



# Le patrimoine passe au numérique

## ISABELLE WESTEEL

Conseil régional Nord-Pas-de-Calais  
Direction de la culture  
i.westeel@cr-npdc.fr

*Archiviste-paléographe, conservateur des bibliothèques, titulaire d'un DEA d'histoire, Isabelle Westeel a travaillé au SCD de Lille-III puis à la bibliothèque municipale de Lille, où elle a organisé la réinformatisation complète du réseau et la mise en place de la bibliothèque numérique. Elle est depuis le 1<sup>er</sup> novembre 2008 en poste à la Direction de la culture du conseil régional Nord-Pas-de-Calais, où elle s'occupe de projets de valorisation par les nouvelles technologies. Responsable de formations sur le numérique et XML à l'Enssib et dans plusieurs centres de formation régionaux, elle mène également des travaux de recherche sur l'hagiographie et sur l'histoire des bibliothèques (centre Gabriel-Naudé/Enssib).*

L'apport des nouvelles technologies et en particulier du numérique aux besoins des bibliothèques patrimoniales se mesure essentiellement en termes d'accessibilité et de diffusion des contenus. L'utilisation massive du web et la prise en compte des besoins de nouveaux publics nécessitent de modifier la gestion des collections patrimoniales et de s'occuper sérieusement du numérique<sup>1</sup>.

## L'informatisation des catalogues

Depuis quinze ans environ, les bibliothèques municipales et universitaires ont largement fait profiter le public des avantages du numérique grâce aux opérations de conversion rétrospective (ou rétroconversion) des catalogues anciens. Le premier bilan de l'enquête DLL-Drac (Direction du livre et de la lecture – directions régionales des affaires culturelles) montre des situations très diverses selon les régions<sup>2</sup>. En 2008, on estime encore à 36 % les collections ni inventoriées, ni cataloguées (26 % pour l'imprimé)<sup>3</sup>.

La visibilité nationale et internationale de ces données patrimoniales passe essentiellement par le portail national CCFr (Catalogue collectif de France), qui permet déjà de localiser plus de vingt millions de notices et qui a pour vocation de fédérer à terme les

catalogues locaux et régionaux pour faciliter l'accès aux collections physiques et numérisées conservées par les bibliothèques sur tout le territoire national.

Ces opérations de conversion de données, désormais parfaitement maîtrisées par un certain nombre de prestataires, restent lourdes pour les établissements : préparation et correction préalable des fichiers ou des inventaires papier quand ils existent, mise au point des spécifications particulières de traitement pour intégration dans les logiciels de gestion des bibliothèques, correction d'anomalies et enfin injection dans le catalogue et création des exemplaires. Ces étapes demandent des compétences croisées et triples : patrimoniales, bibliothéconomiques et informatiques, qui ne peuvent s'envisager que par un travail d'équipe et intergénérationnel au sein des établissements. À ces vertus, il faudra ajouter des connaissances en histoire du livre, en bibliographie matérielle et éventuellement en langues anciennes. Là encore, une connaissance du fonds et de son origine peut se révéler précieuse.

Dans le cas du catalogage à la source, l'importation de notices issues de grands réservoirs comme BN-Opale Plus permet de gagner du temps. En octobre 2008, l'Abes (Agence bibliographique de l'enseignement supérieur) a annoncé son accord avec le Consortium of European Research Libraries<sup>4</sup> (Cerl) pour l'utilisation par huit bibliothèques du réseau Sudoc de la base Heritage of the Printed Book (HPB), afin d'en dériver des notices mais aussi de l'alimenter.

1. Je remercie Raphaële Mouren et François Houste (questions techniques de référencement et d'évolution du web) pour leurs conseils et leur expertise. Je remercie également Thierry Claerr qui a mis au point la partie sur les projets régionaux.

2. [www.bnf.fr/pages/infopro/journeespro/pdf/poles\\_pdf/poles2006\\_pdf/Plazannet.pdf](http://www.bnf.fr/pages/infopro/journeespro/pdf/poles_pdf/poles2006_pdf/Plazannet.pdf)

3. [www.culture.gouv.fr/culture/dll/pat/PAPE/PAPEbilan2008.ppt](http://www.culture.gouv.fr/culture/dll/pat/PAPE/PAPEbilan2008.ppt)

4. [www.cerl.org](http://www.cerl.org)

The screenshot shows a web interface for the digital library of the Lille Municipal Library. The main title is 'Carton 35 : Lille : Château, fortifications, portes'. Below the title, there is a table with the following information:

Date	
Lieu de conservation	Cote : Carton 35
Quantité	29 documents

Navigation elements include 'Page 1 / 2', 'Résultats 1 à 20 sur 30', and 'Dossier: carton\_35'. The main content area displays a grid of 14 image thumbnails, each with a label like 'ca35\_01' through 'ca35\_14'. A tooltip is visible over thumbnail 'ca35\_06', containing the following text:

Bibliothèque Municipale de Lille, carton 35\_6  
Lille : Porte St Pierre en 1525  
Lille (Nord) -- Porte Saint-Pierre. Portes de ville -- France -- Lille (Nord).  
carton\_35/ca35\_06

At the bottom of the interface, there is a search bar with the text 'Rechercher' and 'Ok', and a list of filters: 'Titres', 'Auteurs', 'Sujets', and 'Références'. The URL at the bottom is 'http://numerique.bibliotheque.bm-lille.fr/sdx/num/carton\_35/ca35\_06?p=1&next=carton\_35/ca35\_07'.

