

# Genèse et description des bibliothèques numériques

## The Origins and Description of Digital Libraries

### Origen y descripción de las bibliotecas digitales

Lyne Da Sylva

Bibliothèques numériques

Volume 59, numéro 3, juillet–septembre 2013

URI : <https://id.erudit.org/iderudit/1018843ar>

DOI : <https://doi.org/10.7202/1018843ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (imprimé)

2291-8949 (numérique)

[Découvrir la revue](#)

Citer cet article

Da Sylva, L. (2013). Genèse et description des bibliothèques numériques. *Documentation et bibliothèques*, 59 (3), 132-145. <https://doi.org/10.7202/1018843ar>

Résumé de l'article

Les bibliothèques numériques sont des collections constituées de ressources numériques et des fonctionnalités donnant accès à ces dernières. Cet article décrit d'abord la naissance des bibliothèques numériques, issues de l'informatique mais puisant maintenant aux traditions de plusieurs disciplines et institutions, qui ont chacune un point de vue différent sur celles-ci. Un panorama des bibliothèques numériques existantes est présenté, avec les particularités de chacune. Des enjeux multiples sont soulevés par le développement des bibliothèques numériques : enjeux économiques, technologiques, juridiques, etc. Ces enjeux sont détaillés dans le texte, comme autant de pistes de recherche qui s'ouvrent à la communauté. L'article se termine sur une note prédictive : il évoque les développements prévisibles, selon les perspectives des disciplines et institutions touchées.

# Genèse et description des bibliothèques numériques

LYNE DA SYLVA

Professeure agrégée  
École de bibliothéconomie et des sciences de l'information  
Université de Montréal  
lyne.da.sylva@umontreal.ca

## RÉSUMÉ | ABSTRACTS | RESUME

*Les bibliothèques numériques sont des collections constituées de ressources numériques et des fonctionnalités donnant accès à ces dernières. Cet article décrit d'abord la naissance des bibliothèques numériques, issues de l'informatique mais puisant maintenant aux traditions de plusieurs disciplines et institutions, qui ont chacune un point de vue différent sur celles-ci. Un panorama des bibliothèques numériques existantes est présenté, avec les particularités de chacune. Des enjeux multiples sont soulevés par le développement des bibliothèques numériques : enjeux économiques, technologiques, juridiques, etc. Ces enjeux sont détaillés dans le texte, comme autant de pistes de recherche qui s'ouvrent à la communauté. L'article se termine sur une note prédictive : il évoque les développements prévisibles, selon les perspectives des disciplines et institutions touchées.*

### *The Origins and Description of Digital Libraries*

*Digital libraries are made up of electronic resources and include the functions enabling access to these resources. This article describes the origins of digital libraries from their early beginnings in computer science, drawing upon the traditions of several disciplines and institutions, each with a different point of view. Several types of digital libraries, as well as their characteristics, are described. The economic, technological and legal challenges of digital libraries are identified and the possible research venues for the community are raised. The article predicts the possible future developments for the disciplines and institutions involved.*

### *Origen y descripción de las bibliotecas digitales*

*Las bibliotecas digitales son colecciones formadas por recursos electrónicos y funcionalidades que facilitan su acceso. En primer lugar, este artículo describe el origen de las bibliotecas digitales, que surgieron de la informática, pero que actualmente se basan en las tradiciones de numerosas disciplinas e instituciones, cada una con un punto de vista diferente. Asimismo, se ofrece un panorama de las bibliotecas digitales actuales, con las especificidades de cada una. A partir del desarrollo de las bibliotecas digitales, se han presentado diversos desafíos : económicos, tecnológicos, jurídicos, etc. Estos desafíos se explican detalladamente en este artículo, al igual que los criterios de búsqueda que se ofrecen a la comunidad. El artículo finaliza con una nota predictiva, que alude a los desarrollos previsibles, en función de las perspectivas disciplinarias e institucionales afectadas.*

