

**Réponse des autorités françaises à la consultation publique de la Commission  
sur les bibliothèques numériques :  
le projet de Bibliothèque numérique européenne**

**Introduction**

Le réseau Internet est devenu, pour un grand nombre des citoyens de l'Europe, l'un des principaux moyens d'accès à la connaissance, à la communication et à la diffusion de l'information. La maîtrise de la sélection, de la présentation et de la diffusion des savoirs passe aujourd'hui par une forte présence sur le réseau.

Si nous voulons que notre culture garde toute sa place dans la future géographie des connaissances, il importe donc de réunir les meilleures conditions d'une diffusion en ligne efficace et intelligible de notre patrimoine, à l'attention de tous les citoyens européens et de toutes les cultures.

Des réponses développées ci-après aux questions destinées à évaluer les options appropriées pour concrétiser l'objectif de rendre le patrimoine numérique européen visible et disponible, il ressort qu'une bibliothèque numérique européenne permettrait, bien plus efficacement qu'une simple coordination européenne de projets épars : de donner une forte visibilité pratique et symbolique aux contenus européens, de construire des services innovants en direction de tous les publics, de bénéficier d'une audience suffisamment large auprès du grand public, d'attirer ainsi les investissements privés nécessaires.

Une diffusion de la culture européenne ne pourra, en effet, être efficace si elle demeure disséminée et dispersée : pour avoir une visibilité comparable aux plus grands projets internationaux, une bibliothèque numérique européenne devrait offrir un accès unique à une quantité significative de documents. La création d'une bibliothèque numérique européenne, multiculturelle et multilingue, offrirait un point d'accès unique aux ressources numériques des pays de l'Union européenne, fortement visible pour les citoyens.

Les industriels américains ont pris une longueur d'avance en ces domaines, en particulier pour les moteurs de recherche grand public de la Toile ; les principaux, Google, MSN (Microsoft) et Yahoo, disposent désormais d'importantes sources de revenus publicitaires et de capacités d'investissement considérables. Depuis un an, chacun de ces trois grands acteurs de la Toile s'est lancé dans de grands projets de bibliothèques numériques.

Le plus ancien de ces projets, *Google Book Search* (ex-*Google Print*), lancé le 15 décembre 2004, a d'emblée émis des objectifs très élevés en termes de quantité de livres numérisés diffusés en accès libre sur les réseaux Internet.

Courant 2005, avec une approche plus collaborative et concertée, s'est formé le consortium Open Content Alliance, auquel participent notamment Yahoo et Microsoft.

Enfin, d'autres grands projets de numérisation ont vu récemment le jour sur l'impulsion des pouvoirs publics. En particulier, l'Inde et la Chine entreprennent également de grands programmes de numérisation, tandis que la Bibliothèque du Congrès (Etats-Unis) vient de lancer une *World Digital Library*.

Nous pensons que l'Europe doit être présente en tant que telle dans ce nouveau paysage de la diffusion. Elle doit rassembler et accentuer l'effort de numérisation du patrimoine et des contenus culturels entrepris par chacun des États membres, pour réaliser ensemble une grande bibliothèque numérique européenne, dont les collections culturelles et scientifiques soient le reflet de notre diversité et de notre communauté de civilisation, d'histoire et de pensée. C'est là un premier objectif qu'il faut développer sans tarder.

Parallèlement, il convient de faire porter les efforts sur un second objectif : identifier, reconnaître et cartographier les composantes de cet immense océan de connaissance. Une industrie spécifique s'est développée à cette fin : les moteurs de recherche, qui ont encore de grands progrès technologiques à accomplir, notamment dans les domaines du multilinguisme, de la sémantique et du multimédia. En facilitant l'accès à un plus large public ces outils doivent contribuer à construire la visibilité et la crédibilité du projet dans le domaine concurrentiel décrit ci-dessus. La qualité, la finesse et la pertinence des réponses aux recherches des usagers, sont des leviers valorisés fortement par les publics allant du spécialiste au profane. Il convient de s'appuyer sur les réalisations significatives et pertinentes de l'industrie européenne et française, en profitant de l'existant. Elles pourront être développées par des contacts avec l'activité de recherche ou de R&D conduite dans le cadre du 7<sup>ème</sup> PCRD, des programmes de l'AII (QUAERO), de l'ANR (RNTL) ou des Pôles de compétitivité.

Dans le contexte évolutif de l'Internet, la réflexion sur les moteurs de recherche devrait également être complétée par la mise en place de nouveaux services à valeur ajoutée. Il convient de l'appuyer sur les réalisations des industriels européens, en incorporant sans attendre les résultats pertinents, et en continuant à développer les technologies pour rester compétitif. Il s'agit également d'imaginer et de tester de nouveaux modèles interactifs et collaboratifs permettant l'appropriation, la réutilisation et l'enrichissement des contenus par les utilisateurs. Les travaux du 7<sup>ème</sup> PCRD pourraient être orientés dans ce sens, contribuant ainsi à l'évolution de la société de l'information à une société de la connaissance partagée.

