 (i2010 Digital Libraries)<sup>17</sup>.

La question de savoir si l'on a recours à des solutions de type fédéré ou de type centralisé n'est pas tranchée. L'analyse de l'articulation de plates-formes nationales avec une plate-forme européenne conduit à une appréciation nuancée des choix organisationnels préalables de la BNUE. En replaçant le projet de BNUE dans un tel contexte, il semble d'autant plus impératif de composer avec les principaux partenaires et leurs sensibilités pour mieux faire prévaloir les enjeux essentiels.

La numérisation et la communication en ligne de manière massive de documents sous droits seraient facilitées par un régime d'exception au droit de la propriété intellectuelle, mais toute évolution législative en ce sens porterait atteinte aux intérêts légitimes des ayants droit (auteurs et éditeurs), pour lesquels un niveau de protection élevé a été récemment réaffirmé sur le plan international. Les orientations à privilégier sont donc prioritairement le recours au contrat et à la gestion collective volontaire. Le principe premier est donc que tout document sous droits consultable en ligne est soumis à rémunération pour les ayants droit, ce

que vient de rappeler la Federation of European Publishers dans la réponse qu'elle a adressée à la Commission européenne pour i2010.

La BNUE n'a pas pour objet d'organiser une diffusion exclusive des contenus. Dans le contexte d'Internet, les partenaires garderont aussi leurs propres interfaces de consultation dont les plus-values seront différentes, soit très spécialisées dans le cas de communautés de chercheurs, soit parce que les bibliothèques, les institutions culturelles ou les partenaires privés voudront continuer à animer leurs propres réseaux ou sites, dont les objectifs ne seront pas totalement identiques à ceux de la BNUE.

En résumé, les gages d'efficacité sont de :

- définir les principales attentes et plus-values de la BNUE pour les internautes comme pour les acteurs concernés de la chaîne du livre ;
- solidariser autour de l'objectif de la BNUE le secteur des bibliothèques et celui de l'édition en ligne, qu'il s'agisse de la presse ou de l'édition au sens plus classique, et assigner à chacun son rôle en fonction de son cœur de métier ;
- mobiliser le potentiel d'innovation technologique européen sur ce qu'il y a d'applicable à la BNUE, notamment l'analyse intelligente des contenus en ligne et le multilinguisme ;
- alléger le plus possible les contraintes administratives pour obtenir plus de rapidité dans la mobilisation des financements et dans la réalisation des programmes ou dans le choix des outils ;
- en déduire les scénarios d'organisation politique, sans présupposés d'architecture technique, en ayant pour but d'attirer les participants majeurs ;
- tracer des étapes de mise en œuvre qui, dans un premier temps, se centreront sur les bibliothèques nationales (plans de rattrapage à divers niveaux) et les participants volontaires dans le monde de l'édition, mais qui tiendront compte des situations nationales et qui n'excluront pas, dans

17. [http://europa.eu.int/information\\_society/activities/digital\\_libraries/consultation/replies/index\\_en.htm](http://europa.eu.int/information_society/activities/digital_libraries/consultation/replies/index_en.htm)

les solutions techniques retenues, d'autres partenaires à terme, tels que les archives, dont les liens avec l'écrit sont consubstantiels.

### Attentes et plus-values

Que peut apporter aujourd'hui une contribution européenne de bibliothèque numérique aux produits existant sur le marché (Google, OCA avec Yahoo! ou MSN)? Où se situeraient son originalité et sa valeur ajoutée pour les acteurs de la chaîne du livre concernés, principalement bibliothèques et éditeurs, en sorte qu'ils choisiraient la BNUE plutôt qu'une alliance avantageuse avec ces produits désormais qualifiés pour répondre aux attentes des internautes autant que des professionnels?

#### Le point de vue des internautes

La demande politique est claire: un outil d'abord grand public, mais qui sera grandement utile aussi à la recherche. L'analyse des usages montre que, si les usages « généralistes » du web ne sont pas actuellement centrés sur l'offre de textes en ligne, caractérisée par la dispersion, on ne peut plus attendre. C'est le matériau utilisable par les nouvelles générations, formatées de plus en plus par les outils du marché, qui doit être élaboré maintenant. L'offre conditionnera les usages futurs de l'écrit, en ligne et sur papier.

Si l'on regarde de plus près la demande sociétale, elle doit répondre à quelques critères minimaux: ouverture sur le contemporain, attractivité du produit, rapidité de consultation. Mais, plus que la quantité de documents numérisés qui resteront toujours accessibles par le biais des moteurs existants, les plus-values nouvelles attendues se situent surtout au niveau de services conformes à l'essentiel de la demande. C'est ici que l'analyse de ce qui intéresse certains usages de recherche devrait coïncider avec les attentes d'un plus large public.

### Une interface commune d'accès et de consultation de textes numérisés libres de droits et sous droits

L'accès aux documents contemporains demeure une très forte attente: une bibliothèque numérique européenne constituera la première des réponses pour élargir l'accès au patrimoine culturel écrit, à condition qu'elle prenne en compte à la fois les fonds libres de droits et ceux sous droits, en respectant le droit de la propriété intellectuelle.

Si, à partir d'une même interface d'interrogation, les internautes peuvent localiser les documents numérisés, quel que soit leur statut, si les critères d'affichage sont hiérarchisés

### La demande politique est claire: un outil d'abord grand public, mais qui sera grandement utile aussi à la recherche

et si, dans la foulée, ils peuvent consulter, soit librement, soit moyennant paiement, les textes de leur choix, nul doute que le progrès politique et technique apporté par une telle solution sera considérable<sup>18</sup>.