## Introduction

LES BIBLIOTHÈQUES NUMÉRIQUES REPRÉSENTENT un secteur d'activités et de recherche actif depuis une quinzaine d'années. Né de l'informatique, il est aujourd'hui défini davantage selon des critères issus des sciences de l'information : il vise les systèmes informatiques contenant des collections de ressources numériques (le plus souvent textuelles, comme des livres et articles, mais parfois aussi des images, des vidéos, des composantes logicielles, etc.) ainsi que les services assurés aux utilisateurs sur la base de politiques définies. Si le volet technologique a prédominé dans les débuts (tout récents) des travaux de recherche et de développement, on reconnaît aujourd'hui que ce domaine d'étude est véritablement multidisciplinaire et interpelle autant les chercheurs en sciences de l'information (bibliothéconomie et archivistique), qu'en muséologie, en informatique, en traitement automatique de la langue (pour les traitements automatiques du contenu textuel), en droit (notamment sur la question des droits d'auteur relatifs à la diffusion des contenus), et d'autres disciplines encore ; à ceux-ci s'ajoutent tous les champs d'application, les domaines d'activité qui bénéficient de la mise sur pied d'une bibliothèque numérique spécialisée.

L'univers des bibliothèques numériques est décrit globalement dans Arms (2000) et dans Papy (2005). Différents colloques leur sont consacrés depuis une dizaine d'années : *Joint Conference on Digital Libraries (JCDL)*, *European Conference on Digital Libraries (ECDL)*, maintenant *Theory and Practice of Digital Libraries*, ou *TPDL*, *International Conference on Asian Digital Libraries (ICADL)*, *International Conference for Digital Libraries and the Semantic Web (ICSD)*.

Les bibliothèques numériques présentent des aspects technologiques et sociaux intimement liés : la technologie vise à faciliter le repérage de sources d'information et l'accès à celles-ci dans le but de mieux servir les utilisateurs.

Le présent article brosse un tableau des bibliothèques numériques en trois parties. D'abord, nous décrivons la naissance des bibliothèques numériques dans le monde en général et dans la francophonie en particulier. Ensuite, nous ferons état des projets et progrès réalisés jusqu'à présent, tout en présentant l'éventail très diversifié des collections comprises dans les bibliothèques numériques ; nous énumérerons au

passage les enjeux qui façonnent le développement des bibliothèques numériques. En guise de conclusion, nous établirons ce que sera vraisemblablement leur avenir, du moins dans un horizon rapproché.

## Naissance des bibliothèques numériques

Pour décrire la genèse des bibliothèques numériques (dorénavant BN), nous présenterons d'abord un bref historique de celles-ci, détaillant également les facteurs qui ont permis leur croissance. Après ce tour d'horizon, la présentation de quelques définitions permettra de bien circonscrire le domaine.

### Historique

L'histoire des BN est très récente. Le bref historique détaillé dans ce qui suit servira à identifier les apports de diverses disciplines au domaine et nous amènera à définir ce que l'on entend aujourd'hui par ce terme.

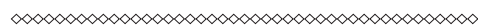
#### Débuts

Les premiers jalons de l'histoire des BN appartiennent à l'ère pré-WWW. En effet, on a vu se créer, dès les années 1970, des collections de ressources numériques qui peuvent se prétendre les précurseurs des BN actuelles.

En 1971, le projet Gutenberg voit le jour. Il s'agit d'une collection d'œuvres numérisées, essentiellement des livres. En 1985, le projet Perseus est fondé par Gregory Crane, professeur d'informatique formé en études classiques, sur le thème de la Grèce antique. Il cherche non seulement à donner accès à ses étudiants à des textes classiques, mais également à mettre ces textes en relation avec d'autres documents de l'époque, des cartes géographiques, des photos de sculptures et autres œuvres d'art, etc. C'est également l'année de la création de NetLib, répertoire de logiciels mis sur pied par des informaticiens.