Pour envisager l'évolution des pratiques et des technologies, nous pouvons proposer des scénarios techniques qui préservent à la fois la cohérence de l'image et l'extension du projet à d'autres partenaires. Sur le premier point, une interface unifiée présente l'avantage d'une meilleure visibilité conformément aux ambitions politiques du projet. Sur le second, il devrait être possible, si on prend garde à préserver l'avenir, d'adopter une architecture extensible, évolutive et distribuée auprès des principaux acteurs. En procédant de la sorte, il sera à la fois possible d'améliorer la dynamique du projet et de s'assurer de l'adhésion d'un plus grand nombre de partenaires pays ou entreprises.

Relever le défi culturel posé par l'Internet, c'est savoir être à la pointe de la performance et de l'innovation. C'est pourquoi, qu'il s'agisse de la production, du traitement ou de l'indexation des données numériques, une bibliothèque numérique européenne constituerait un véritable laboratoire de l'innovation en même temps qu'un phare de la culture et des savoir-faire européens.

Dans ce but, le Président de la République française avec ses homologues allemand, espagnol, hongrois, italien et polonais a écrit le 28 avril 2005 à M. Barroso pour proposer une initiative européenne concertée qui permette aux cultures européennes de prendre pleinement leur place dans ce vaste mouvement mondial.

En France, un comité de pilotage a été créé pour esquisser les orientations d'une future bibliothèque numérique européenne, qui seront ensuite proposées aux autres États-membres et à la Commission. Ses premiers travaux d'analyse et de réflexion, qui ont porté sur des enjeux culturels, scientifiques, technologiques, économiques, juridiques et européens, sont en voie d'achèvement et de synthèse.

En outre, dès à présent, la France travaille sur une maquette de démonstration et la Bibliothèque nationale de France a lancé une étude de faisabilité pour une numérisation en masse de ses collections et pour l'indexation des contenus numériques. Cette étude, qui doit aboutir vers le mois de mai, présentera des scénarios techniques et organisationnels, avec des estimations financières, grâce auxquels le comité de pilotage sera à même de faire des propositions concrètes et illustrées à la Commission européenne et aux États membres.

*1) Quelles mesures supplémentaires pourraient être prises au niveau national et au niveau européen pour encourager la numérisation et l'accessibilité en ligne du matériel dans toutes les langues européennes ?*

Les réalisations européennes en matière de numérisation sont souvent de très grande qualité mais fréquemment issues d'initiatives ponctuelles. Leur dimension modeste et leur spécialisation les destinent à des publics restreints et ne leur permettent pas de bénéficier d'une audience suffisamment large auprès du grand public.

C'est pourquoi la France propose de créer une bibliothèque numérique européenne. En offrant un point d'accès central et en développant les ressources numériques des pays de l'Union européenne, cette bibliothèque donnerait une forte visibilité pratique et symbolique aux contenus européens et permettrait de construire des services innovants en direction de tous les publics.

Une telle bibliothèque rassemblerait à la fois des documents issus d'opérations de numérisation ou créés nativement sous une forme numérique. Elle accueillerait un grand nombre de textes de toutes les langues européennes, numérisés dans tous les pays par des opérateurs privés et publics, et pourrait ultérieurement s'élargir à d'autres documents (documents iconographiques, sonores, audiovisuels, etc.).

Afin d'illustrer l'identité européenne, les contenus pourraient être organisés autour de quelques grands axes thématiques facilement identifiables, par exemple :

**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.

**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>ème</sup> siècle en rapport avec cette problématique) ;

- le patrimoine artistique européen (par exemple : les grands musiciens) et l'apport de la création artistique dans le débat identitaire (par exemple : les convergences en Europe des avant-gardes au XXème siècle).

Seront particulièrement privilégiés les textes les plus souvent traduits dans les différentes langues européennes, ce critère manifestant à l'évidence l'importance de l'impact d'une pensée, d'une oeuvre au-delà des frontières nationales.

### **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 ;
- les sources du débat économique et social : des recueils de statistiques aux revues critiques de sciences humaines et sociales, en y ajoutant les textes fondateurs de l'économie et de la sociologie.

### **L'espace européen :**

- les histoires nationales et les relations internationales: principaux historiens et principales sources ;
- la construction de l'espace européen, dans sa triple dimension (continentale, nationale et régionale) à travers un choix de cartes et d'atlas notamment.

Le noyau central de ces collections pourrait être constitué de plusieurs milliers d'œuvres emblématiques du patrimoine culturel européen issues de chaque pays, et qui seraient présentées dans le plus grand nombre possible de traductions dans les langues européennes.

Dans le domaine des archives, un objectif de la numérisation et de la diffusion pourrait être de parvenir, pour quelques documents de base (concernant l'histoire des personnes et des territoires par exemple), à une couverture totale du territoire de l'Union.

Dans le domaine de la recherche, la création d'espaces d'archivage de prépublications ou de documents de travail contribuerait à un meilleur partage des avancées les plus récentes.

La cohérence de cette bibliothèque numérique européenne pourrait être garantie par un conseil scientifique où siègeraient tous les Etats-membres. Le conseil émettrait des prescriptions techniques, normatives et multilingues, notamment en s'appuyant sur les projets européens *Minerva* et *Michael*. Il validerait les principaux axes de contenu thématiques.