Ce n'est pas une interrogation par fonds comme le propose Michael avec des liens sur les sites des établissements, mais par titre ou par auteur. Une telle interface de localisation donne tout son sens à une accélération nécessaire des programmes de numérisation rétrospective. Le fait de pouvoir croiser d'un même mouvement une interrogation sur des

<sup>18</sup>. MSN et Yahoo! investissent le champ du patrimoine libre de droits et de l'information en ligne. Amazon ne traite pas les documents dépourvus d'ISBN, c'est-à-dire antérieurs à 1970. Google liste tout ce qu'il indexe, libre de droits ou non, en vrac.

fonds anciens ou contemporains est une plus-value majeure, qui impulse d'autant plus les efforts à consentir sur le patrimoine.

### Un affichage hiérarchisé et organisé des résultats et des provenances

Le deuxième service à forte valeur ajoutée concerne la fiabilité des résultats. Beaucoup déplorent aujourd'hui l'opacité des critères d'affichage de Google (*page ranking*) et de ses algorithmes. C'est un entrepôt. La transparence des critères de classement des résultats ainsi que la certification de l'origine de la source de numérisation sont des atouts permettant de rallier l'essentiel de la communauté des chercheurs et des médiateurs de l'information, avec valeur d'usage pédagogique pour le grand public.

Pour atteindre cet objectif, sont prioritaires l'utilisation de possibilités d'interrogation en langage naturel, croisée avec un affichage intelligent des résultats, qui élargissent les réponses aux domaines connexes, comme le permettent les possibilités de recherche du web sémantique.

C'est dans ce service qu'on distinguera la BNUE des entrepôts numériques et qu'elle gagnera sa légitimité en tant que bibliothèque numérique organisée des savoirs en ligne. La force d'une étiquette européenne est de garantir, par une certaine maîtrise des contenus, une certification d'origine d'une part et d'offrir la possibilité de consulter des ensembles organisés de façon plus transparente d'autre part.

#### Le multilinguisme

Le multilinguisme est enfin au cœur de la problématique d'une bibliothèque numérique européenne pour favoriser l'accès le plus large à la diversité et à la richesse du patrimoine culturel exceptionnel de l'Europe. Les interfaces multilingues de qualité, appliquées au champ d'une bibliothèque numérique d'une part,

les offres d'un même texte numérisé en plusieurs langues d'autre part, sont pratiquement inexistantes actuellement.

Il existe en France et en Europe des entreprises qui ont une certaine expérience en matière de recherche avancée multilingue, du fait que les Anglo-Saxons, forts du tropisme majoritaire de l'anglais, ont moins investi dans ce domaine. Il convient, dans cette perspective, d'encourager parallèlement les standards qui facilitent les connexions avec des outils linguistiques existants. En effet, l'accessibilité en ligne tient essentiellement à la possibilité d'interpréter des contenus automatiquement. Sans une normalisation du codage des contenus, on ne peut développer des outils génériques valables dans plusieurs pays.

Par ailleurs, s'il va de soi que l'outil commun doit offrir des passerelles d'accès multilingues telles qu'évoquées plus haut, la question des traductions automatiques des textes numérisés reste encore sujette à caution, pour différentes raisons à la fois juridiques, politiques et techniques. Il semble raisonnable, dans un premier temps, de mettre en correspondance les traductions existantes, libres de droits, des grands textes numérisés du patrimoine culturel et de développer des outils spécifiques issus du web sémantique en appoint (aide lexicale, contextualisation intelligente...).

#### **Le point de vue des professionnels (bibliothécaires et éditeurs)**

Les bibliothèques savent que le seul patrimoine libre de droits ne constitue pas, dans la durée, un élément assez attractif pour les usagers distants, et qu'il faut développer des passerelles avec d'autres portails ayant une offre sous droits: c'est toute la problématique développée actuellement dans le monde des bibliothèques de recherche avec des portails de revues utilisant notamment un système de barrières mo-

biles. Les bibliothèques nationales savent de surcroît que le présent et encore plus l'avenir les lient étroitement, par le dépôt légal du papier et de la Toile notamment, aux acteurs de la chaîne du livre, en premier lieu aux éditeurs.

Elles rejoignent sur ce terrain les éditeurs qui déplorent le manque de transparence et de certification des critères de tri et d'affichage des

### **Les bibliothèques nationales savent que le présent et encore plus l'avenir les lient étroitement, par le dépôt légal du papier et de la Toile notamment, aux acteurs de la chaîne du livre, en premier lieu aux éditeurs**

grands moteurs de recherche. Plus encore, la crainte d'être pillés plaide pour le développement d'outils apportant une alternative. Une interface commune d'interrogation est pour les bibliothèques comme pour les éditeurs un élément important de fidélisation de la clientèle.

Mais cela ne suffit pas. Chez les éditeurs, il y a également une forte attente pour d'autres outils, puisque toute consultation de document numérisé en ligne sous droits doit donner lieu à rémunération des ayants droit, qu'il s'agisse des auteurs ou des éditeurs. Dans cette perspective, à l'instar de l'exemple allemand, outre l'interface d'interrogation et de navigation précédemment évoquée et bien évidemment une accessibilité démultipliée par le recours à un moteur de recherche performant, une

plate-forme mutualisée de gestion rémunérée des contenus sous droits offrirait:

- un espace auquel on ne pourrait accéder qu'à certaines conditions et notamment en contrepartie d'un ou de plusieurs modes de rémunération donnés (sur abonnement, à la consultation, impressions, téléchargement...), déclinés par les éditeurs en fonction de leurs besoins. Les éditeurs reverseraient à leurs auteurs un pourcentage de ces recettes proportionnel à l'exploitation des œuvres;
- des outils de sécurisation pour les documents communiqués en ligne, ainsi que pour les systèmes de paiement (y compris le micropaiement).

