



Tous les chemins mènent au numérique :

→ ARCHIVAGE PÉRENNE, NUMÉRISATION DE MASSE ET COOPÉRATION NUMÉRIQUE À LA BIBLIOTHÈQUE NATIONALE DE FRANCE

EMMANUELLE BERMÈS

emmanuelle.bermes@bnf.fr

MARIE-ÉLISE FRÉON

marie-elise.freon@bnf.fr

FRÉDÉRIC MARTIN

frederic-d.martin@bnf.fr

Bibliothèque nationale de France

Archiviste-paléographe, titulaire du DCB, Emmanuelle Bermès, après avoir été responsable fonctionnel de la bibliothèque numérique de la BnF, est actuellement chef du service « Prospectives et services documentaires » au sein du département de l'Information bibliographique et numérique, à la Direction des services et des réseaux de la BnF. Elle a écrit de nombreux articles sur les ressources électroniques.

Titulaire du CAFB, Marie-Élise Fréon est chef du service numérisation au département de la Conservation de la BnF, après y avoir été chef de projet pour la conservation des documents numériques. Elle a contribué à différents ouvrages et dirigé Contrôler la qualité et la cohérence d'un catalogue (IFB, 1996).

Titulaire d'un DEA de lettres modernes et du DCB, Frédéric Martin est adjoint au chef de service Pôles associés/Gallica de la BnF, après avoir été responsable des techniques documentaires pour les collections numériques. Il est l'auteur d'articles pour diverses revues professionnelles.

La constitution de bibliothèques numériques, dont la masse documentaire est sans cesse croissante, amène à plusieurs réflexions essentielles. Le numérique a-t-il changé la vision du patrimoine pour les bibliothèques qui se sont lancées, depuis la mise en chantier de programmes de numérisation, dans des réalisations d'importance ? Comment faire du patrimoine numérique un bien pérenne, cohérent, aussi fondamental que les collections physiques ? La numérisation, qui apparaissait à la fin des années 1990 comme un chantier de reproduction, est maintenant une pièce maîtresse du dispositif documentaire. Aussi faut-il faire le point sur les enjeux et les évolutions qui conduisent des bibliothèques à constituer des ensembles numériques interopérables. Les réflexions qui ont été conduites à la BnF comprennent plusieurs phases, d'abord la phase de constitution des collections et la mise au point des chaînes de réalisation en lien avec des prestataires de service, puis la pérennisation de ces mêmes collections. Quelques éclairages sur l'activité de la BnF sont proposés dans ce domaine.

* Cet article réunissant trois grandes parties (« La transmission du patrimoine numérique », rédigée par Emmanuelle Bermès ; « La numérisation de masse : rigueur des procédures et contrôle qualité », rédigée par Marie-Élise Fréon et « La coopération numérique : nécessités et opportunités », rédigée par Frédéric Martin) a été coordonné par Thierry Cloarec.

La transmission du patrimoine numérique

La transmission du patrimoine est la mission qui, plus que toute autre, est inhérente à l'identité d'une bibliothèque nationale. Une communauté se construit par l'existence et l'appropriation de son patrimoine, par les objets patrimoniaux dans lesquels elle se représente¹. Le rôle du bibliothécaire, comme celui du chercheur et de l'historien, est de fournir des repères à la communauté à travers ce patrimoine qu'il contribue à forger ; notre rôle est donc de collecter et de conserver au mieux tous les supports de la mémoire quels qu'ils soient, pour pouvoir ultérieurement utiliser ces traces et mieux comprendre notre culture². Ainsi le patrimoine se définit avant tout comme un bien que l'on transmet, richesse héritée des anciens, ou fortune constituée au cours du temps. Cet acte s'avère un véritable défi dès lors que ce patrimoine se compose d'objets dont nous maîtrisons encore mal la durée de vie : les fichiers numériques.