La création d'une bibliothèque numérique européenne faciliterait ainsi la mise en place d'une coordination européenne en matière de contenus numérisés, afin de mieux organiser les projets et de proposer ainsi au citoyen européen une vision plus complète de sa culture.

A cet égard, le catalogue des collections numérisées *Michael*, soutenu par la Commission européenne, pourrait être étendu à l'ensemble des Etats de l'Union. Il permettrait de mieux constater les besoins en matière de numérisation et les lacunes à combler. Sur cette base, il serait souhaitable que des services correspondant à des usages d'intérêt européen puissent être

identifiés, sur des types de documents ou des thèmes précis, et que des normes techniques et de contenu puissent être définies pour leur exploitation, ainsi que des critères de qualité.

Un projet de l'ampleur de la bibliothèque numérique européenne constituerait également une vitrine exceptionnelle en matière de normes pour la diffusion, la description ou la conservation des données numériques, par exemple l'OAI-PMH ou le format de description METS. Il encouragerait la promotion des standards existants soutenus par la Commission européenne, favoriserait le développement de nouvelles normes et par là même l'interopérabilité entre les autres projets de numérisation européens ou nationaux, en cours d'élaboration ou en projet.

Pour mieux illustrer les enjeux techniques d'une telle bibliothèque, la France présentera prochainement une maquette. Les fonctionnalités de ce premier prototype devraient permettre de prendre la mesure des progrès qu'une bibliothèque numérique européenne permettrait d'accomplir, notamment en matière de services au public : elle constituerait un champ d'application précieux pour le développement d'outils de recherche d'information interlingues, la traduction automatique ou assistée, la mise en correspondance de textes bilingues ou multilingues, etc. Au regard de l'évolutivité des technologies, le projet devrait adopter une stratégie reposant sur une offre évolutive, capable d'intégrer les améliorations et les services innovants. (Ce point n'est pas particulièrement original, il est au centre de la démarche d'acteurs industriels tels que Google, Yahoo et Microsoft) En agissant de la sorte, il est possible de tester des technologies, des services et des « Business Models » en émergence, dans la période où le plus grand effet levier s'applique à l'investissement intellectuel ou financier dans le projet.

En outre, le test en vraie grandeur des usages innovants, leur insertion dans une offre intégrée et lisible, et leur intégration dans les pratiques (information, recherche, éducation, veille documentaire, ) seront ainsi facilités.

### **S'agissant plus particulièrement du multilinguisme**

L'accessibilité en ligne du matériel dans toutes les langues européennes constitue à la fois une force du projet et un enjeu de taille pour l'Union Européenne. Le multilinguisme est en effet au cœur de la problématique de la création d'une bibliothèque numérique européenne afin de favoriser l'accès le plus large à la diversité et à la richesse du patrimoine culturel de l'Europe.

Dans cette perspective, ce qui est déterminant n'est pas tant le volume numérisé que la capacité d'un utilisateur non spécialisé à retrouver les ressources disponibles. La conception de moteurs de recherche capables de dépasser les barrières linguistiques, par une traduction des termes de l'interrogation et de la réponse, pourrait apporter une nette amélioration.

L'enjeu du multilinguisme concerne de nombreux secteurs applicatifs au delà du projet de Bibliothèque numérique européenne lui-même, et y compris les besoins propres des instances européennes. Une action d'envergure visant à s'assurer de la disponibilité de technologies de la langue facilitant le multilinguisme dans toutes les applications concernées serait donc une absolue nécessité. Des bases substantielles existent, mais il convient de soutenir leur amélioration ; c'est un domaine où l'Europe peut être particulièrement motivée à développer des solutions compétitives, de par sa diversité linguistique et culturelle.

Concernant spécifiquement la bibliothèque numérique européenne, on pourrait mentionner le développement d'outils de recherche d'information interlingues, de traduction automatique ou

assistée, de mise en correspondance de textes bilingues ou multilingues, ou de correction orthographique permettant d'améliorer la reconnaissance de caractères dans les tâches de numérisation. Il convient également d'encourager les standards qui permettent l'échange de données et leur archivage pérenne, facilitent les connexions avec les outils linguistiques existants et leurs développements.

L'importance de l'effort à accomplir pour couvrir les différentes technologies nécessaires, dans les différentes langues européennes justifierait que la Commission Européenne et les Etats-Membres coopèrent étroitement. La nécessité à présent établie de disposer de données linguistiques en quantité suffisante pour pouvoir développer des technologies de qualité plaiderait pour que la Commission prenne en priorité la charge de tâches relativement indépendantes des langues traitées (coordination générale, standards, développement et évaluation de technologies génériques), et que les Etats-membres prennent en priorité les tâches relatives à la langue, ou aux langues, qui sont les leurs (constitutions de ressources linguistiques (lexiques, corpus, bases terminologiques, adaptation d'une technologie aux spécificités des langues parlées).

Ces différentes propositions contribueraient très fortement à améliorer la visibilité et la diffusion auprès du plus large public des ressources offertes par l'Europe et les Etats membres.