Ces orientations peuvent s'accompagner de mesures techniques et organisationnelles qui inciteraient les éditeurs à proposer à la vente en ligne certains des ouvrages protégés de leurs fonds. C'est dans cet axe que s'engage la réflexion du Börsenverein allemand, qui a le mérite d'apporter une réponse intégrant les préoccupations de la librairie traditionnelle en matière de commerce en ligne.

Pour les fonds d'éditeurs encore protégés par le droit d'auteur, mais pour lesquels une hypothèse de commercialisation en ligne ne présenterait aucun espoir de rentabilité commerciale, il serait aussi intéressant d'inciter les maisons d'éditions à donner les droits de communication en ligne des œuvres à une société de gestion collective. Les institutions auraient alors la possibilité de recourir à ce « réservoir » autant que de besoin et de verser des rémunérations correspondant à leurs utilisations à ladite société de gestion (par exemple rémunération forfaitaire par l'institution en cas de gratuité d'accès pour l'utilisateur final, ou rémunération proportionnelle en cas de paiement par l'utilisateur final).

Enfin, il s'agit d'optimiser la conservation des données numériques en garantissant le moins de déperdition dans les transferts informatiques successifs. Il est nécessaire pour cela de se plier à des standards précis.

De l'avis de l'ensemble des professionnels consultés (spécialistes des moteurs de recherche, éditeurs, bibliothécaires), ce poste financier à terme sera beaucoup plus coûteux que l'investissement initial dans la numérisation proprement dite. Sur ce point, auquel réfléchissent les principaux moteurs de recherche, il n'existe pas aujourd'hui de formule opérationnelle, ni même, en amont, de recherche aboutie. Là résiderait le dernier potentiel innovant d'une réponse européenne, qui reprendrait

### L'analyse des attentes et des plus-values par rapport à l'existant conduit à la définition d'une bibliothèque numérique comme un ensemble organisé de contenus en ligne sélectionnés et non comme un entrepôt

certaines des solutions proposées par un consortium tel qu'Internet Archive en l'appliquant au niveau européen et qui, en tout cas, impliquerait les différentes bibliothèques nationales de façon solidaire.

La Commission européenne encourage tous les développements de standards propices à une organisation rationnelle de l'archivage numérique et à un partage des responsabilités. Les bibliothèques nationales ont un rôle majeur à jouer dans cet archivage, complémentaire de celui des éditeurs en matière de diffusion, car ce sont des institutions qui peuvent s'engager dans la durée, à la différence du secteur privé.

Ces quatre aspects (interface de localisation des textes, navigation in-

telligente dans l'interface, gestion mutualisée des rémunérations des ayants droit et organisation rationnelle de l'archivage numérique) pourraient constituer la feuille de route prioritaire du moteur de recherche de la BNUE. Ils correspondent à une attente partagée des professionnels et des internautes. En cela, ils constituent une réponse politique à une demande sociétale de services et de contenus.

Même si maîtriser les données en ligne et avoir la responsabilité de leur sauvegarde pérenne peut paraître secondaire aux yeux des internautes, ces enjeux culturels et économiques concernent aussi le contribuable qui, par ailleurs, attachera le plus grand prix à une interface intelligente de navigation. On répond ici aux critères d'ouverture sur le domaine contemporain, d'attractivité des contenus et de l'interface et, pour une grande part, de rapidité d'obtention des résultats.

#### *Une interface commune de visualisation et de recherche plein texte ?*

Portée par l'exemple de Google qui place les internautes dans un contexte d'affichage uniforme des textes en ligne, qu'il s'agisse d'extraits sous droits ou de documents de type patrimonial, la proposition française doit bien cerner ce que doit apporter une interface commune de visualisation dans les attentes du public. Faut-il copier Google ou bien faire valoir d'abord d'autres plus-values telles que celles décrites ci-dessus ?

Distinguons deux choses : d'une part une interface unique de visualisation des textes numérisés, qui impose de fortes contraintes techniques en particulier pour les bibliothèques, et d'autre part l'enjeu de la recherche plein texte. Quelles sont les attentes du public ? La recherche plein texte est une exigence de plus en plus répandue, qui s'applique non seulement aux outils que sont les tables des matières par exemple mais aux textes eux-mêmes. Par ailleurs, les internautes attachent une certaine

importance au fait de ne pas être dépayés par une navigation dans des interfaces différentes de consultation sur des sites différents. Enfin, ils sont sensibles au fait de rester le plus possible dans un environnement comportant une visualisation unifiée des textes consultés.

La première exigence, celle de la recherche plein texte, est aujourd'hui la priorité. Pour les fonds récents des éditeurs qui, peu ou prou, existent sous forme numérique depuis une dizaine d'années, un affichage en mode texte ne pose pas de difficulté insurmontable, même si des opérations de reformatage sont à prévoir selon les cas.