Se promener dans les collections numériques : la bibliothèque numérique de la BM de Lille [<http://numerique.bibliotheque.bm-lille.fr>]

## La bulle du numérique ?

Le catalogue des collections numérisées Patrimoine numérique<sup>5</sup> dresse un panorama de la numérisation du patrimoine écrit en France : plus de 500 institutions culturelles impliquées dans plus de 1400 opérations de numérisation. Ces chiffres, qui attestent une activité certaine dans ce domaine, sont néanmoins trompeurs. Une observation précise de la situation des bibliothèques (157 établissements) montre un bilan plutôt mitigé. 40 % des projets signalés dans l'ensemble des bibliothèques (plus de 50 % des projets des bibliothèques municipales) concernent la campagne nationale de numérisation d'enluminures des manuscrits médiévaux menée par l'IRHT (Institut de recherche et d'histoire des textes) en partenariat avec la Direction du livre et de la lecture. Les projets en ligne et les véritables bibliothèques numériques (on entend par là des collections organisées) sont fina-

lement assez peu nombreux. On peut compter une trentaine de projets pour les bibliothèques municipales : Troyes, Lyon, Lille, Roubaix, Valenciennes, Toulouse...

D'autres types d'institutions gérant le patrimoine, comme les services de l'inventaire, montrent aussi des situations contrastées, mais ont l'avantage de bénéficier de la visibilité de bases nationales (Mérimée pour l'architecture, Palissy pour les objets mobiliers, Mémoire pour les illustrations des œuvres) accessibles depuis début 2007 par le moteur de recherche « Collections » (onglet « Collections » du portail Culture.fr) donnant accès à trois millions de notices et deux millions d'images de données patrimoniales et permettant une recherche dans 30 sources documentaires différentes.

Quelles sont les raisons de ce bilan en demi-teinte malgré la multiplication des journées d'étude et des stages de formation consacrés à la numérisation, malgré les efforts financiers importants de l'État par le biais des « appels à projets de numérisation » ? Dans les petits et moyens établissements,

construire et surtout maintenir dans la durée des projets de valorisation du patrimoine par le numérique est un travail compliqué. La cartographie des projets est encore trop souvent subordonnée aux tableaux de mutation des personnels de bibliothèques. Il est vrai que la matière est complexe.

Aux difficultés liées au respect du calendrier budgétaire, s'ajoute la carence de l'offre logicielle dans le domaine de la mise en ligne des contenus numérisés. Les sociétés spécialisées dans l'informatique documentaire peinent à inventer et à développer des systèmes de navigation et des moteurs de recherche satisfaisants pour tous les types de documents patrimoniaux. Le feuilletage d'un ouvrage ancien et la consultation de photographies cataloguées à la pièce au sein d'un dossier imposent des systèmes de navigation différents et parfois difficiles à mettre en cohérence. Au lieu de s'obstiner à utiliser des outils intégrés, lourds à maintenir et répondant difficilement aux contraintes du web et aux envies du public, ne faudrait-il pas adopter une stratégie

5. [www.numerique.culture.fr](http://www.numerique.culture.fr)

technique pragmatique, consistant à mettre en œuvre des outils de consultation adaptés aux supports ?

La fabrication de mini-sites pour mettre en valeur un support, un fonds, une collection ou un thème paraît être une solution viable, quitte à les intégrer au sein d'un portail et à permettre des recherches fédérées. On peut prendre exemple sur quelques sites anglo-saxons (par exemple la bibliothèque de l'université d'Utah<sup>6</sup> ou celle de l'université de Columbia<sup>7</sup>) qui préfèrent proposer une offre thématique en utilisant des logiciels indépendants pour la recherche et la consultation. Cette démarche participe presque de l'éditorialisation des collections. En France, quelques établissements complètent la mise en place de bases de données traditionnelles par des sites particuliers qui permettent d'ajouter du sens aux documents<sup>8</sup>.

On touche là à la nécessaire réflexion sur l'accès aux documents, qui ne saurait calquer les catalogues des bibliothèques : il faut donner à voir une collection numérique avant d'offrir une recherche précise, les promenades libres ou guidées ont une signification. Cette démarche multiplie les difficultés de conception et alourdit le travail de mise en place d'une collection numérique sur internet.

## Les plans régionaux de numérisation

Dans le domaine de la valorisation du patrimoine culturel, la région est un échelon particulièrement pertinent<sup>9</sup> pour la mise en place de politi-

ques concertées sous la forme de plans régionaux de numérisation et de portails régionaux de ressources numériques, pour le déploiement du volet numérisation du PAPE (Plan d'action pour le patrimoine écrit) ou encore pour l'articulation des programmes de coopération menés par la BnF.

Les objectifs des plans régionaux de numérisation sont principalement de favoriser la conservation, la connaissance, la diffusion ainsi que l'exploitation intellectuelle et économique du patrimoine dispersé physiquement dans les bibliothèques, les services d'archives et les musées, de garantir la cohérence, l'accessibilité et la pérennité des réalisations, de développer de nouveaux services pour les citoyens, enfin de valoriser le patrimoine immatériel et d'inscrire les territoires dans l'économie du numérique en renforçant l'offre culturelle et touristique.