Ces collections sont consultées par un public restreint, adeptes avertis des technologies de partage de fichiers en réseau (i.e. Internet avant le Web). Il est important de noter que plusieurs autres collections de textes sont présentes sur le réseau Internet ou sur des réseaux intranet privés avant le début des années 1990. Mentionnons les archives privées de sociétés commerciales contenant de la documentation technique (pour l'aviation, pour le secteur des appareils électriques et électroniques, etc.), des articles scientifiques (IBM, compagnies du secteur médical, etc.) et les corpus de textes créés pour la recherche en linguistique-informatique (par exemple, le corpus Brown 1965 — Kučera et Francis 1967). Une collection particulière qui influencera de manière importante l'évolution de la recherche en physique aussi bien que le développement des BN est

*Les premiers jalons de l'histoire des BN appartiennent à l'ère pré-WWW. En effet, on a vu se créer, dès les années 1970, des collections de ressources numériques qui peuvent se prétendre les précurseurs des BN actuelles.*



l'archive de prépublications du Los Alamos National Laboratory<sup>1</sup>, nommé à ses débuts dès 1991 « *Los Alamos E-print Archive* » et qui deviendra par la suite simplement *arXiv* (écrit avec la lettre grecque khi et donc prononcé comme « archive » en anglais).

Les premiers efforts des bibliothèques ont été déployés par la Bibliothèque du Congrès américain et par quelques bibliothèques universitaires. Les objectifs de ces systèmes : permettre une meilleure circulation de l'information qu'auparavant (Arms 2000, 4).

Le prochain jalon important est sans contre-dit la création du World Wide Web (WWW) en 1991 et sa démocratisation subséquente qui s'étalera sur une dizaine d'années. Les technologies nécessaires au déploiement de BN (dont nous parlerons à la section suivante) étaient suffisamment mûres pour permettre leur éclosion.

#### Naissance véritable

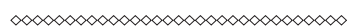
C'est au tournant des années 1990 que le terme « *digital library* » apparaît. Certains l'attribuent à la Corporation for National Research Initiatives en 1988 (Kahn et Cerf 1988), d'autres à l'initiative NSF/DARPA/NASA pour les bibliothèques numériques en 1993 (Fox 1999). Une chose est certaine : l'histoire des BN ne date que d'une quinzaine d'années. En 1994, aux États-Unis, se tient le premier congrès sur les BN, le *First Annual Conference on the Theory and Practice of Digital Libraries*<sup>2</sup>. Plusieurs prototypes et systèmes ont émergé depuis. La plate-forme Érudit, qui diffuse au départ des thèses de doctorat de l'Université de Montréal et qui se consacre rapidement aux revues scientifiques, est introduite en 1998 (voir l'article de Paquin dans ce numéro thématique). Les débuts de la collection « Les Classiques des sciences sociales » (présentée ailleurs dans ce numéro) remontent à 1999. En 2000, Douglas Arms publie la première monographie sur les BN, intitulée simplement *Digital Libraries*.

On constate une explosion du phénomène lorsque Google lance, en 2004, sa collection *Google Books*, dans la controverse sociale, scientifique et judiciaire (voir notamment Jacquesson 2010). Dès lors, les institutions

1. <<http://www.lanl.gov/index.php>>.

2. <<http://www.csdl.tamu.edu/DL94/>>.

## *Plusieurs technologies textuelles ont permis de créer et de rendre disponibles les ressources textuelles attrayantes et efficaces.*



publiques ne peuvent ignorer le phénomène et doivent se positionner et agir quant à la diffusion numérique de leurs collections. Gallica, la BN de la Bibliothèque nationale de France (<<http://gallica.bnf.fr>>), voit le jour en 1997 et la bibliothèque européenne Europeana (<<http://europeana.org/portal>>) est lancée en 2008.

En mai 2012, le colloque scientifique sur la BN à l'ACFAS (Association canadienne-française pour l'avancement des sciences) a enclenché l'essor des travaux sur les BN au Québec et dans la francophonie nord-américaine.

### **Facteurs de croissance**

On peut identifier quatre facteurs technologiques importants qui ont favorisé le développement des BN.