En revanche, la recherche plein texte implique pour les bibliothèques, en particulier pour la BnF, une importante mise à niveau technique puisque Gallica est majoritairement en mode image. La solution qui serait la plus opératoire et permettrait à la fois recherche plein texte dans les textes numérisés et maintien de l'apparence originale des textes, est l'OCR (reconnaissance optique de caractères). Or, le coût d'un tel chantier est loin d'être négligeable. L'OCR aujourd'hui ne donne pas des résultats satisfaisants pour des corpus de l'époque moderne (XVI-XVIII<sup>es</sup> siècles), pour lesquels il faudra proposer d'autres solutions. Notons aussi que, selon les types de documents, il faudra adapter les formats de visualisation (textes illustrés, textes non illustrés, presse...). Ce chantier, qui concerne Gallica, a été considéré comme prioritaire dans le plan d'action gouvernemental.

#### **Les contenus de la BNUE : qualité et quantité**

L'analyse des attentes et des plus-values par rapport à l'existant conduit à la définition d'une bibliothèque numérique comme un ensemble organisé de contenus en ligne sélectionnés et non comme un entrepôt. C'est l'alliance des services et d'un positionnement des contenus qui parachève la mutation.

### Une charte ouverte et porteuse d'identité

Quatre thèmes rassembleurs se dégagent autour de ce qui fonde l'identité européenne (cf. encadré). Ces thèmes permettent de définir des corpus (existants, à consolider ou à bâtir), qui peuvent être mis à disposition en ligne immédiatement ou dans un bref délai en ce qui concerne l'apport français afin de constituer une masse critique.

Ils rencontrent des initiatives actuelles de plusieurs pays européens, notamment en matière de presse ancienne ou de numérisation de classiques. Ils demandent à être traités dans une approche comparatiste. Ils permettent de réunir des textes fondamentaux de la culture européenne : classiques de la pensée, de la littérature ou de l'art susceptibles d'intéresser un large public<sup>19</sup>, tout en concernant un public de chercheurs déjà familiarisé avec une offre de textes plus difficiles à se procurer, disponibles dans Gallica ou dans certains portails de revues (ex. textes du XVI<sup>e</sup> siècle, de la Révolution française, articles d'histoire du XIX<sup>e</sup> siècle). C'est l'alliance entre ces classiques et des textes moins connus mais importants qui, grâce à Internet, multiplie les appropriations possibles du patrimoine culturel européen par des publics variés.

Les axes thématiques décrits dans l'encadré ci-contre recouvrent volontairement des catégories documentaires. L'inclusion de documents tels que la presse, les cartes, les dictionnaires notamment est le meilleur gage d'un succès dépassant les publics actuellement connus du web documentaire : ce sont en effet ces types de documents qui suscitent le gros de la demande de nos contemporains, au-delà de la littérature, des sciences et des arts, domaines pour lesquels il y a un intérêt avéré.

19. Comme le fait actuellement Gallica classiques (1 000 textes de la littérature française) très utilisé en milieu scolaire.

#### Axes thématiques

##### Le mouvement de la pensée : l'Europe, creuset fondateur

- Les idées : les grands textes des philosophes, essayistes, penseurs de toute obédience.
- L'apport des sciences et techniques : leurs outils majeurs de diffusion à travers les individus et les institutions qui les ont portés et continuent à le faire, principalement traités et revus dans un premier temps.

##### L'émergence et le développement des démocraties et des libertés publiques : les sources du débat public

- La construction politique : constitutions, discours politiques, principales sources juridiques.
- L'essor de l'opinion publique : les revues de débats et les journaux d'opinion, du *Mercure de France* au *Figaro* et à *L'Humanité*.
- Les sources du débat économique et social : des recueils de statistiques aux revues critiques de sciences humaines et sociales.

##### Identités nationales/identité européenne : le patrimoine culturel

- Les langues : dictionnaires, outils d'analyse de la production éditoriale.
- Le mouvement des lettres (la création littéraire dans toute sa diversité au fil des siècles ; une sélection pour le XX<sup>e</sup> siècle en rapport avec cette problématique).
- Le patrimoine artistique européen et l'apport de la création artistique dans le débat identitaire (par exemple, les convergences en Europe des avant-gardes au XX<sup>e</sup> siècle).
- Accent mis sur les livres très souvent traduits dans d'autres langues européennes.

##### L'espace européen

- Les histoires nationales et les relations internationales : principaux historiens et principales sources (par exemple, traité de Verdun).
- La construction de l'espace européen, dans sa triple dimension (continentale, nationale et régionale), ses populations, à travers entre autres un choix de cartes et d'atlas.

### Des contenus structurés par des outils intelligents d'indexation et de classification

Il n'y a pas d'antinomie entre une construction classique de charte documentaire qui indique les axes (*guidelines*) organisant les choix de numérisation et un moteur de recherche efficace ratisant très largement les contenus en ligne ; le point de rencontre est justement la capacité de ce moteur à présenter de façon organisée, avec une navigation en même temps intuitive utilisant les ressources du langage naturel, des contenus qui ont de fortes connexions entre eux. Une charte documentaire aide à structurer des services nouveaux.

*A priori*, les bibliothèques se retrouvent davantage dans les contenus proposés dans la charte, qui recourent des points forts des collections patrimoniales dont elles ont la charge, ceci aussi bien en France que dans les différents pays européens, même si les partenaires ne seront pas toujours, selon les thèmes, les autres bibliothèques nationales.