*2) Quelles mesures pourraient être prises pour promouvoir les investissements privés et de nouvelles stratégies d'entreprise telles que les partenariats entre le secteur public et le secteur privé pour numériser et rendre les collections historiques accessibles ?*

Un projet phare tel qu'une bibliothèque numérique européenne serait susceptible d'attirer les investissements privés bien plus efficacement qu'une simple coordination européenne de projets épars.

Il reste que, compte tenu de l'enjeu politique d'un tel projet, un apport significatif de fonds publics, au niveau national et européen, est absolument indispensable pour démarrer le projet, le rendre crédible et créer ainsi les conditions nécessaires aux partenariats avec le secteur privé.

Une bibliothèque numérique européenne, par l'ampleur de son champ et le grand nombre de visites qu'elle susciterait, pourrait être source de revenus, qu'ils soient liés à des prestations liées aux collections elles-mêmes ou, plus indirectement, au mécénat ou à la publicité.

Pour les documents libres de droits, des pistes de rentabilité pour des investisseurs pourraient être trouvées grâce à des services supplémentaires ou périphériques, susceptibles de s'agréger à l'offre documentaire gratuite : reproduction de documents, recherches spécifiques ou thématiques, offre documentaire élaborée, nouveaux services pour de nouveaux usages (éducatifs ou touristiques par exemple).

Les professionnels de l'édition pourraient trouver dans cette bibliothèque un espace de diffusion et de publicité. La présence de produits éditoriaux, consultables par extraits, faciliterait la visibilité des œuvres et leur vente en ligne, par l'intermédiaire des libraires ou en direct, par un système de commerce de fichiers électroniques. Par ailleurs, la coopération entre éditeurs et bibliothèques pourrait couvrir également le champ du stockage partagé des documents numériques.

Il est concevable que les professionnels s'intéressent à développer de nouveaux services ou documents écrits ou virtuels, permettant d'exploiter ou de faire référence à des documents libres de droit ou non. Ainsi par exemple, des catalogues raisonnés, de études historiques ou de littérature comparée et des critiques couvrant des sources multiples pourraient être conçus, suivant l'intérêt qu'ils sont susceptibles de générer. La valorisation de travaux de recherche pourra fournir un point de départ à certaines de ces entreprises qui nourriront les grands axes thématiques décrits précédemment.

Quant au mécénat, il s'exercerait à la fois sur des innovations technologiques et sur la numérisation de contenus textuels, par exemple des corpus d'œuvres ou de sources. La réédition de prestige de textes patrimoniaux, sous forme numérique ou physique, pourrait donner lieu à un parrainage spontané d'entreprises de natures très diverses, désireuses d'associer leur nom à de grandes œuvres, relatives ou non à leur activité (par exemple un corpus économique parrainé par une banque) .

D'une manière générale, de nombreuses sociétés privées seraient nécessairement associées à une telle réalisation, notamment par le biais des marchés publics (développements informatiques, innovations matérielles, hébergement de données, etc.), qui représenteraient pour elles des opportunités tant économiques que technologiques. Il y a lieu d'être vigilant au plan des cahiers des charges, de manière à encourager l'innovation et l'investissement, tout en développant la concurrence lorsque cela est souhaitable.

Dans un second temps, la constitution d'un vaste ensemble de textes multilingues fournirait un terrain favorable à la recherche et développement d'applications Internet et multimédia, ainsi qu'à l'amélioration d'outils tels que des moteurs de recherche multilingues, des dictionnaires, des ontologies, etc. Au-delà du projet de bibliothèque numérique européenne, ces progrès pourraient profiter aux institutions, services publics et entreprises de l'Europe en améliorant les conditions de la diffusion de l'information dans toutes les langues européennes sur la Toile mondiale.

Cette bibliothèque numérique européenne devrait donc être dotée d'une forme juridique adaptée permettant d'associer dans les meilleures conditions des partenaires de différents statuts.

*3) Quelles mesures de nature législative, technique, organisationnelle ou autre, pourraient faciliter la numérisation et l'accessibilité ultérieure du matériel protégé par le droit d'auteur, tout en respectant les intérêts légitimes des auteurs ?*

Les utilisations en cause, reproduction, communication au public et mise à disposition du public, relèvent de l'exercice des droits exclusifs du droit d'auteur. Les Etats Membres et la Commission pourraient rechercher les moyens de rapprocher les parties en présence. Les orientations à privilégier sont donc prioritairement le recours au contrat et à la gestion collective volontaire.

Ces orientations doivent s'accompagner de mesures techniques et organisationnelles qui inciteront les éditeurs à proposer à la vente en ligne certains des ouvrages protégés de leurs fonds. Les formes d'intervention doivent en effet être appropriées pour lever les réticences et les difficultés des éditeurs aux différentes étapes du processus.