Sur la période 2000-2006, plusieurs contrats de plans État-Région (CPER) ont pris en compte la numérisation des collections des institutions culturelles. La médiathèque Les silos de Chaumont en Champagne-Ardenne par exemple a pu numériser les affiches de la collection Dutailly<sup>10</sup>, avec un taux de subvention de l'État et de la Région atteignant 70 % de l'opération. En Picardie, le conseil régional et la direction régionale des affaires culturelles ont mis en place, dès 2001, un plan régional de numérisation, sous la forme d'un protocole d'accord et d'un appel à projets annuel à destination des musées et de trente bibliothèques éligibles (bibliothèques municipales ou intercommunales regroupant au moins 10 000 habitants, dotées d'agents qualifiés et disposant de fonds anciens ou régionaux spécifiques).

La région Aquitaine a développé sur la même période le projet original de Banque numérique du savoir en Aquitaine (BNSA)<sup>11</sup>. Le dispositif,

bibliothèques et centres de documentation de Bretagne. Les actes sont consultables à partir du site Britalis : [www.britalis.org](http://www.britalis.org)

10. <http://silos.ville-chaumont.fr>

11. Site « Aquitaine patrimoines, le portail culturel des patrimoines de l'Aquitaine » <http://bnsa.patrimoines.aquitaine.fr/apd/web>

bénéficiant également de subventions européennes (Feder – Fonds européen de développement régional), est reconduit pour 2007-2013. Ce programme innovant s'inspire aussi des objectifs définis par les Conseils européens de Lisbonne et de Göteborg (recherche, innovation, attractivité territoriale).

En Rhône-Alpes, Lectura, portail des bibliothèques des villes-centres de Rhône-Alpes lancé en septembre 2006, a bénéficié d'une inscription dans plusieurs dispositifs : le contrat de plan État-Région 2000-2006, la convention cadre de réseau 2003-2007 signée par la Région et les huit villes, la convention triennale État/Région/Arald 2000-2006. La maîtrise d'ouvrage a été confiée à l'Arald et le réseau insiste sur le partage des compétences.

Pour la période 2007-2013, deux nouvelles régions (la Lorraine et la Franche-Comté) bénéficient de contrats de projets État-Région incluant la numérisation des collections de bibliothèques. En Lorraine<sup>12</sup>, un programme fédérateur concernant le patrimoine écrit, documentaire et filmique, patrimoine commun aux musées, aux archives et aux bibliothèques de Lorraine, mais qui intéresse aussi les universités de Nancy et de Metz, les laboratoires de recherche (Inalf – Institut national de la langue française par exemple) et les entreprises, est mis en œuvre.

En Franche-Comté, dans le prolongement du CPER 2000-2006 qui avait contribué à la création d'un portail régional des musées<sup>13</sup> et à la réalisation d'un « musée virtuel » du patrimoine muséographique avec l'aide de crédits Feder, le contrat de plan État-Région 2007-2013 prévoit la mise en place d'un fonds d'intervention régional interdisciplinaire de numérisation pour les bibliothèques, les archives et les musées.

Ces exemples paraissent d'autant plus intéressants que leur modèle organisationnel (état des lieux, concertation entre les acteurs, mutualisation

12. [www.lorraine.eu/jahia/webdav/site/e-internet/shared/documents/conseil\\_regional/publications/cper.pdf](http://www.lorraine.eu/jahia/webdav/site/e-internet/shared/documents/conseil_regional/publications/cper.pdf)

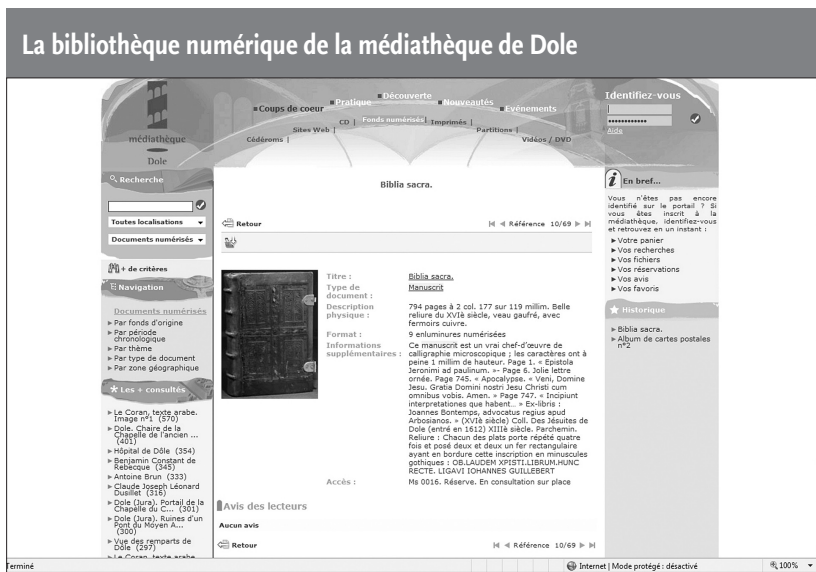
13. [www.musees-franche-comte.com](http://www.musees-franche-comte.com)