Des axes documentaires fédérateurs permettent aussi aux éditeurs volontaires de participer et de profiter des avantages offerts par la BNUE, entre autres des outils d'indexation et de classification qui pourraient être mis en place pour la BNUE et structurés par les axes de la charte. Voici quelques exemples de recherches possibles et souhaitées : la récupération, dans une même interrogation, de la liste des titres de revues politiques numérisées publiées entre 1920 et 1950 en France, Allemagne et Grande-Bretagne, ou encore la production d'un écrivain mort en 1920 avec ses traductions, ses biographies et la littérature critique publiée ensuite dans les différentes langues européennes.

Outre la presse, les sciences humaines et l'édition de référence sont aussi plus concernées par les enjeux de l'accessibilité des textes en ligne que la fiction. La littérature générale n'est pas pour autant à exclure. Le

numérique ouvre aux auteurs et éditeurs une nouvelle porte en termes de notoriété et de mémoire. Enfin, il ne faut pas exclure la perspective d'une contribution de certains corpus de quelques éditeurs, détenteurs d'un patrimoine intellectuel sous droits, qu'ils estiment ne plus avoir les moyens de rentabiliser économiquement et dont ils confieraient la valorisation, après accord des ayants droit, à la BNUE. Cette piste évoquée par plusieurs éditeurs de sciences humaines n'est pas à exclure par principe. Mais, de l'avis de la plupart des éditeurs consultés, l'idée d'une plateforme de gestion de la rémunération des ayants droit est la priorité.

Ont été écartés de la réflexion les bases de données très spécialisées (les bases de données d'articles de recherche par exemple), qui bénéficient déjà d'un circuit et d'une identité propres, les sites de pré-prints et les produits scolaires de type manuel; mais, concernant ces derniers, libre aux éditeurs d'en décider, notamment en lien avec les dictionnaires et encyclopédies.

Le rôle de ces axes thématiques doit être envisagé pragmatiquement. Dans une première phase, leur mérite serait de structurer des priorités de politiques documentaires et éditoriales en matière de numérisation et d'indexation et d'avancer ainsi de façon concertée. Il est évident qu'ils ne doivent pas constituer un frein et qu'ensuite, ils seront appelés à évoluer. Mariés à des outils intelligents d'indexation et de classification qui organisent les réponses aux demandes des internautes, ils prennent un poids structurel qui leur confère une dimension concrète, donnant sens à un mot d'ordre politique mobilisateur.

### Les seuils quantitatifs

L'ambition consiste à réaliser dans des temps resserrés des programmes crédibles. Les millions de titres annoncés par Google font basculer dans le champ des textes non libres de droits, pour lesquels les solutions dépendront

des éditeurs. Quels seuils peut-on fixer pour répondre à la fois aux objectifs d'une charte et à l'impérieuse nécessité d'offrir un matériau suffisant?

La situation de l'offre en ligne de patrimoine écrit a été décrite comme inégale. Si Gallica doit alimenter de façon importante la partie française d'une bibliothèque numérique européenne, elle fait aujourd'hui face à des critiques sur la qualité de sa numérisation, majoritairement en mode image à un moment où l'industrialisation des techniques de reconnaissance optique de caractère plaide pour une mise à niveau technique de l'existant. Cette priorité qui s'impose à la BnF n'est pas la même pour la British Library, par exemple, qui ne disposait pas jusqu'à présent des crédits autorisant une numérisation massive. Il y a donc nécessité de plans de rattrapage qui ne sont pas tous de même nature.

Pour Gallica, il est assurément tout aussi important d'améliorer l'interface de consultation des contenus existants que de multiplier de façon conséquente son offre. Une première estimation montre en effet que s'il y a des lacunes dans le choix des 80 000 documents actuels, le différentiel pour disposer des fondamentaux de la culture européenne écrite, presse non comprise, ne permettrait guère d'aller au-delà d'une multiplication par 4 ou 5. Ceci inclurait évidemment des documents difficiles à trouver mais d'intérêt documentaire avéré et également un choix de traductions des grandes œuvres européennes libres de droits disponibles dans les collections de la BnF<sup>20</sup>. C'est dans cette perspective que la BnF réfléchit à un vaste plan de numérisation de masse de l'écrit pour les toutes prochaines années. L'effort est déjà bien engagé puisque la numérisation de la presse nationale quotidienne devrait

permettre la mise en ligne, en accord avec les organes de presse, d'environ 26 titres des origines à 1940 d'ici cinq à sept ans. L'autre axe du plan d'action concernant la BnF consistera à mettre à niveau l'actuelle offre en monographies et revues. Pour en clarifier les enjeux techniques et financiers, la BnF entreprend un test de numérisation de masse dont les résultats seront disponibles à l'été 2006.

À terme, la contribution de la France pourrait atteindre le double en nombre de documents, soit un million. Quel est le point d'équilibre des plans de numérisation massive? Jusqu'où aller? Quelles sont les conditions pour une politique documentaire? À l'échelle du million de documents pour la France, on bascule *de facto* dans le champ des documents non libres de droits et une association élargie de partenaires avec la BnF et le noyau initial. Cela renforce la pertinence d'une réponse européenne consacrant l'alliance avec les éditeurs pour localiser l'ensemble des textes numérisés qu'ils soient libres de droits ou non. Puis, progressivement, de façon contractuelle avec les éditeurs, la BNUE, *via* ses structures nationales, augmentera le potentiel de textes intégraux récents en ligne.