En matière de diffusion en ligne de revues scientifiques, l'expérience française de la coexistence de sites privés (Cairn) et publics (revues.org et Persée, ce dernier pour la numérisation rétrospective) permet de tirer quelques recommandations. Il convient :

- d'aider techniquement (et marginalement financièrement) les producteurs d'information scientifique à utiliser les ressources numériques pour diffuser leurs travaux (exemple revues.Org). Il s'agit ici du volet diffusion d'une politique scientifique publique, avec un aspect culture scientifique et technique.
- de susciter la diversification des modèles et des sources (la concurrence du privé, stimulante pour les projets publics). L'aide à l'édition numérique scientifique n'a pas encore pris corps mais elle se justifie pleinement aujourd'hui.
- de soutenir l'offre publique en complémentarité à l'offre privée, pour la numérisation de revues ayant un marché limité, notamment dans une perspective patrimoniale.
- de favoriser l'interopérabilité et la communication des systèmes, afin de permettre à l'utilisateur de faire des recherches transversales à l'ensemble des sites existants.
- d'éviter des situations conduisant à limiter la concurrence, par exemple par des offres combinées « système d'exploitation » - « client de recherche » insuffisamment ouvertes à l'interopérabilité.

Dans cette perspective, une Bibliothèque numérique européenne offrirait le concept de plate-forme de services aux éditeurs : outils de sécurisation des documents communiqués en ligne, interfaces d'interrogation et de navigation, systèmes de paiement (y compris micro paiement) et bien évidemment une accessibilité démultipliée par le recours à un moteur de recherche performant. Cette plate-forme pourrait prévoir 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 abonnements, à la consultation, impressions, téléchargement, etc). Les éditeurs reverseraient à leurs auteurs un pourcentage de ces recettes proportionnel à l'exploitation des œuvres.

Pour les fonds encore protégés par le droit d'auteur mais pour lesquels une hypothèse de commercialisation en ligne ne présenterait pas de rentabilité commerciale, il serait essentiel d'inciter les maisons d'éditions à apporter 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 » en tant que de besoin et de verser des rémunérations correspondant à leurs utilisations à la dite 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).

Afin d'élargir l'accès au patrimoine culturel écrit, notamment encore sous droits, tout en respectant le droit de la propriété intellectuelle, une mission a été confiée en 2004 par le ministre français de la culture et de la communication à François Stasse, conseiller d'Etat. Son rapport offre des pistes de réflexion intéressantes en particulier pour ce qui concerne l'accès numériques aux œuvres dans les bibliothèques publiques. L'auteur du rapport cherche à concilier l'accès le plus large à la création culturelle avec la protection du droit d'auteur et s'intéresse notamment à l'accès aux œuvres de la « zone grise », c'est à dire aux œuvres encore protégées par le droit d'auteur mais qui ne sont plus présentes dans le réseau de distribution commerciale. Si les craintes d'usage déloyal des ressources numériques paraissent devoir être prises en compte attentivement lorsqu'elles sont relatives aux publications les plus récentes, on peut en revanche s'interroger sur leur pertinence s'agissant des publications qui, dans les deux à cinq ans après leur mise sur le marché ont été retirées des circuits de distribution commerciale tout en continuant d'être régies par la législation protectrice du droit d'auteur. Quatre propositions sont avancées dans ce rapport :

- permettre la consultation sur place, dans les bibliothèques d'une certaine importance, d'œuvres de la zone grise sous forme numérique,
- permettre, à titre expérimental, la consultation à distance de cette zone grise par un public de chercheurs accrédités et moyennant rémunération,
- autoriser à des fins d'étude, de recherche ou d'enseignement le déchargement payant d'extraits restreints d'un document numérique dans les locaux des bibliothèques publiques,
- permettre la sauvegarde de données numérisées sur des supports ou dans des formats devenus obsolètes.

Le concept de zone grise laisse aux bibliothèques une marge de manœuvre non négligeable. Cependant l'entrée d'un document dans la zone grise ne peut être déduite de manière automatique du seul critère de la date d'édition et doit faire l'objet d'un accord contractuel avec l'éditeur agissant en son nom et en celui de l'auteur. Par ailleurs les titulaires de droits gardent la faculté de retirer de la zone grise toute œuvre qui, après avoir été retirée des circuits de distribution des librairies, connaîtrait un regain d'intérêt susceptible de lui ouvrir une nouvelle carrière commerciale.

Les organisations représentatives des auteurs se sont montrées dès à présent très intéressées par le contenu du rapport de François Stasse tandis que les négociations avec le Syndicat national de l'édition se poursuivent.

*4) Le problème des œuvres orphelines est-il économiquement important et est-il concrètement pertinent de l'aborder ? Si oui quels mécanismes techniques, organisationnels et juridiques pourraient être utilisés pour faciliter une plus large utilisation de ce matériel ?*

L'évaluation du nombre d'œuvres protégées non identifiées et qui seraient susceptibles d'intégrer une bibliothèque numérique est difficile à opérer. L'expérience de constitution d'un fonds numérisé par la Bibliothèque nationale de France en relation avec les représentants des titulaires de droits vers la fin des années 90 tend cependant à démontrer que ce nombre n'est pas négligeable. Ces œuvres semblent être le plus souvent des monographies dont les éditeurs ont disparu ou encore des revues assimilables juridiquement à des œuvres de collaboration et pour lesquelles il est très contraignant de rechercher l'ensemble des auteurs ayant participé à la publication.