## Les structures régionales pour le livre

Pour mettre en place un plan régional de numérisation, l'existence d'une structure de coopération apparaît comme un atout indéniable et, dans certaines régions, comme une condition *sine qua non*. En Bretagne, la Cobb, agence de coopération des bibliothèques et centres de documentation de Bretagne, devenue Livre et lecture en Bretagne, anime depuis 2002 une commission numérisation qui a proposé, dans le cadre de la préparation du contrat de projets État-Région 2007-2013, un programme régional particulièrement ambitieux « pour la numérisation et la valorisation du patrimoine écrit et graphique de Bretagne », concernant aussi bien les aspects archivistiques, sonores (notamment en langue bretonne) et musicaux qu'ethnologiques et audiovisuels, et leur mise en ligne par le portail régional documentaire Britalis<sup>16</sup>.

En Basse-Normandie, c'est le centre régional des lettres (CRL) qui anime la bibliothèque numérique Normannia<sup>17</sup>. L'agence de coopération Interbibly<sup>18</sup> en Champagne-Ardenne organise régulièrement des journées d'étude sur les questions de la numérisation.

Pour répondre au besoin de coordination régionale en matière de numérisation, plusieurs structures régionales du livre ont élargi les activités de leur commission Patrimoine à la numérisation, comme Picasco<sup>19</sup>, agence régionale du livre en Picardie, ou se sont dotées d'une commission numérisation, comme la C2LR<sup>20</sup> (Agence de coopération pour le livre en Languedoc-Roussillon), devenue association Languedoc-Roussillon Livre et lecture, qui a réuni, dès 2006, une commission « numérisation-réseau régional ».



La bibliothèque numérique de la médiathèque de Dole<sup>1</sup> rend accessibles, sur la même interface, en recherche ou par navigation, des manuscrits médiévaux et modernes, des incunables, des livres imprimés anciens, une documentation sur l'histoire locale et régionale, des documents d'archives privées, des cartes et plans et des documents iconographiques sur Dole et le Jura (actuellement 3 000 documents en ligne : 90 % d'iconographie et 10 % de textes), tous supports catalogués en Dublin Core. La recherche fédérée avec le catalogue de la médiathèque est possible. En 2008-2009, un programme d'enver-

ture a été lancé concernant 60 000 images environ avec le concours de la BnF (pôle associé Franche-Comté), du programme PAPE et de la Région (CPER). Une bibliothécaire, chargée de mission sur la numérisation et la coopération, suit l'ensemble du projet au quotidien : cahiers des charges, inventaire, catalogage et mise en ligne. Des perspectives de collaboration sont envisagées avec le portail numérique de Franche-Comté (à l'étude auprès de l'agence régionale de coopération Accolad) ainsi qu'avec Gallica par l'intermédiaire de l'OAI (Open Archives Initiative)<sup>2</sup>.

I.W.

1. [www.dole.org](http://www.dole.org)

2. [www.culture.gouv.fr/culture/dll/OAI-PMH.htm](http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm)

des compétences et des moyens...) est transposable à d'autres régions.

La richesse patrimoniale d'autres régions justifierait la mise en place d'un dispositif régional. Les potentialités<sup>14</sup> de la région Nord-Pas-de-Calais ne sont plus à démontrer : six bibliothèques municipales classées (Boulogne-sur-Mer, Cambrai, Douai, Lille, Roubaix, Valenciennes), d'autres fonds importants – Aire-sur-la-Lys, Arras, Calais, Dunkerque, Saint-Omer, Tourcoing, la réserve ancienne des trois universités lilloises. Le recensement Patrimoine numérique cite une soixan-

taine de projets et de réalisations dans différentes institutions culturelles. Les réalisations concrètes sont déjà nombreuses : Musenor, PôLiB, NordNum, Libris, bibliothèques numériques de Lille, Roubaix et Valenciennes. La complémentarité des fonds est évidente notamment pour l'iconographie locale. Le manque de coopération en matière de numérisation, malgré la volonté des établissements, tient en grande partie à l'absence de structure fédératrice<sup>15</sup>.

15. Le centre régional du livre et des lettres, actuellement en phase de préfiguration, envisage de développer une politique régionale du patrimoine écrit en liaison avec les établissements concernés.

14. Le site Eulalie propose une histoire des bibliothèques du Nord-Pas-de-Calais : [www.eulalie.fr/spip.php?rubrique120](http://www.eulalie.fr/spip.php?rubrique120)

16. [www.britalis.org](http://www.britalis.org)

17. [www.normannia.info](http://www.normannia.info)

18. [www.interbibly.fr](http://www.interbibly.fr)

19. [www.picasco.org](http://www.picasco.org)

20. [www.c2lr.net/html/1\\_c2lr/1\\_commissions.html](http://www.c2lr.net/html/1_c2lr/1_commissions.html)



## Sur le plan national : le « Schéma numérique pour les bibliothèques »

Face aux difficultés des établissements de construire, maintenir et pérenniser une offre numérique, la solution est sans doute d'accentuer les collaborations entre établissements et avec la BnF qui s'est dotée de moyens pour relayer une offre nationale diversifiée. Ainsi l'exposition des données sur Gallica par le biais du protocole OAI, réalisée pour la Bibliothèque interuniversitaire de médecine (Bium), le Conservatoire numérique des arts et métiers (Cnum), la bibliothèque numérique du SICD universités de Strasbourg et les bibliothèques virtuelles humanistes (BVH), est la garantie d'une visibilité accrue au niveau national et, par conséquent, au niveau européen sur Europeana.