1. Michel Melot, « Qu'est-ce qu'un objet patrimonial ? », *BBF*, 2004, n° 5, p. 5-10.

2. Got, « En haut de Delphes, l'histoire nous regarde », *Les petites cases*, 1^{er} octobre 2006, en ligne : www.lespetitescases.net/en-haut-de-delphes-l-histoire-nous-regarde

Un référentiel commun

Dès la fin des années 1990, la société prend conscience que le passage au « tout-numérique » constitue une menace pour la constitution du patrimoine. C'est ainsi que naissent des métaphores poétiques, comparant la conservation des sources³ aujourd'hui, à un « Moyen Âge documentaire ». Dans les professions directement touchées par le phénomène, archivistes, bibliothécaires, mais aussi scientifiques du domaine aérospatial, concernés par le risque de perte de données qui ont été fort coûteuses à acquérir, cette prise de conscience débouche sur la mise au point d'un modèle conceptuel décrivant un système qui permettra de franchir ce fossé, de conserver les objets numériques sur le long terme : un Système ouvert d'archivage d'information. Ce modèle, l'OAIS⁴, mis au point par le CCSDS⁵ en 1999 et accepté par l'ISO comme norme 14721 en 2002, a été adopté dès 1999 par les grandes bibliothèques patrimoniales impliquées dans des programmes de préservation d'objets numériques⁶. Cette norme constitue aujourd'hui un référentiel commun pour les différentes professions confrontées au défi de la préservation numérique et un guide pour comprendre et maîtriser les enjeux de la conservation d'un patrimoine qui se présente largement sous une forme nouvelle, le numérique.

La démarche de constitution des collections patrimoniales de la BnF, dont le dépôt légal est l'activité la plus

emblématique, s'est ouverte depuis plusieurs années maintenant au numérique, d'abord sur support au sein du département de l'Audiovisuel, puis en ligne avec le dépôt légal de l'internet⁷. En intensifiant ses programmes de numérisation, la BnF constitue également un patrimoine numérique à plusieurs visages qui impose une réflexion approfondie sur les moyens de sa conservation à long terme ; cette réflexion s'appuie en premier lieu sur l'OAIS et débouche aujourd'hui sur la mise en place d'un système d'archivage pérenne, Spar (Système de préservation et d'archivage réparti⁸).

Un système d'archivage pérenne : Spar

Avant d'entrer dans les considérations techniques et fonctionnelles d'un système de préservation numérique, il faut rappeler que l'OAIS, en définissant les fonctions et les responsabilités d'une archive⁹, représente un apport fondamental pour la mission de transmission du patrimoine numérique. La conservation n'est pas un but en soi ; elle doit permettre aux utilisateurs présents et futurs, dans la durée, d'accéder aux objets que l'on a collectés. Conçu suivant ces principes, Spar n'est en aucun cas un entrepôt de données inerte. Il s'agit, tout au contraire, dans la tradition du dépôt légal, d'assurer dans un même mouvement les trois volets : collecte, préservation et diffusion des données.

Dans un premier temps, et conformément aux termes de l'OAIS, un transfert de responsabilité s'opère entre le producteur (dans de nombreux cas, le producteur est représenté par le gestionnaire de collection numérique, c'est-à-dire que ce n'est pas le créateur des fichiers numériques

qui est en prise directe avec Spar, mais l'entité au sein de la BnF qui se charge de leur collecte) et l'archive. Ce transfert de responsabilité fait l'objet d'un accord de qualité de service, un contrat qui décrit de manière concrète les conditions de réalisation des trois moments du cycle de vie des données numériques :

- *le versement* : qui est autorisé à réaliser un versement, quand celui-ci peut-il avoir lieu, et sous quelle forme ?

- *la préservation* : quel niveau de garantie souhaite-t-on (nombre de copies, localisation des copies, fréquence des audits), quelles opérations sont autorisées afin de garantir la lisibilité sur le long terme des données (simple recopie, ré-encodage, migration) ?