- i) L'évolution dans les interfaces : aux débuts de l'informatique, des interfaces purement textuelles offraient des fonctionnalités d'interactions très limitées et peu conviviales. Les systèmes de menus à options numérotées offraient une alternative mnémotechnique et plus directe pour lancer des commandes. Ces systèmes étaient la norme lorsqu'on a vu le développement d'interfaces de recherche telles les gophers (voir la norme Internet RFC1436, <<http://tools.ietf.org/html/rfc1436>>) et Archie (Li 2002). Mais il a fallu attendre le mode graphique et la présentation des fonctionnalités interactives en deux dimensions (plutôt que l'unidimensionnalité de la ligne de texte), puis en trois dimensions (l'hypertexte), pour que les utilisateurs envisagent ce mode d'interaction plutôt que la consultation directe (physique) de ressources textuelles comme des articles ou des monographies.
- ii) La miniaturisation et l'augmentation de la performance de la micro-informatique : elles ont permis l'appropriation de la technologie par le grand public, et non seulement par les experts ou les scientifiques. L'ajout de la souris comme périphérique d'interaction se superposant au clavier a augmenté d'autant l'aisance de manipulation et de navigation dans l'interface.
- iii) La sophistication des technologies de réseau : les BN sont d'autant plus consultées qu'elles sont accessibles, bien sûr. Les premières configurations réseau, locales (LAN, *local area network*), reliaient des ordinateurs physi-

quement rapprochés les uns des autres. Les réseaux WAN (*wide-area network*) ont étendu ce champ d'action et ont permis des configurations de plus en plus complexes, mais c'est avec la création du réseau mondial, propulsé par le WWW et ses technologies hypertextuelles, que l'on a vu un saut qualitatif dans le déploiement des ressources.

iv) La prolifération de technologies textuelles : plusieurs technologies textuelles ont permis de créer et de rendre disponibles les ressources textuelles attrayantes et efficaces :

- l'extension des jeux de caractères au-delà de l'ASCII standard qui ne permettait guère que l'encodage des caractères de l'anglais (avec des limites importantes même pour cette langue) ;
- les divers logiciels ont permis le traitement et la création des textes (bureautique, éditique) ;
- les langages de balisage (SGML, XML, HTML) et leurs « grammaires de formes de documents », les DTD (*Document Type Definition*) dont on peut citer trois célèbres exemples, la TEI (*Text Encoding Initiative*, pour l'encodage de textes de divers genres littéraires et scientifiques ; Burnard et Bauman 2007), DocBook (pour la documentation technique ; Walsh 2010) et EAD (*Encoded Archival Description*, pour la description des outils de recherche archivistiques ; Library of Congress 2006) ;
- le format PDF qui, de surcroît, a permis de restituer l'apparence de documents numérisés, ce qui est utile dans le cas où ce n'est pas seulement le contenu informatif d'un document que l'on veut rendre disponible en version numérique, mais aussi son apparence originale ;
- les technologies d'indexation automatique et de compression de fichiers qui ont donné l'opportunité de développer des index performants pour des collections de très grande taille et ainsi permis le repérage et la consultation dans des temps réalistes (voir notamment Witten et al. 1999) ;
- les technologies de numérisation et de reconnaissance optique des caractères qui ont autorisé un déploiement plus facile et plus rapide des collections imprimées existantes ;
- enfin, la technologie hypertextuelle dont les avantages ont dépassé le seul champ textuel puisqu'elle donne accès à des ressources numériques sous d'autres formats.

À ces facteurs technologiques de croissance, on peut ajouter certains facteurs sociaux qui ont déterminé la manière dont les BN se sont développées et, surtout, la manière dont les utilisateurs se les sont appropriés :

- un accès direct aux collections, sans intermédiaire, au moyen des technologies de repérage de l'information (*information retrieval*) ;
- la création de communautés virtuelles, géographiquement dispersées, constituées de chercheurs et scientifiques, de communautés de pratique ou de groupes centrés autour d'un intérêt commun ;
- le multilinguisme croissant sur le Web, exigeant le développement des technologies en langues multiples et non seulement en anglais.