Une solution est actuellement à l'étude avec des sociétés de gestion collective afin de permettre l'accès à des œuvres protégées pour lesquelles les coordonnées des ayants droit ne peuvent être identifiées. Cette solution consisterait à constituer un fonds de réserve pour répondre aux réclamations formulées pendant la durée de protection des œuvres.

*5) Comment rendre plus transparente et élargir la diffusion de l'information sur le matériel qui est tombé dans le domaine public et d'autre matériel disponible en libre accès (avec l'accord des ayants droit) afin de faciliter leur accessibilité en ligne pour une utilisation ultérieure ?*

Le projet de BNUE a pour objectif premier d'accroître la visibilité et l'accessibilité du patrimoine culturel écrit européen. La performance de l'outil technique de recherche, liée à la reconnaissance de standards partagés est un bon moyen pour relever ce défi. Le moteur retenu devra notamment répondre aux exigences d'utilisateurs aux compétences très variées et permettre d'enrichir les questions en puisant dans l'information mise à disposition.

Sur le plan technique, afin d'élargir l'accès au patrimoine culturel écrit, l'un des enjeux majeurs est d'encourager les institutions patrimoniales, qui ont développé une forte tradition de coopération nationale et internationale en matière de normes, à utiliser les standards, particulièrement ceux intéressant d'autres domaines comme l'enseignement, les éditeurs, les acteurs commerciaux, les développeurs de moteurs de recherche, etc.

Plusieurs pistes sont à signaler, notamment le méta-langage RDF, le format de métadonnées en XML « ALTO », le protocole OAI-PMH (Open Archive Initiative Protocol for Metadata Harvesting), ou les Web services.

Le RDF (Resource description framework), standard proposé par le W3C (World Wide Web Consortium) pour décrire des ressources, encoder des métadonnées et échanger des descriptions, est capable de répondre à certains besoins dans le domaine de l'interopérabilité, de l'échange de données et de l'accès à l'information. L'interprétation automatique des contenus est nécessaire à une meilleure accessibilité en ligne. Il faut un codage normalisé des contenus pour développer des outils génériques valables dans plusieurs pays. Par exemple, le format de métadonnées en XML « ALTO » est un format spécialement dédié au traitement d'une version textuelle obtenue automatiquement (donc peu coûteuse) en rapport avec l'image à laquelle elle correspond.

Le protocole OAI-PMH (Open Archive Initiative Protocol for Metadata Harvesting) est utilisé pour la création de portails d'accès à des ressources documentaires réparties. Ce standard permet la collecte des métadonnées décrivant des ressources hétérogènes indépendamment des applications utilisées. Il repose sur des standards basés sur XML, tels que le format de métadonnées Dublin Core, et il est couramment utilisé dans les domaines du patrimoine culturel (archives, bibliothèques...)

Les Web services fournissent un cadre pour trouver, décrire et exécuter ces applications. L'architecture étendue des Web services, qui repose notamment sur deux standards basés sur XML, est à l'heure actuelle largement adoptée sur le Web par des acteurs du secteur commercial et industriel. Ils peuvent être une solution pour mettre en place une interrogation commune sur des bases distribuées, transparente pour l'utilisateur, sans modifier ni intégrer les applications. Les Web services permettent également d'exploiter directement ce qui relève des applications développées : moteur de recherche, interface de consultation, et les contenus eux-mêmes.

Enfin, pour favoriser la diffusion et l'exploitation des documents numérisés non susceptibles de reconnaissance optique de caractères traditionnelle (documents manuscrits ou imprimés anciens), les projets de recherche portant sur la reconnaissance de l'écriture manuscrite, et plus généralement la reconnaissance des formes, devraient être encouragés.

*6) Quelles mesures prioritaires - notamment de nature organisationnelle et juridique - devraient être prises au niveau national et au niveau européen pour optimiser la préservation du contenu numérique avec les ressources limitées disponibles ?*

La question de l'archivage et de la préservation des sources ou œuvres numériques, est inséparable d'une maintenance pour permettre aux sources d'être réexploitables ultérieurement. Le choix de conservation doit répondre à des principes généraux et spécifiques, parmi lesquels les possibilités d'utilisation pour des travaux de nature scientifique ultérieurs.

Dans de nombreux pays européens, des travaux sont ainsi menés pour la conservation des documents électroniques.

En France, ces travaux, qui ont abouti au développement de plates-formes d'archivage électronique opérationnelles, sont notamment le fait :

- de la direction des archives de France (et en particulier du Centre des archives contemporaines, depuis une trentaine d'années) et de l'Agence pour le développement de l'administration électronique (ADAE), pour la conservation des documents administratifs ;
- de la Bibliothèque nationale de France, pour la conservation du dépôt légal et de documents numérisés ;
- des grandes institutions scientifiques (CNES, IGN, etc.), pour la conservation de données scientifiques issues d'observations et de calculs ;
- de la Direction de la Documentation française (DDF), pour l'ensemble des rapports et des discours publics ;
- de la Direction des Journaux officiels, pour les textes législatifs.

D'autres projets ont pu être menés dans le secteur privé.

Même si des structures informelles de coordination existent entre ces différentes initiatives, elles mériteraient d'être renforcées au niveau national.