Constituée à partir des collections de la BnF, l'offre documentaire de Gallica s'est construite aussi grâce à plusieurs partenariats : la numérisation partagée du fonds Philidor de musique baroque, réparti entre la BnF et la bibliothèque municipale de Versailles<sup>21</sup> ; le partenariat avec la Bibliothèque de documentation internationale contemporaine (BDIC), autour de la numérisation des journaux français des tranchées de la guerre 1914-1918<sup>22</sup> ou avec la médiathèque de l'agglomération troyenne, pôle associé pour le signalement et la numérisation de la littérature de colportage<sup>23</sup>. On peut citer surtout le programme national de numérisation des revues des sociétés savantes. Les revues des régions Aquitaine, Auvergne, Bourgogne, Bretagne, Lorraine, Poitou-Charentes sont déjà intégralement en ligne, soit plus de 1,2 million de pages provenant de cent dix-huit revues parues de 1750 à 1914 et émanant de cent trois sociétés. Les publications de huit autres régions sont partiellement consultables en ligne ou en cours de numérisation : Île-de-France (hors Paris), Languedoc-Roussillon, Limousin, Basse- et Haute-

Normandie, Pays de la Loire, Picardie, Rhône-Alpes. La presse locale et régionale est une matière propice aux projets de collaboration.

Le rapport Livre 2010 publié en juin 2007 a préconisé la mise en place d'un « schéma directeur du numérique en bibliothèque » afin « d'assurer la cohérence du projet BnF et les autres initiatives du secteur public (enseignement supérieur et recherche, collectivités territoriales) » et de « concevoir une organisation partagée et un développement cohérent des ressources numériques pour les bibliothèques françaises ». « Ce schéma directeur du numérique en bibliothèque pourrait rassembler un guide des bonnes pratiques et des normes à utiliser, qui organiserait le recensement et l'analyse des projets de numérisation menés en bibliothèque, proposerait des solutions pérennes en matière de conservation numérique, et enfin favoriserait l'émergence de relais régionaux en matière de coordination des politiques numériques. » Cette idée a été reprise par l'atelier<sup>24</sup> réuni le 30 juin 2008 sur le thème « Les services de bibliothèque et de documentation, acteurs de la chaîne numérique » par l'Interassociation archives bibliothèques documentation (IABD) dans le cadre des assises du numérique.

Le rapport « France numérique 2012 : plan de développement de l'économie numérique<sup>25</sup> », publié en octobre 2008 et listant 150 propositions pour réduire la fracture numérique, reprend presque mot pour mot cette proposition dans l'action n° 55. Cette action, majeure pour les bibliothèques, est complétée par d'autres visant à accroître la production et l'offre de contenus numériques et à développer la visibilité, l'accessibilité et l'interopérabilité des services numériques et intéressant directement les bibliothèques (actions 39, 40, 41, 42, 52, 54, 94, 95). Les questions juridiques liées à la reproduction d'œuvres du domaine public détenues par les institutions culturelles françaises feront l'objet d'une étude particulière. Des groupes de travail préparent des propositions pour construire la future « Bibliothèque numérique de France ».

Dans les établissements, le recadrage des procédures de subventionnement est souhaité, tout comme la diffusion de recommandations techniques pour assurer l'interopérabilité des systèmes par le protocole OAI en particulier.

Les enjeux du numérique sont évidemment politiques. Le web et son fonctionnement conditionnent cependant certaines stratégies.

## La base : rendre visibles les contenus des moteurs de recherche

Les internautes ne commencent presque jamais leurs recherches sur les sites des bibliothèques patiemment et sagement construits par les bibliothécaires. Ils y arrivent très souvent par hasard et suite à l'interrogation d'un moteur de recherche généraliste. À une époque où plus de 90 % des recherches sur le web en France passent par l'intermédiaire du moteur Google<sup>26</sup>, il est important de rendre les données accessibles et d'élaborer des stratégies de visibilité. Pour ce faire, on peut utiliser les méthodes du référencement naturel ou Search Engine Optimization (SEO) consistant à placer en meilleure position possible dans les moteurs de recherche une page web pour un mot ou une expression donnés. « Bien plus qu'une simple opération technique, le référencement naturel est une stratégie d'acquisition d'audience globale, qui couvre les étapes allant de la définition même de cette audience (quel public ?) à la satisfaction de celle-ci par le biais des contenus disponibles sur un site. »

Les méthodes complexes permettant d'arriver à un résultat sont maîtrisées par des agences spécialisées dans le référencement qui travaillent sur-

21. [www.bibliotheques.versailles.fr/Statique/pages/collections-numerisees/presentation.htm](http://www.bibliotheques.versailles.fr/Statique/pages/collections-numerisees/presentation.htm)

22. [www.bdic.fr](http://www.bdic.fr)