- *la diffusion* : qui peut accéder aux données, comment communique-t-on ces données ?

Une fois les documents numériques versés dans l'archive, celle-ci se doit de garantir leur intégrité et leur authenticité, mais aussi de collecter toutes les informations nécessaires pour que ces documents restent compréhensibles pour la communauté d'utilisateurs. Pour cela, l'archive peut appliquer des stratégies de préservation telles que :

- *la migration*, qui implique une transformation du format des objets vers un format plus pérenne et/ou mieux maîtrisé par l'archive ;

- *l'émulation*, qui permet de restituer les objets numériques en simulant leur environnement informatique d'origine.

La mise en place de telles stratégies nécessite une surveillance constante des documents et l'accumulation de nombreuses informations complémentaires : métadonnées techniques, de structure, de droits, informations sur les formats, description de plateformes de référence sur lesquelles on peut faire fonctionner les objets, outils de validation pour contrôler l'intégrité des fichiers et leur conformité aux exigences des accords de qualité de service, etc. Dans Spar, toutes ces données sont stockées dans des fichiers de métadonnées dites d'emballage, au format METS, qui permettent de stocker en XML toutes les informations concernant un

3. Par exemple chez l'écrivain Steward Brant dans *The Clock of the Long Now: Time and Responsibility*, New York, Basic Books, 1999, ou dans le documentaire *Digital Dark Age*, de Jörg Daniel Hissen et Peter Moers, Société Park Film, 2003.

4. Open Archival Information System : <http://public.ccsds.org/publications/archive/650xob1.pdf> (consulté en septembre 2008).

5. Consultative Committee for Space Data Systems, organisme de normalisation dans le domaine de l'aérospatiale.

6. Cf. Catherine Lupovici, « Gestion de la pérennisation des objets numériques : aspects fonctionnels et techniques », *Pérenniser le document numérique*, séminaire Inria, 2-6 octobre 2006, Amboise, ouvrage coordonné par Lisette Calderan, Bernard Hidoine et Jacques Millet, ADBS, 2006.

7. Voir l'article de Gildas Illien dans ce numéro, p. 20-27.

8. À la suite du marché d'acquisition de l'infrastructure, la Bibliothèque nationale de France a lancé en juin 2007 un appel d'offres pour la réalisation de la partie logicielle de Spar, remporté par la société Atos Origin.

9. Au sens de l'OAIS : organisation chargée de conserver l'information pour permettre à une communauté d'utilisateurs cible d'y accéder et de l'utiliser.

objet et garantissant sa pérennité. On transfère ensuite une copie de ces métadonnées vers le module « gestion de données », un entrepôt de métadonnées encodées en RDF (Resource Description Framework), sur lequel on peut effectuer des requêtes complexes qui vont permettre de piloter les opérations de préservation.

Différents acteurs seront chargés de suivre l'évolution des documents au cours de leur cycle de vie dans Spar et de vérifier leur adéquation avec les besoins de la communauté d'utilisateurs :

- *les gestionnaires de collection numérique*: personnes ou entités responsables du contenu des objets archivés ;
- *les administrateurs*: personnes chargées de veiller à la bonne marche du système ;
- *les experts de préservation*: personnes assurant, en lien avec leurs pairs, une veille technique sur la préservation numérique et capables d'édicter des bonnes pratiques et de mettre en place des plans de migration ;
- *les gestionnaires de filière*: personnes ayant délégation pour engager la Bibliothèque vers la préservation à long terme de collections numériques particulières.