Tout ceci a favorisé l'autonomie croissante des utilisateurs ainsi que l'apparition d'exigences également croissantes de leur part quant aux possibilités d'interaction avec les systèmes.

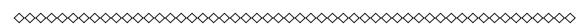
## Définitions

La définition la plus répandue de ce qu'est une BN est celle de Borgman (2000), présentée déjà dans l'introduction de ce numéro thématique : « *Un ensemble de ressources numériques et technologies associées pour créer, repérer et utiliser l'information ; les bibliothèques incluent des données, métadonnées et liens (métadonnées) vers d'autres ressources.* » (42, notre traduction). Il est utile d'insister sur le fait que les BN ne sont pas nécessairement des collections bibliothéconomiques, mais qu'elles peuvent inclure des ressources de types très différents. Ceci représentera peut-être, ultimement, un handicap pour ce terme, qui semble exclure les collections archivistiques et muséales, par exemple. Or il n'en est rien. Les besoins de description et les systèmes existants puisent véritablement tout à la fois dans les traditions bibliothéconomiques, archivistiques et muséales.

Les BN ont la particularité de conjuguer des aspects technologiques (l'infrastructure informatique) et institutionnels. Borgman (2000, 52) les décrit comme une extension, une mise en valeur et une intégration de systèmes de repérage de l'information et de plusieurs institutions liées à l'information, incluant les bibliothèques mais d'autres également.

Il est important de souligner aussi que ce sont les fonctionnalités développées pour exploiter les ressources qui définissent les BN, ces dernières ne pouvant se résumer à de simples assemblages de ressources variées. Et que les BN sont constituées de collections rassemblées de manière intentionnelle et suivant des politiques ou une mission définies. En ce sens, le Web dans sa globalité, à cause de l'absence de concertation pour l'assemblage des ressources, ne peut être considéré comme une véritable BN.

*Borgman décrit [les BN] comme une extension, une mise en valeur et une intégration de systèmes de repérage de l'information et de plusieurs institutions liées à l'information, incluant les bibliothèques mais d'autres également.*



Chacune de ces caractéristiques revêt une grande importance. En fait, des modèles théoriques ont été développés pour définir la portée exacte du terme « bibliothèque numérique ». Ils servent aussi à baliser les efforts d'implémentation. Les deux modèles principaux sont le *5S Framework* : *Streams, Structures, Space, Scenarios, Societies* (Gonçalves *et al.* 2004) et le modèle DELOS (Candela *et al.* 2008) en trois strates.

## Institutions

Les institutions qui ont façonné le développement des bibliothèques numériques (les archives, les bibliothèques et les musées) ont chacune leurs caractéristiques particulières. On peut brosser un tableau rapide de celles-ci pour les distinguer et les opposer.

Les archives sont créées dans le cadre des activités d'un organisme, soumises à un « cycle de vie » et doivent souvent répondre à des exigences légales. Leurs utilisateurs sont, selon l'étape du cycle de vie, les créateurs ou les destinataires visés (archives actives ou semi-actives), ou bien les chercheurs et historiens (archives définitives). Parmi les problématiques particulières aux archives, on peut noter l'importance déterminante du contexte de création des archives (pour leur interprétation) et les effets du calendrier de conservation sur leur durée de vie. Sur le plan de la description de ces documents, ils ont la particularité d'être décrits souvent au niveau du dossier et non du document. Un autre défi est la nécessité de se soumettre à la loi sur la protection des renseignements personnels et aux contrats de donation.

Les bibliothèques sont habituellement créées pour diffuser de l'information, des connaissances, des œuvres, auprès d'une communauté d'utilisateurs ciblés ; leur profil est défini selon la mission du service d'information. La problématique de description relève d'une approche bibliothéconomique traditionnelle. Dans le monde numérique, la diffusion des œuvres devient très facile, même trop, enfreignant souvent les lois liées au droit d'auteur.