La veille technologique sur les formats et les supports pourrait être mutualisée au niveau européen, comme cela avait été l'ambition du projet DigiCULT sur une période limitée, et comme c'est l'ambition de certaines initiatives déjà en cours (par exemple la base PRONOM des Archives Nationales du Royaume-Uni sur les formats de fichiers). La création d'un répertoire européen, voire international, des formats de fichiers constitue un objectif stratégique (cf. 8).

Les projets de bibliothèques numériques devraient pouvoir bénéficier, davantage que ce n'est le cas actuellement, des savoir-faire et des infrastructures créés à l'occasion de programmes d'archivage électronique développés initialement à d'autres fins (conservation de documents administratifs ou de données scientifiques par exemple).

D'un point de vue organisationnel, il est utile de rappeler que seule la duplication et la mise en place d'archives en « miroirs » (sur plusieurs sites) offre aujourd'hui une garantie minimale de conservation pour les documents numériques.. La mise en place d'une archive organisée et ouverte suivant le modèle OAIS (norme ISO 14721) représente un effort technologique et organisationnel important qui ne peut être rentabilisé que par la masse volumétrique du matériau conservé, donc par la mutualisation des pôles de conservation.

A cet égard, la création d'un standard commun d'échange de données pour l'archivage, destiné à faciliter les échanges de données entre un producteur de données et une plate-forme d'archivage d'une part, et entre deux plates-formes d'archivage différentes d'autre part, est un objectif important. La France fera prochainement une proposition en ce sens.

En terme de gestion globale du risque, le fait de conserver un maximum de documents, y compris sous forme de doubles et de miroirs mais aussi en limitant autant que possible les processus de sélection manuelle qui réduisent considérablement le volume des collections, est la seule garantie qu'au moins une partie des documents collectés aujourd'hui sera préservée sur le long terme. Les calculs de niveaux de risque prennent en compte les facteurs généraux comme l'obsolescence technologique, pour lesquels une réflexion commune est possible, mais aussi les facteurs de risque imprévisibles et ponctuels tels que les accidents physiques subis par les collections. Seule la duplication peut être une réponse à ce second point. Il y a lieu d'inclure dans le cadre de « l'obsolescence technologique » l'indisponibilité de logiciels capables d'exploiter certains formats de données. Le recours aux formats ouverts (de

spécification publique) et au logiciel ouvert (dont le code est accessible, libre ou non de droits) peut permettre de mieux prendre en compte cette préoccupation.

7) Y a-t-il un risque que les systèmes nationaux de dépôt légal conduisent à une multiplication des exigences de dépôt pour les sociétés actives au niveau international ? Une législation européenne aiderait-elle à l'éviter ?

L'objet du dépôt légal est de conserver et de donner accès sur le long terme à l'ensemble du patrimoine culturel d'un pays. Le risque de disparition massive de sources numériques que l'on peut redouter aujourd'hui constitue en effet un véritable problème pour la connaissance de notre société.

Il semble souhaitable d'appliquer le principe de subsidiarité au dépôt légal. Chaque Etat-membre définit un périmètre du patrimoine culturel national à préserver dans le cadre du dépôt légal en fonction de critères qui lui sont propres et des moyens qu'il peut y consacrer. Il est important de préserver cette diversité afin que chaque Etat assure la continuité des collections nationales.

En France, le champ du dépôt légal s'étend à tous les documents publiés et diffusés en nombre sur le territoire national, non seulement la production « officielle » des éditeurs mais aussi la littérature grise accessible au public, donc publiée, telles que les publications institutionnelles, les « pré-prints », les « post-prints » ainsi que les blogs et l'auto-édition qui sont tous rassemblés dans la catégorie des « e-print » sur l'Internet. Dans des domaines comme l'audiovisuel, l'internationalisation de l'édition a eu pour conséquence la duplication des dépôts (qui prennent en compte l'importation) plutôt que des tentatives d'harmonisation supranationales.

Chaque nouvelle technologie a été intégrée dans la réglementation nationale du dépôt légal, élargissant chaque fois son étendue.

Le dépôt légal ne constitue pas une politique d'acquisition, mais une collecte quasi exhaustive des documents pour donner la meilleure image des contenus distribués. Cela a deux conséquences :

- que toutes les différentes versions d'un document, tout comme les œuvres interdites, sont concernées, comme elles l'ont toujours été ;
- que la sélection documentaire n'a pas sa place dans un tel processus et que le dépôt légal doit être autant que possible englobant. S'il ne peut prendre en compte toute la production pour des raisons quantitatives, il procède par échantillonnage mais jamais par sélection documentaire. Car il est impossible de définir ce qui dans la production documentaire actuelle sera utile aux lecteurs de demain.

Les problèmes techniques posés par le dépôt légal des documents numériques ne dépendent pas de l'importance volumétrique de la collecte, ni pour le stockage ni pour le traitement. En effet, pour les documents numériques, il est possible d'automatiser les processus de façon qu'ils s'effectuent de manière simple et peu coûteuse. Pour les documents en ligne, dans la plupart des cas, le dépôt est entièrement à la charge des organismes dépositaires, qui se chargent eux-mêmes de la collecte automatique des contenus par un système de robot. Pour des types de documents qui ne peuvent être collectés de cette manière (Web profond, documents hors ligne), il est souhaitable de demander aux producteurs de greffer sur leur chaîne de production une sortie supplémentaire destinée au dépôt légal.