Quel peut être le rythme pour la contribution française? Dans un premier temps, un objectif de 80 % de textes libres de droits, fournis par les bibliothèques et de 20 % de textes sous droits est plausible, car les éditeurs viendront progressivement, mais il faut néanmoins tendre ensuite rapidement vers un objectif où les apports de textes sous droits et libres de droits seront équilibrés.

### La BNUE mode d'emploi: propositions

#### Une démarche par étapes

La BNUE réunit un nombre limité d'acteurs volontaires ayant des intérêts convergents autour d'un objet commun: l'écrit diffusé.

20. Une projection sur le seul champ de la littérature (traductions en français libres de droits des classiques italiens, hispano-lusophones, germaniques et anglophones) conduit à une estimation de 4 000 volumes à numériser dans un premier temps, en plus de ce qu'il y a déjà dans Gallica (1 500).

Ce sont d'abord les bibliothèques nationales ou les grandes bibliothèques associées dans des programmes de numérisation de leurs collections patrimoniales, en France et Europe. Ce sont les éditeurs volontaires en France et en Europe. Ces derniers retireraient les avantages suivants: une plate-forme commune de gestion des droits, la mise en valeur des produits et services numériques existants, une plate-forme de mise en ligne pour les éditeurs n'en ayant pas les moyens propres, un espace de promotion très visible. Ce sont également les prestataires associés directement: sociétés de recouvrement de droits, prestataires de numérisation, moteurs de recherche, intégrateurs-développeurs. Parmi ces derniers, tous n'ont pas le même rôle dans la structure BNUE et certains resteront prestataires.

Y a-t-il place pour d'autres acteurs? Dans une première étape, l'Europe culturelle pourrait se concentrer sur l'idée de bibliothèque, popularisée par la campagne médiatique actuelle, et une fois confortée par cette réalisation, elle élargirait une recette qui aura fait ses preuves à d'autres contenus comme les archives<sup>21</sup>, les images et le multimédia<sup>22</sup>. Une démarche par étapes renforce la crédibilité de l'outil qui doit rester assez ouvert techniquement pour ne pas être bloqué ensuite.

Du point de vue de la chaîne du livre, le positionnement des libraires demande à être approfondi. On peut concevoir que les libraires traditionnels y trouvent, à l'instar de ce qui est proposé en Allemagne, une occasion pour investir davantage le commerce en ligne: une librairie comme Decitre, à Lyon, donne l'exemple en France d'un investissement réussi dans ce

secteur. Quant aux librairies de taille moins importante, il y a place, pour elles particulièrement, pour la distribution des produits papier qui continueront même avec Internet à faire l'objet d'une demande importante. Il est donc recommandé d'instruire dans la concertation entre les libraires, les éditeurs et les services de production des grands établissements culturels une stratégie qui évite au premier chef de se porter mutuellement préjudice, et de répartir dans un partenariat privé/public bien compris la fourniture des produits dérivés issus de la consultation en ligne (par exemple les reprints...).

Que faire vis-à-vis des partenaires des bibliothèques dans leurs réseaux nationaux? Les outils qui ont fait leurs preuves et qui seront partie prenante de la BNUE serviront de courroie d'entraînement pour la mise à niveau technique et scientifique qui s'impose. Les réseaux tireront vers le haut les bibliothèques intéressées qui voudront participer. La BNUE doit aider en ce sens les réseaux nationaux, en s'appuyant sur eux, faute de quoi les acquis d'harmonisation risquent d'être découragés.

Des structures telles que Gallica doivent continuer à fonctionner et garder un rôle essentiel pour éviter la multiplication des doublons et améliorer la visibilité d'ensemble des initiatives de numérisation des bibliothèques françaises, en complément d'outils tels que Michael. En d'autres termes, Michael signale les fonds, et Gallica, utilisant entre autres le protocole OAI, permet de naviguer vers les documents numérisés d'autres bibliothèques complémentaires des propres programmes de la BnF.

### Répartition des rôles et des fonctions des principaux acteurs

1. Les fournisseurs de contenus à la BNUE sont les bibliothèques et les éditeurs, qui:

- sélectionnent les documents selon les axes d'une charte;
- fournissent des documents nu-

mériques ou bien numérisent des documents papier, éventuellement en sous-traitant à des prestataires de numérisation.

Les fournisseurs sont responsables des données et métadonnées; ils sont prescripteurs de l'indexation unifiée; ils fournissent les documents et leurs descriptions selon un cahier des charges.

2. L'interface BNUE entre contenus et services aux utilisateurs est réalisée par un intégrateur, qui:

- contrôle les contenus fournis;
- gère le portail web;
- gère l'extranet entre les partenaires BNUE;
- met en ligne et assure l'interopérabilité;
- intègre le moteur de recherche (multilingue);
- indexe les données fournies (classification, traduction, moteur de recherche);
- garantit les performances, la disponibilité et la robustesse du système;
- assure la sécurité des comptes utilisateurs.

3. La gestion de l'espace payant (par abonnement ou tout autre moyen), celle des droits pour les ayants droit (SPRD - Société de perception et de répartition des droits) et celle des autres services commerciaux (fonction de promotion), consistent à:

- offrir des possibilités de micro-paiement;
- sécuriser les paiements des utilisateurs;
- appliquer les prix définis par les éditeurs;
- redistribuer les revenus aux ayants droit;
- gérer l'espace de promotion en lien avec les distributeurs traditionnels.

Pour l'archivage pérenne et la sauvegarde des données, un cahier des charges commun est défini, les rôles sont répartis, les bibliothèques nationales assurant une part majeure de l'archivage.