Devant la masse (560 téraoctets aujourd'hui) que représentent les données à archiver et leur croissance (1,4 pétaoctets à la fin de la décennie), il est nécessaire de procéder par ensembles. Ainsi, Spar prend en compte les différents types d'objets numériques préservés par la BnF, en les organisant en différentes filières, suivant les exigences de conservation (par exemple, s'agit-il d'un patrimoine faisant l'objet d'une obligation légale, comme dans le cas du dépôt légal, ou de biens propres de l'établissement, comme dans le cas de la numérisation?). Ce point de vue correspond à la logique, générique, de l'archivage et non à celle, circonstancielle, de l'organisation. Un premier recensement a permis de distinguer les filières suivantes : numérisation de conservation, numérisation de reproduction, dépôt légal automatique (web de surface), dépôt légal négocié, archivage légal des documents administratifs et techniques de la BnF, acquisition et don. Par ailleurs, au fur et à mesure de leur

développement pour les collections de la BnF, les modules de Spar seront utilisés pour le tiers-archivage dans le cadre de la politique de coopération nationale de la BnF.

La mise en œuvre en 2008-2009 de la filière traitant la numérisation de conservation permettra de disposer rapidement de l'infrastructure nécessaire pour soutenir les initiatives de numérisation de masse engagées dans le cadre de la mise en place de la bibliothèque numérique européenne, et pour permettre l'évolution de la bibliothèque numérique de la BnF vers plus de performances dans le cadre de Gallica 2¹⁰.

La numérisation de masse : rigueur des procédures et contrôle qualité

Dès le lancement de la numérisation pour la constitution de sa bibliothèque numérique Gallica, la BnF a passé des marchés pour la production d'images numériques. Elle a mis en place des outils et des procédures qu'elle a améliorés au cours du temps pour évaluer l'exhaustivité et la qualité des prestations exécutées.

Une organisation précise

Le bon fonctionnement des projets est assuré par une organisation précise des relations avec les fournisseurs, qui s'appuie sur un échange d'informations nécessaires à la production des différents types de données :

- fichiers images (fac-similé numérique de l'original) ;
- métadonnées ;
- fichiers textes issus de l'OCR, reconnaissance optique de caractères (conversion en haute qualité ou en automatique) ;
- fichier des tables des matières et index (conversion en haute qualité avec lien vers les pages d'entrée).

Au départ des objets à traiter, la BnF fournit deux éléments : un borde-

reau de traitement au format XML¹¹ qui donne les identifiants des documents et précise les traitements demandés (images couleur, OCR haute qualité...). Il est extrait des bases de sélection numériques alimentées à partir du catalogue Bn-Opale Plus qui signale tous les exemplaires existants pour un document, y compris les exemplaires numériques. Ce bordereau contient également des informations bibliographiques permettant l'identification du document, ainsi que le signalement de son état physique. Le second document est un bon d'enlèvement sous forme papier destiné au transport et aux assurances.

Le prestataire accuse réception des objets par bordereau.

Au retour, les livraisons sont accompagnées de plusieurs éléments : un bon de retour papier accompagne les originaux retournés après numérisation et un fichier de métadonnées accompagne chaque document numérique ; il contient :

- la reprise des informations bibliographiques, en particulier l'identifiant de l'original (code à barres) et celui de la notice bibliographique permettant de lui rattacher l'exemplaire numérique créé automatiquement par la BnF lors du chargement ;
- l'identifiant du document numérique attribué par le prestataire à partir d'une liste donnée par la BnF ;
- le cadre de classement Dewey et, éventuellement, la numérotation pour les périodiques, la tomaison... ;
- la table de correspondance entre les images et les pages physiques de l'original, des données de production (fichiers associés aux images tels la table des matières ou les fichiers textes issus de l'OCR, informations sur les traitements effectués, par tel atelier de production, matériels, logiciels) ;
- un bordereau de traitement retour au format XML¹² donne la liste des objets traités ;
- un bordereau de livraison fournit la liste des documents numériques livrés.