Il est essentiel que ce processus d'extraction soit normalisé au niveau européen. Ainsi, la charge du dépôt pour les producteurs dont l'activité couvre plusieurs pays serait limitée à l'installation initiale d'une chaîne d'extraction normalisée.

Pour ces raisons, la duplication des obligations de dépôt légal pour les documents numériques ne doit pas être évitée ni même limitée, mais doit être optimisée et organisée de façon à permettre une collecte normalisée et multiple des contenus dans la perspective d'une meilleure conservation à long terme de ces documents. Loin d'être un inconvénient, elle doit être considérée comme un avantage de sécurité et de redondance pour l'accès.

*8) Comment la recherche pourrait-elle contribuer à améliorer la préservation ? Quels axes de travail devraient être suivis en priorité dans les prochains programmes de recherche spécifiques du 7ème programme-cadre ?*

Les programmes de recherche et les réflexions des institutions patrimoniales ont privilégié, jusqu'à une période très récente, les questions d'accessibilité par rapport aux problématiques de préservation. Ces dernières doivent être développées.

La préservation des données numériques est particulièrement dépendante de trois facteurs :

- d'une part, les métadonnées de contexte générées au moment de la collecte de chaque fichier et permettant de comprendre le document d'un point de vue à la fois historique et thématique de sa création ;
- d'autre part, les métadonnées relatives au format de chaque fichier, qui permettront de le relire ultérieurement voire de le convertir dans un autre format ;
- enfin, la fiabilité des supports et des systèmes de stockage.

La recherche doit être encouragée dans ces trois directions.

Concernant la préservation du Web, des outils de collecte et d'accès importants ont été réalisés récemment par l'International Internet Preservation Consortium (IIPC) auquel participent huit bibliothèques nationales européennes. Ces outils permettent de réaliser des collectes nationales de plusieurs dizaines de téraoctets par an et sont déjà utilisés dans plusieurs pays européens. Un travail de normalisation internationale des formats et des métadonnées de préservation propres aux archives de l'Internet accompagne le développement de ces outils ouverts.

Le développement d'un standard d'échange de données pour l'archivage, évoqué plus haut, peut également, si ce standard est implémenté dans les logiciels de gestion utilisés dans les différents secteurs d'activités, faciliter la récupération des métadonnées nécessaires à l'archivage.

Un des axes de travail pour ces outils de collecte, qui concerne à la fois l'accès et la préservation des documents électroniques, pourrait être leur capacité à enregistrer et à exploiter des informations sémantiques sur les contenus collectés. La sélection intelligente des contenus et l'ajout de métadonnées thématiques lors de la collecte découlent des technologies mises en œuvre actuellement dans l'industrie des moteurs de recherche, et une convergence technologique pourrait être envisagée. Un autre enjeu est l'analyse des contenus collectés pour faire émerger des niveaux d'accès logiques (niveau du « site »).

Le deuxième axe de recherche plus générique est celui de la gestion des formats. La collecte des documents numériques implique la prise en compte par les organismes de conservation de nombreux formats de représentation des données numériques, parfois ouverts et largement répandus, mais aussi très souvent confidentiels ou propres à un producteur spécifique. La caractérisation des données est un élément indispensable pour leur préservation ; or l'élément fondamental de l'identification d'un document numérique est le format. Les axes de recherche dans les prochaines années devraient donc inclure une réflexion sur la mise en place d'un répertoire européen, voire international, des formats, qui contribuerait à la détection et à l'analyse des formats – des outils génériques pourraient être développés pour effectuer des vérifications de format (pour avoir la certitude qu'un fichier respecte bien le format prévu et documenté) ou des conversions d'un format à un autre (pour passer d'un format obsolète à un format de conservation) – et à la préservation de leur documentation. Un tel outil serait une formidable garantie de conservation sur le très long terme de toute la production documentaire numérique européenne.

Le troisième axe de recherche pourrait concerner les supports de stockage. En France, par exemple, un groupement d'intérêt scientifique a récemment été constitué entre plusieurs laboratoires de recherche pour étudier les mécanismes de vieillissement des supports optiques afin de prévenir la perte de données. Au-delà des questions scientifiques et techniques, s'ajoute une dimension politique afin d'obtenir des producteurs de supports d'archivage qu'ils fournissent les informations nécessaires à la détermination de la durée de vie.

S'agissant plus généralement du 7<sup>ème</sup> PCRD, il faudrait intégrer à la phase de préparation dans laquelle se trouve ce programme à la fois :

- des recherches en sciences et technologies de l'information et de la communication qui permettent d'améliorer les outils existants pour l'ensemble de la chaîne de numérisation (numérisation des documents, archivage, diffusion),
- des recherches sur les pratiques des utilisateurs qui permettent d'améliorer en permanence le dispositif, ainsi que de définir et de tester des nouveaux usages (outils collaboratifs, plates formes d'annotation, enrichissement des contenus...).

Des observatoires des usages et des technologies pourraient être soutenus en Europe et mis en réseau.