Dans chaque cas, y compris l'archivage, il peut y avoir de la sous-traitance à un ou plusieurs prestataires. Quel

21. Il importe en effet pour les archives de trouver une place dans la BNUE pour différentes raisons: les fonds des bibliothèques patrimoniales et des archives sont intimement liés (manuscrits, documents officiels, presse...). D'un point de vue technique toutefois, la gestion de séries nominatives ou topographiques appelle des développements précis qui n'autorisent pas une intégration d'emblée de certains fonds.

22. Les archives sonores, les fonds des musées ne sont pas au même niveau de numérisation que les bibliothèques.

que soit le schéma organisationnel qui s'impose en bout de course, ces rôles seront tenus par les acteurs cités ci-dessus, car cela correspond à des cœurs de métiers.

### Architectures

Il est prématuré de détailler les différents scénarios étudiés par les professionnels dans le cadre des groupes de travail, car ils requièrent une instruction technique et financière complémentaire, préalable aux arbitrages politiques. Toutefois, on peut indiquer qu'il existe plusieurs scénarios possibles d'organisation et d'architecture, dont on cerne aujourd'hui les implications techniques.

L'adoption d'un système totalement décentralisé a été *a priori* exclue du fait de sa trop forte dépendance à l'égard des systèmes locaux. Le scénario préconisé est à mi-chemin entre une architecture totalement décentralisée et une architecture fortement centralisée de type Google, parce que ses implications politiques et financières sont plus adaptées au contexte européen, qu'il peut s'appuyer sur des technologies récemment développées en termes de performances et d'interopérabilité, qu'il est plus rapidement opérationnel et, enfin, parce qu'il n'exclut pas la mise en place, dans un second temps, d'un schéma plus centralisé.

Il reste à poursuivre l'état des lieux technique précis des données des fournisseurs potentiels, en dehors des bibliothèques, à préconiser des standards et normes, et à proposer un chiffre précis des infrastructures à mettre en place, et, en particulier, à approfondir la question du moteur.

La question qui se pose enfin est celle de la nécessité ou non de plateformes nationales complémentaires de la plate-forme BNUE, en fonction de l'architecture politique que l'on vient de proposer. Il est difficile de trancher aujourd'hui. Les besoins diffèrent selon la taille des pays : si l'on peut juger opportun d'avoir une plate-forme nationale au niveau d'un

petit pays unissant à la fois éditeurs et bibliothèque nationale (question de taille critique), cette nécessité se fera sans doute moins sentir dans le cas de pays plus grands comme la France, la Grande-Bretagne, l'Espagne ou l'Allemagne. Pour les éditeurs, il semble qu'une structure de plate-forme nationale soit utile, à l'instar de ce que fait le Börsenverein.

### Politique de financement

Les États ont un rôle décisif pour impulser le processus. Ils interviennent pour mettre à niveau les programmes de numérisation des bibliothèques et lancer le développement des plateformes nationales des éditeurs. Ils participent à la mise en place de la plate-forme européenne et sont partie prenante de la structure pour encourager l'association d'autres partenaires. Ils encouragent les établissements publics ou les structures type SNE (Syndicat national de l'édition), ou associées type SPRD (sociétés de perception et de répartition des droits d'auteur et droits voisins), en leur donnant des moyens fléchés pour poursuivre leurs engagements. Ils les aident, le cas échéant, à trouver des mécènes.

La Commission européenne n'est pas en première ligne dans le lancement du dispositif, mais elle est associée sur un plan politique pour que les outils type Michael et surtout TEL soient adaptés : les principes politiques en sont bons, la stratégie et les techniques employées sont à adapter au nouveau contexte. Elle encourage les standards ouverts utilisables par l'interprofession et les programmes de type « e-content + ».

L'association d'acteurs crédibles du privé est essentielle pour inspirer confiance et susciter l'adhésion des éditeurs, notamment des grands groupes éditoriaux, des industriels, des opérateurs de réseaux, des prestataires de service. L'autre élément de confiance est la participation à un niveau important de la structure des bibliothèques nationales (pérennité des missions de ces institutions garantes du long

terme). Le modèle économique des éditeurs se déduit de la notoriété de la BNUE et de la mutualisation des outils offerts par la plate-forme. Là sont les sources d'équilibre et de profit.

Si les industriels sont avant tout prestataires, ceux qui feront directement partie de la BNUE, et qui sont les partenaires du moteur de recherche en particulier, y gagnent en retour d'image par rapport à un investissement qui reste raisonnable en regard d'autres secteurs comme celui du multimédia. On peut imaginer que les outils mis au point puissent être rentabilisés sur des produits plus rémunérateurs comme les moteurs multilingues appliqués aux brevets.

Les bibliothèques ont tout à gagner d'un partenariat avec des industriels, aussi bien en ce qui concerne des financements éventuels de parties de programmes de numérisation (mécénat), que de soutien en matière de moteurs. La mise en commun d'un certain nombre de problématiques de commercialisation entre privé (éditeurs) et public (bibliothèques) peut relancer une dynamique du côté des produits dérivés. Enfin et surtout, cette alliance garantit une politique de marketing efficace de la BNUE.

S'il y a convergence sur le principe d'un partenariat public/privé, il reste à arbitrer la forme qu'il peut prendre. L'idée d'une fondation ne recueille pas un consensus, et une étude est à prévoir pour examiner sur un plan juridique quelle forme est la plus opératoire à l'échelle européenne.