L'ensemble des fichiers d'un document est regroupé dans un répertoire

11. <http://bibnum.bnf.fr/demandePrestationConservation>

12. <http://bibnum.bnf.fr/transmissionMessages>

10. <http://gallica2.bnf.fr>

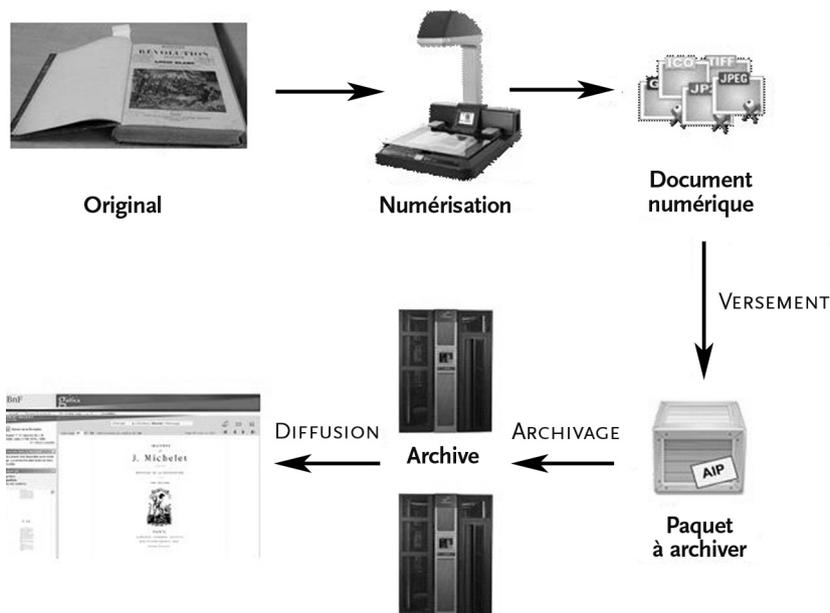
de livraison faisant office de paquet de versement, auquel le prestataire ajoute un fichier d’empreinte permettant de contrôler l’intégrité des données livrées.

Par ailleurs, la chaîne d’entrée BnF effectue un certain nombre de contrôles de structure des fichiers et ajoute dans les métadonnées les résultats et les étapes d’intégration du document (admission partielle ou complète en fonction des résultats des contrôles effectués). Tous ces fichiers alimentent des bases de données BnF afin d’assurer la traçabilité des originaux, des documents numériques versés et leur statut jusqu’à la mise en ligne.

Pour les marchés de numérisation de masse et afin d’assurer une bonne compréhension de ses attentes, la BnF a demandé à ses prestataires de fournir un plan assurance qualité (PAQ), qui détaille les procédures et les outils mis en place pour produire les données requises et le niveau de qualité exigé. Toutes les étapes de la chaîne sont décrites avec les éléments reçus en entrée et les produits fournis en sortie. Un certain nombre de chartes et de référentiels fournis par la BnF figurent en annexe de ce document (règles de numérisation et de production des fichiers texte, des métadonnées, des données de l’exemplaire numérique, de la table de correspondance...). De son côté, la BnF a mis en place son PAQ interne, la diversité des acteurs et la complexité des opérations nécessitant une démarche qualité décrite dans un document dont l’objet est de définir les procédés et outils de travail de chacun en conformité avec le pilotage du projet. Il énonce les principes et objectifs de ce PAQ, et recense un ensemble de procédures révisables correspondant aux étapes clés de la chaîne de traitement (sélection, extraction, pistage, conditionnement, intégration, contrôle, mise en ligne, facturation...), afin que les lots soient prêts en temps voulu.

Formats et fichiers

La charte de numérisation de la BnF permet d’obtenir un document numérique maître pour l’archivage, à partir duquel elle décline les fichiers



Exemple de la chaîne Numérisation de conservation

de diffusion mis en ligne. Tout document numérisé comporte la reproduction en mode image de ses pages, auxquelles on ajoute éventuellement des fichiers XML pour la conversion de la table des matières, des pages de texte en OCR, et pour les métadonnées qui gèrent l’ensemble.