23. [www.mediathèque-agglomeration-troyes.fr/bmtroyes](http://www.mediathèque-agglomeration-troyes.fr/bmtroyes)

24. [www.iabd.fr/spip.php?article54](http://www.iabd.fr/spip.php?article54)

25. [http://francenumerique2012.fr/pdf/081020\\_FRANCE\\_NUMERIQUE\\_2012.pdf](http://francenumerique2012.fr/pdf/081020_FRANCE_NUMERIQUE_2012.pdf)

26. [www.xitimonitor.com](http://www.xitimonitor.com)

Étude web XiTiMonitor, « Paramètres des moteurs », septembre 2008 ; Google domine le marché de la recherche en France avec 91,21 %, Yahoo 2,69 %, Live Search 2,16 %, AOL 1,43 %, Orange 1,24 %. ComScore publie les résultats des parts de marché des moteurs de recherche aux États-Unis, en juillet 2008 : Google 61,9 %, Yahoo 20,5 %, Microsoft 8,9 %. La part de Yahoo est loin d'être négligeable dans un contexte international.



La presse locale et régionale: un axe de travail prioritaire au niveau national et local. Montage: C. Soyez (BM de Lille)

tout pour des entreprises intéressées à ces aspects pour des motifs économiques. En ce qui concerne les bibliothèques, les financeurs sont aussi en droit d'exiger un retour sur investissement en termes d'audience. D'une façon générale, les bases de données culturelles comme les catalogues de bibliothèques, les inventaires d'archives, les photothèques font partie du web profond ou invisible (*deep web*), cette partie du web abritée par les formulaires de recherche que les moteurs ne peuvent atteindre<sup>27</sup>.

Le contenu d'une page de résultats généré en fonction des informations disponibles dans une base de données doit pouvoir compléter de façon dynamique les balises de l'entête HTML de la page et ainsi être visible des moteurs. Cela ne demande

souvent qu'une adaptation minime du site déjà développé. D'autres méthodes imposent des modifications d'architecture des systèmes. D'une façon générale, une attention particulière portée à ces aspects – optimisation sémantique et technique – en début de projet permet d'assurer une visibilité sur le web. Les prestataires informatiques traditionnels des bibliothèques sont souvent peu informés de ces domaines nouveaux et rechignent à modifier l'architecture technique de leur système pour quelques demandes émanant de bibliothécaires mieux informés et sensibilisés à ces questions, mais le sujet est critique pour rendre les données visibles et prêtes pour le web de demain. Il faut bien entendu mobiliser des budgets annuels – souvent raisonnables – et surtout écrire des procédures d'appels d'offres pour travailler avec des agences web spécialisées dans les questions de référencement.

## Intégrer les contenus

Le web constitue un changement de paradigme dans la diffusion de l'information. Tous les administrateurs de progiciels de bibliothèques savent bien que faire avancer et adapter les outils informatiques professionnels aux besoins du public et à l'évolution des technologies du web représente un travail ardu. Les demandes d'améliorations sont difficiles à obtenir et il faut souvent se conformer aux priorités du club utilisateur. Pendant ce temps, le web avance et invente à grande vitesse. Les outils traditionnels peinent à suivre. Des solutions moins figées, parfois issues du monde du libre, existent, mais cette souplesse n'empêche pas les chefs de projet de préciser leur demande. Une des tendances actuelles du web est l'ouverture des services en ligne vers l'extérieur par des connexions multiples transparentes pour les internautes. En général, cela

27. En avril 2008, Google a annoncé qu'il pourrait indexer à l'avenir certaines données du web invisible.

ne nécessite pas de développements lourds mais plutôt le paramétrage de passerelles liant des services (services web ou *web services*).

Les solutions d'interfaces de programmation ou API (Application Programming Interface) permettent de lire et de lier des documents exposés à l'extérieur. L'exemple le plus célèbre est sans doute celui des API proposés par Google : par exemple, pour insérer Google Maps dans un site ou pour faire un lien vers les ouvrages numérisés dans Google Book Search. Le catalogue WorldCat fournit une API XML de recherche permettant l'accès aux notices. Cette technique fonctionne dans les deux sens : des programmes peuvent ainsi être utilisés et/ou construits pour récupérer et/ou offrir de l'information. D'autres techniques basées sur la « recherche verticale » ou sur l'agrégation de contenus à la Netvibes permettent de rassembler les sources sur une même thématique dans une même page pour en avoir la meilleure vision possible à la manière de FoxyTunes dans le domaine de la musique.

## Faire participer le public

Dans les institutions culturelles, l'idée de permettre au public d'amender la description des documents n'a pas attendu la popularité ni la labellisation du web 2.0 pour être mise en œuvre. Depuis plusieurs années, les archives départementales des Yvelines<sup>28</sup> proposent aux lecteurs de leur salle de lecture de participer à la validation et à l'enrichissement de l'indexation des matricules militaires par l'annotation des actes. Une plateforme d'annotations collectives sur les actes d'état civil de 1850 à 1860 est proposée en ligne. Cette participation des internautes est le fruit d'une collaboration avec le laboratoire Irisa (Institut de recherche en informatique et systèmes aléatoires) de Rennes, qui apporte son savoir-faire dans les techniques de reconnaissance de

documents et de formes. Les archives municipales de Rennes<sup>29</sup> proposent une indexation collaborative des registres d'état civil. L'opération, commencée en juin 2005, a permis de relever 102 855 patronymes sur 75 registres de naissances de 1807 à 1880. Autres régions expérimentant les technologies du web 2.0 : la Picardie, qui propose depuis mars 2008 l'encyclopédie participative picarde Picardia<sup>30</sup> à la manière de la célèbre Wikipédia ou encore Wiki-Brest<sup>31</sup> sur le même modèle.