### La coordination en Europe

Les pays européens ont développé selon leurs traditions nationales propres différentes solutions. Une BNUE ne peut en faire table rase. La première étape a consisté à faire des propositions de stratégie de partenariats, de contenus et d'organisation. À ce stade de la réflexion, le relais doit être pris par les instances politiques et professionnelles (bibliothèques, éditeurs...).

La BnF est en contact permanent avec la Conférence des directeurs

des bibliothèques nationales d'Europe (CENL) à qui elle a soumis une proposition écrite, développant notamment les axes thématiques ainsi que les réponses à l'appel de la Commission européenne, dont la plateforme mutualisée de gestion des rémunérations des ayants droit fait partie. Au niveau de la Commission européenne, il reste également à instruire comment un outil comme TEL s'harmonise politiquement et techniquement avec les propositions faites pour le développement d'une BNUE. Les enjeux d'une stratégie culturelle du web, tels que la force de frappe des moteurs les a délogés, obligent à une concertation renforcée.

Des rencontres avec le SNE ont permis de dégager des relais à saisir au sein de la FEP (Federation of European Publishers). Un calendrier de consultations est à l'œuvre, se fixant des échéances au moment des grands rendez-vous de l'interprofession (Salon du livre de Paris, Foire de Francfort...). De plus, les grands groupes d'éditeurs intéressés par le projet, qui ont tous une assise européenne, peuvent aussi servir de relais pour approfondir et faire des propositions. Ils le feront d'autant mieux que ce sont eux qui ont le plus investi dans les outils en ligne et qu'ils pèseront les avantages politiques, techniques et économiques en termes de visibilité sur le web de l'articulation d'une telle plateforme avec les leurs. Ils ont en outre la faculté de peser eux aussi sur les instances européennes pour promouvoir des standards et des normes utiles aux partenaires concernés, qu'il s'agisse des éditeurs ou des bibliothèques.

L'idée est de démarrer avec un noyau fondateur de quelques pays. La BNUE doit rendre interopérable ce qui entre dans les objectifs prioritaires. Par exemple, en France, les contenus de Gallica et ses développements vers les revues et la presse, les principales offres numériques comme celles de la Bibliothèque inter-universitaire de médecine ou Persée, les outils des éditeurs ou les groupes d'édition volontaires.

Le schéma organisationnel doit rester souple, pour tenir compte des traditions nationales tout en garantissant à terme l'interface de recherche et de consultation commune. Le maître mot doit être la garantie d'une interopérabilité technique et intellectuelle des efforts existants qui entrent dans la stratégie générale de la BNUE.

Le plan d'action détaille des préconisations de méthode pour la poursuite du projet tant au niveau national qu'au niveau européen :

- l'élaboration d'un plan de financement et d'un modèle économique;
- la constitution d'une plateforme commune en termes de contenus et d'infrastructure technique;
- un recentrage des politiques de soutien aux programmes de numérisation susceptibles de participer à la BNUE et une aide pour accélérer la mutation technique et quantitative nécessaire pour Gallica (conversion des contenus numériques existants et numérisation de masse). Une première approche budgétaire pour cette dernière permet d'évaluer à 3,6 millions d'euros le coût du passage en mode OCR de l'existant et à une base de 8 millions d'euros celui de la numérisation annuelle pour atteindre les objectifs quantitatifs fixés;
- l'intensification des échanges en Europe et la poursuite du portage du projet autour d'un noyau fondateur de pays.

Il y a en effet urgence à avancer au niveau français, tandis que s'étoffera au niveau européen une stratégie d'alliances qui s'enrichira au fur et à mesure que la BNUE se construit. Les premières échéances concrètes sont prévues fin 2006.

## Un défi majeur

Mettre en place une stratégie culturelle du web à travers le domaine de l'écrit est un véritable enjeu démocratique pour l'Europe. Si la stratégie des grands moteurs de recherche est

d'abord commerciale, n'oublions pas que, dans cette quête de la maîtrise de l'ensemble des contenus en ligne, le secteur culturel et scientifique a une valeur toute particulière. Puisque les moteurs cherchent un supplément d'âme, disons de légitimité auprès du grand public, à travers le domaine de l'écrit, il faut se donner les moyens d'une politique culturelle de l'Internet. C'est bien ce sentiment qui a suscité une mobilisation inédite autour de l'annonce de l'idée de bibliothèque numérique européenne, renforcé par le sentiment d'un certain gâchis dans la dispersion des initiatives et des investissements.

Le domaine de l'écrit est un enjeu politique de première importance. On y retrouve le cœur de métier des bibliothèques et des éditeurs, de façon plus générale des médiateurs de savoirs : l'organisation des savoirs et leur mise à disposition de façon intégrale afin que chacun se fasse son propre jugement sont les fondements de toute démocratie. La mise à disposition d'informations consultables de façon fragmentaire, en vrac, selon des critères opaques, a de lourdes conséquences dans la formation des esprits, non seulement parce qu'il y a risque de prédominance d'une culture sur les autres, mais aussi parce que des textes tronqués peuvent conduire à tous les égarements.

La seule réponse est d'élaborer une stratégie culturelle de mise à disposition de ce qui fait la culture de référence de l'Europe. L'accès aux sources de la pensée à travers une bibliothèque organisée en ligne est un défi majeur, qui a le mérite de faire redécouvrir ce qui doit unir les acteurs de la chaîne de l'écrit, au-delà des intérêts particuliers.

Mars 2006