Les images du document maître sont fournies au format Tiff V.6 non compressé, sauf pour celles des textes en noir et blanc compressées en IUT groupe IV (compression entièrement réversible permettant la restitution de tous les pixels à leur place initiale). Le format JPEG a été utilisé jusqu’en 2006 pour la compression des images en niveaux de gris et en couleur, puis a été abandonné (pas de restitution complète et exacte des pixels supprimés). Chaque fichier Tiff comporte un en-tête incluant des informations techniques, de production et de propriété propres à la gestion des images, sous forme codée selon le standard Adobe¹³. La résolution standard est de 300 dpi, mais peut monter jusqu’à 600 si l’original le nécessite.

Les textes convertis par OCR sont conformes au schéma XML Alto (Ana-

lyzed Layout and Text Object¹⁴) qui permet de stocker à la fois la présentation et le contenu d’information. Chaque page convertie produit un fichier « Alto » présentant les coordonnées de chaque élément identifié par l’OCR (blocs de textes y compris coordonnées de chaque ligne et de chaque mot, mais aussi blocs illustrations et graphiques). Ces coordonnées permettent de faire correspondre le texte et l’image originale lorsqu’on les superpose afin de mettre en surbrillance les termes trouvés à l’issue d’une requête.

Le fichier de table des matières permet d’accéder directement à certaines sections du document. Il respecte le schéma tdmNum¹⁵.

Ce format permet d’encoder les niveaux hiérarchiques de la table des matières ou des listes d’index dans un seul fichier, il s’inspire de la TEI (Text Encoding Initiative¹⁶) sous un formalisme très simplifié en utilisant des en-têtes (<head>) pour les intitulés des niveaux répartis dans des divisions (éléments <div>) qui peuvent être typées « T » pour table ou « I » pour index. Au sein d’une division, cha-

13. <http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf> (consulté en septembre 2008).

14. http://bibnum.bnf.fr/ns/alto_prod.xsd

15. <http://bibnum.bnf.fr/tdmNum>

16. www.gutenberg.eu.org/publications/autres/TEILITE

que entrée fait le lien vers l'image qui contient la partie référencée et vers le numéro de page concernée.

L'encodage TEI est également utilisé pour publier une édition textuelle d'un document dont le contenu sera restitué avec toute sa richesse typographique et sa mise en page d'origine. La BnF encode les données selon un profil d'application local de la DTD TEI, c'est le cas par exemple pour la revue de synthèse.

Les métadonnées au format XML suivent les règles du schéma refNum¹⁷. Les informations sont regroupées dans trois éléments de base : bibliographie (description et identifiants), production (nombre de pages et de vues, historique des traitements...) et structure (table de correspondance avec pour chaque page des commentaires éventuels ou une légende).

La coopération numérique : nécessités et opportunités

À mesure que s'accroît l'expérience des bibliothèques en matière de numérisation, de nouveaux champs de collaboration émergent. Des questions se posent aussi : comment mieux articuler les projets entre eux ? Quelle mutualisation des efforts peut être envisagée ? Quelle cohérence et quelle complémentarité doit-on rechercher au-delà des priorités fixées par chaque établissement ? Les trois axes suivants tentent d'y répondre.

La numérisation concertée

La constitution d'une bibliothèque numérique suppose plusieurs étapes (sélection des corpus à numériser, mise en ligne, signalement des collections, conservation numérique), où la démarche collaborative est amenée à jouer un rôle de plus en plus grand. Tout d'abord, les bibliothèques trouveront un intérêt majeur à établir conjointement des listes de docu-

ments à numériser, en particulier dans les thématiques susceptibles d'être portées par plusieurs établissements. Les avantages sont évidents. Non seulement on évite que la même édition soit numérisée plusieurs fois, mais on peut utiliser l'exemplaire qui présente les meilleures caractéristiques pour cette opération. Enfin, l'établissement de listes communes permet une plus large couverture de la thématique, à un niveau de complétude qu'une bibliothèque seule peut rarement atteindre, notamment en ce qui concerne les séries. L'offre documentaire aux utilisateurs s'en trouve accrue, et les moyens financiers et humains mieux employés.