À la Bayerische Staatsbibliothek, on utilise le potentiel du monde virtuel Second Life pour présenter les incunables<sup>32</sup>. Dans la perspective salutaire d'aller là où sont les internautes, la bibliothèque municipale de Toulouse présente une collection de 500 images numérisées du début du xx<sup>e</sup> siècle<sup>33</sup> sur le site de partage de photographies en ligne de Yahoo Flickr sous la licence « Flickr Commons<sup>34</sup> » utilisée par la Bibliothèque du Congrès<sup>35</sup> en janvier 2008 pour le versement de photos d'actualité des années 1910-1930. Cette initiative française avait été précédée en février 2007 par le versement sur Flickr de 2763 photos de la bataille de Normandie (juin-août 1944) dans le cadre du projet PhotosNormandie.

Même si les interventions des internautes sur ces documents peuvent sembler anecdotiques, ces projets permettent de créer de véritables communautés d'utilisateurs. Cependant, pour que la mise en place des techniques issues de la mouvance du web 2.0 dé-

passé l'effet de mode et conserve son formidable impact en termes de diffusion de l'information, il convient de cibler son utilisation et de relativiser l'impact des médias participatifs sur la vie d'une base de données. La mise en place d'une interface de marquage sur un contenu très qualifié et au sujet déjà défini n'apportera sans doute rien à la qualité de la base documentaire diffusée. Permettre aux visiteurs extérieurs de qualifier une information ne veut pas dire que ceux-ci vont profiter réellement de cette opportunité, c'est même parfois l'inverse qui se produit. Sur Wikipédia, on estime que 80 % du contenu est fourni par seulement 10 % des utilisateurs de l'encyclopédie gratuite.

D'une façon générale et en particulier pour les métiers de la documentation, on obtient de meilleurs résultats en matière de collaboration en ligne en ciblant une communauté d'intérêt proche (généalogistes, historiens amateurs...) et en permettant à celle-ci de s'exprimer, qu'en essayant de cibler un public très large mais peu impliqué sur un sujet. Le projet PhotosNormandie a obtenu des résultats sur la base de ces technologies parce que les photographies mises en ligne relevaient d'un patrimoine local et relativement récent.

Sans utiliser des techniques très complexes, la simple mise en place d'une adresse mail ou d'un formulaire sur un site web permet aux internautes de réagir. Par ce moyen, la bibliothèque numérique de la bibliothèque municipale de Lille reçoit régulièrement des demandes de correction sur l'indexation de l'iconographie locale mise en ligne. Les internautes s'approprient très vite le patrimoine local mis en ligne. À la joie de toucher au cœur un public captif s'ajoute la satisfaction pour les professionnels, qui indexent souvent dans l'ombre, de voir enfin leur travail reconnu car accessible au-delà des frontières.

## Le web sémantique : les données sont de retour

Esquissé dès 1998 par l'inventeur du web, Tim Berners-Lee, le web sémantique est un concept étudié

28. [www.cg78.fr/archives/home.html](http://www.cg78.fr/archives/home.html)  
<http://img-dad.cg78.fr/archives/fonds-numerises/internet/pac-etat-civil.htm>

29. [www.archives.rennes.fr/pdf/presentation\\_index\\_ec.pdf](http://www.archives.rennes.fr/pdf/presentation_index_ec.pdf)

30. [www.encyclopedie.picardie.fr](http://www.encyclopedie.picardie.fr)  
Un exemple : la page multimédia « Roses de Picardie », [www.encyclopedie.picardie.fr/index.php/Roses\\_de\\_Picardie](http://www.encyclopedie.picardie.fr/index.php/Roses_de_Picardie)

31. [www.wiki-brest.net](http://www.wiki-brest.net)

32. Incunable digitization at Munich : from the Gutenberg Bible to mass digitization  
<http://slurl.com/secondlife/Insel%20der%20Information/132/23/22>

33. Ce projet concerne le fonds Trutat, une collection de 5 000 photographies sur plaques de verre sur les paysages et les habitants de Toulouse et de sa région.

34. [www.flickr.com/commons/usage](http://www.flickr.com/commons/usage)

35. [http://flickr.com/photos/library\\_of\\_congress](http://flickr.com/photos/library_of_congress)  
Sur le projet : [www.loc.gov/rr/print/flickr\\_pilot.html](http://www.loc.gov/rr/print/flickr_pilot.html)

depuis 2001 par le W3C<sup>36</sup> et qui se veut l'avenir du web. Il s'appuie sur une qualification avancée de l'ensemble des documents du web et sur une compréhension précise du sens des documents plutôt que des mots qu'ils utilisent. L'idée du web sémantique est finalement assez difficile à expliquer à l'écrit puisqu'elle veut dépasser les possibilités des mots. En web sémantique, une recherche sur « Queen » et « Mercury » éliminera d'office toutes les réponses relatives à la reine d'Angleterre ou à la planète Mercure, le sens même des deux mots recherchés en même temps ne pouvant renvoyer qu'au groupe de rock.