C'est dans cette optique que s'est constitué, en mai dernier, le groupe pour la numérisation concertée en sciences juridiques¹⁸. À l'initiative de la BnF et de la bibliothèque interuniversitaire Cujas, ce programme réunit plusieurs établissements désireux de numériser leurs fonds juridiques, tout en les insérant dans une offre globale au niveau national. Il ne s'agit pas seulement de rapprocher les bibliothèques, mais d'associer aussi en amont les utilisateurs finaux des futures collections numériques, chercheurs et professionnels du droit (notaires, avocats, documentalistes...).

De manière pragmatique, le travail débute par l'examen de listes d'ouvrages fournies par les chercheurs et la répartition de l'effort de numérisation en fonction des collections présentes dans chaque établissement. À terme, nous devrions aboutir à la mise en place d'un cadre institutionnel au niveau national, réunissant les principales tutelles ministérielles. L'objectif

sera de définir les modalités scientifiques, administratives et financières de la numérisation dans le domaine juridique, en validant des axes thématiques de travail. Cette entreprise de

“Gallica joue le rôle de métamoteur numérique généraliste, une vocation confirmée par l'expérimentation avec les éditeurs dans Gallica 2”

numérisation concertée, une première pour la BnF à cette échelle, permettra de tester une nouvelle méthodologie en vue de l'étendre à d'autres disciplines.

Le signalement fédéré

La convergence des initiatives de numérisation ne signifie pas pour autant une confiscation des fonds numériques au profit de tous, ni la négation de la part apportée par chacun. Nous disposons en effet de solutions techniques qui permettent à la fois de créer des portails thématiques et largement fédérateurs, tout en laissant la possibilité à chaque bibliothèque de valoriser ses fonds sur son propre site internet. Nous multiplions ainsi les points d'accès aux documents numériques en donnant le maximum de visibilité aux ressources. L'interopérabilité est la déclinaison technique de la notion de coopération.

À cet égard, l'utilisation du protocole technique OAI-PMH (Open Archive Initiative Protocol for Metadata Harvesting) doit être encouragée.

17. <http://bibnum.bnf.fr/refNum>

18. Les actes de la 1^{re} Journée d'information et d'échange sont disponibles à l'URL : www.bnf.fr/pages/infopro/journeespro/jp_num_juridique.htm (consulté en septembre 2008). Voir le compte rendu de la journée dans le *BBF*, 2008, n° 5, p. 93-94.

Cette technique de diffusion et de collecte des métadonnées facilite le référencement des documents numérisés dans des contextes très différents (bibliothèque numérique locale, portail régional, portail thématique, Gallica, et plus tard Bibliothèque numérique européenne...). Par ce procédé, le site Gallica proposé par la BnF référence déjà plus de 6 000 documents numérisés par les bibliothèques partenaires. Seules les notices, comportant le lien vers le document, sont indexées dans le moteur de Gallica : les documents sont consultables sur les sites des bibliothèques associées. Gallica joue ainsi le rôle de méta-moteur numérique généraliste, une vocation confirmée par l'expérimentation avec les éditeurs dans Gallica 2.

Reposant également sur le protocole OAI-PMH, cette initiative vise à compléter les collections du domaine public par une offre sous droits, dont les modalités de consultation sont établies par les éditeurs ou les *e*-distributeurs qui procèdent en leur nom. Lancée à l'occasion du Salon du livre de Paris 2008, en collaboration avec le ministère de la Culture et de la Communication (Direction du livre et de la lecture), le Centre national du livre et le Syndicat national de l'édition, l'opération réunit une centaine d'éditeurs donnant accès à 4700 documents¹⁹. Selon le titre et le site partenaire, on peut feuilleter le document gratuitement, le louer pour une durée déterminée, le télécharger sur son ordinateur, son PDA ou son téléphone portable ou le commander en ligne. Cette expérimentation fera l'objet d'une évaluation à la fin de l'année 2008, à la fois sur les plans technique et documentaire.