Travaillant sur le sens des mots, le web sémantique devrait permettre aux machines de travailler plus efficacement en traitant automatiquement les données exposées aux moteurs de recherche, le corollaire étant l'exposition massive de données sorties du web invisible, celui des bases de données. Manquant actuellement de résultats réels et concrets, les outils qui soutiennent le web sémantique sont encore suspects, ou mal compris. L'idée est de construire un immense graphe (Giant Global Graph) qui relierait par le sens l'ensemble des données présentes sur le web. Le standard indispensable à la construction du web sémantique est l'URI (Uniform Resource Identifier = identifiant uniforme de ressource) qui permet d'identifier d'une façon certaine et sans équivoque une ressource. Le modèle de base du web sémantique est RDF (Resource Description Framework), « cadre » théorique et formel pouvant englober différents jeux de métadonnées (par exemple Dublin Core, XMP...) et dans des implémentations différentes (le plus souvent la syntaxe XML).

Dans les prochaines années, il serait souhaitable que les bibliothèques puissent informatiser les fichiers papier plus spécialisés, biographiques par exemple. La conversion des données pourra profiter de la puissance

du langage XML par un encodage en Dublin Core XML ou par une utilisation de FOAF (Friend Of A Friend), vocabulaire RDF permettant de représenter des informations sur les groupes ou les personnes.

Les efforts faits vers la rigueur et la complétude des données produites, associés à leur exposition massive ne peuvent être vains<sup>37</sup>. Le web affirme la place du traitement de l'information au cœur du métier. Mettre « la bonne donnée au bon endroit » est la seule façon de faire fonctionner correctement les moteurs de recherche et de faire avancer les outils en synergie. Peu importe les logiciels et les systèmes de consultation. L'essentiel est dans la donnée.

## Place au numérique

La réussite des projets numériques dépend de compétences variées que seules quelques années d'expérience permettent de maîtriser : questions techniques, juridiques, scientifiques, administratives (cahiers des charges)... Leur fragilité tient souvent, dans l'ordre, au manque de moyens humains, techniques et financiers<sup>38</sup>. La gestion de l'information numérique reste souvent le parent pauvre des services d'un établissement et la rigueur imposée par l'informatique peut être considérée comme un obstacle au développement des services. On ne mesure pas suffisamment son impact sur les fonctions bibliothéconomiques traditionnelles dont elle relève également (acquisition, catalogage/métadonnées, médiation/interface, conservation). Les bibliothèques ont absolument besoin de développeurs informatiques.

Ces sujets, qui s'improvisent difficilement au cœur d'un projet, sont très rarement enseignés dans les for-

mations initiales. Leur maîtrise s'acquiert sur le terrain par le goût du web et l'observation précise et attentive de ce médium en perpétuel mouvement. Une culture commune se met progressivement en place par le biais des stages proposés dans le cadre de la formation continue, à l'Enssib (École nationale supérieure des sciences de l'information et des bibliothèques) ou dans les centres de formation régionaux.

Le problème essentiel en France est bien, actuellement, l'acquisition d'un savoir avancé et complexe nécessaire pour mettre en œuvre de manière correcte et complète ces nouvelles technologies dans l'objectif du signalement, de la communication et de la valorisation du patrimoine, pourtant missions de base des bibliothèques. Cette difficulté, trop souvent méconnue ou mal évaluée par les responsables, bloque les projets, ou empêche leur développement complet. Les projets numériques mis en œuvre restent souvent très loin des possibilités réelles de services offertes par l'informatique.

Ces problématiques doivent être portées par un discours stratégique et politique. Il est souhaitable que les établissements mettent en place une véritable « politique de valorisation et de diffusion » basée sur un service stratégique et substantiel composé de bibliothécaires et d'informaticiens aptes à mettre à profit les technologies du numérique, soucieux de respecter les aspects normatifs et convaincus de l'importance de la diffusion des contenus vers les publics. Le numérique est un outil qui permet de nouvelles formes de valorisation, de diffusion, d'accès et d'étude du patrimoine : un outil au service d'un patrimoine vivant. ●

Novembre 2008

36. L'article fondateur est : Tim Berners-Lee, James Hendler et Ora Lassila, « The semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities », *Scientific American*, mai 2001. [www.sciam.com](http://www.sciam.com)

37. Voir la journée « Web sémantique » organisée par Aula (Association des utilisateurs de logiciels Archimed) le 27 juin 2008. [www.aula.asso.fr/index.php?option=com\\_content&task=view&id=56&Itemid=1](http://www.aula.asso.fr/index.php?option=com_content&task=view&id=56&Itemid=1)

38. Comme le montre l'étude de Karine Delvert dans le cadre de son PPP/enquête DLL « Numérique et mise en ligne de documents numérisés : enjeux, pratiques et besoins des bibliothèques municipales. Méthodologie et premiers éléments d'enquête », 2008.