La conservation mutualisée

La numérisation patrimoniale pose de façon cruciale la question de la conservation des données ainsi produites. Sommes-nous certains que, dans dix ans, une période somme toute assez courte, les documents pro-

duits seront encore consultables et que les efforts consentis à leur création ne seront pas définitivement perdus ? Certaines bibliothèques font déjà état d'une perte de leurs données, de façon plus ou moins marginale. L'absence de stratégie dans ce domaine conduit inéluctablement au pire. Si les coûts de stockage informatique ne cessent de décroître, cela ne doit pas faire illusion. L'archivage à long terme ne se

“L'archivage à long terme ne se résume pas au stockage mais nécessite la mise en place d'un dispositif plus complexe”

résume pas au stockage mais nécessite la mise en place d'un dispositif plus complexe, capable de réaliser des opérations spécifiques (empreinte numérique, conversion de formats, migration de supports).

Pour ses propres collections numériques, le système Spar de la BnF sera capable de réaliser de telles opérations, sur un ensemble important de documents. D'emblée s'est posée la question de l'ouverture de ce système à d'autres acteurs, confrontés aux mêmes problèmes et souhaitant bénéficier des technologies et des savoir-faire de la BnF dans ce domaine. Afin de mieux cerner les besoins, la BnF s'est associée à la CDC numérique, filiale de la Caisse des dépôts, pour réaliser une étude durant le premier semestre 2008, sur la valorisation numérique du patrimoine relevant des organismes publics territoriaux. Cette étude, qui donnera lieu à restitution publique, est une première étape dans la mise en œuvre en 2009 d'une offre de services aux bibliothèques partenaires de la BnF. L'objectif

est de proposer à coûts maîtrisés un service de tiers archivage comportant les mêmes garanties de sécurité et de pérennité que celles mises en œuvre pour les collections patrimoniales de la BnF, en l'articulant avec des services de valorisation (identifiants pérennes, entrepôt OAI, enrichissement des métadonnées, etc., intégration des fichiers dans les interfaces de visualisation de Gallica 2, etc.).

Dans un domaine particulièrement évolutif, nécessitant le renouvellement régulier des technologies, la mutualisation peut être une réponse aux institutions qui ne peuvent ou ne veulent supporter les efforts de façon isolée.

Le monde numérique est un espace privilégié pour la coopération, pour lequel il reste encore beaucoup à accomplir. Les trois axes décrits ici sont des orientations qu'il s'agira de consolider dans un avenir proche, à un niveau très opérationnel, en les articulant avec

les actions menées par la Direction du livre et de la lecture, la Mission de la recherche et de la technologie et la Sous-direction des bibliothèques et de l'information scientifique (signalement des programmes et carte numérique documentaire, élaboration de guides et de recommandations, soutien aux programmes de numérisation, etc.).

Le partage d'expériences, la complémentarité des compétences demeurent les ingrédients de la réussite collective. Pour paraphraser Tim Berners-Lee, « *Intelligent digitization requires co-operation* »²⁰. ●

Septembre 2008

20. « Une numérisation intelligente requiert la coopération ». « Berners-Lee: intelligent web requires co-operation » est le titre d'un article de ZDNet résumant l'intervention de l'inventeur du web à une conférence à Southampton en septembre 2006. <http://news.zdnet.co.uk/internet/0,1000000097,39283443,00.htm> (consulté en septembre 2008).

19. Pour un objectif de 10 000 documents fin 2008. Plus d'informations sont disponibles sur Gallica 2 : <http://gallica2.bnf.fr/moreInfos>