Enfin, les musées sont dépositaires d'objets de types variés dans un but d'éducation et de loisir : « *Un musée est une institution permanente sans but lucratif au service de la société et de son développement, ouverte au public, qui acquiert, conserve, étudie, expose et transmet le patrimoine matériel et immatériel de l'humanité et de*



**Tableau 1**  
Différentes visions de la bibliothèque numérique

SELON CE POINT DE VUE...	LA BIBLIOTHÈQUE NUMÉRIQUE REPRÉSENTE...
La bibliothéconomie	une institution et non une technologie une extension logique des bibliothèques traditionnelles ayant pour but d'acquérir, d'organiser, de disséminer l'information
L'archivistique ou la muséologie	de nouvelles manières de préserver le patrimoine
La technologie	une extension des technologies de réseau, de bases de données, de systèmes de recherche d'information
Le droit	un nouveau domaine d'application de lois existantes, notamment le droit d'auteur
L'aspect social	le théâtre de nouvelles utilisations, de nouveaux comportements
L'économie	de nouvelles occasions d'affaires de nouveaux modèles économiques à développer
La communication	un nouveau médium de diffusion et de médiation de l'information
La linguistique	de nouvelles formes de discours des possibilités de traitement automatique de la langue
La culture	de nouvelles manifestations de contenu culturel
L'éducation	de nouvelles ressources pédagogiques

son environnement à des fins d'études, d'éducation et de délectation. » (Desvallées et Mairesse 2011, 271). Leurs utilisateurs peuvent avoir des profils plus flous que ceux des bibliothèques, si l'on considère que les musées sont souvent des lieux touristiques qui peuvent accueillir des gens de passage. Parmi les problématiques particulières à la description des œuvres de musée, notons les dimensions spatiales, expérientielles et parfois dynamiques et interactives, étrangères aux deux autres types d'institution (sans même aborder la question délicate du patrimoine immatériel).

De manière simpliste, et caricaturale peut-être, on peut distinguer les défis particuliers auxquels sont confrontées les trois institutions de base liées aux BN :

- Archives numériques : assurer la pérennité ;
- Bibliothèques numériques : assurer la diffusion dans le respect du droit d'auteur ;
- Musées virtuels : transmettre l'aspect spatial, interactif et expérientiel.

Mais dans les faits, tous ces défis se présentent (ou peuvent se présenter) pour toute BN, peu importe son type.

### Visions de la BN

Divers acteurs œuvrent dans l'univers des BN : au départ, des informaticiens, des littéraires, des linguistes et des linguistes-informaticiens ont jeté les premières bases. Les bibliothécaires se sont joints plus tard aux efforts, avec les archivistes, les muséologues et les muséographes. Puis, des utilisateurs de tout acabit s'y sont intéressés : entreprises commerciales, gouvernements, éducateurs, éditeurs... Par conséquent, les bibliothèques numériques peuvent être envisagées de différents points de vue, chacun ayant sa propre vision.

Ce panorama met en lumière le caractère véritablement multidisciplinaire des BN.

## Réalisations

Il existe aujourd'hui des centaines de BN. Certaines sont liées à des bibliothèques publiques (la Chicago Public Library Digital Collection<sup>3</sup>) ou universitaires (comme la BN Calypso, de l'Université de Montréal<sup>4</sup>), alors que d'autres (en fait la plupart) sont indépendantes de toute bibliothèque physique. Certaines ont une mission très ciblée, d'autres se donnent une vocation plus large, comme la (somme toute modeste) Bibliothèque numérique mondiale, qui « met à disposition sur Internet, gratuitement et en plusieurs langues, une documentation considérable en provenance des pays et des cultures du monde entier » (<<http://www.wdl.org/fr>>). Elle contient près de 8 000 documents : du texte (dont des manuscrits), des images (dont des cartes et photos), des enregistrements sonores, des films. Le répertoire de Martin et Pécout (2011), recensait déjà 185 BN importantes et il y a toutes les raisons de croire que le nombre a augmenté de manière importante depuis la publication de cet ouvrage.

Pour apprécier la diversité des BN, il convient de faire l'inventaire des types de collections qui y sont hébergées.

3. <<http://digital.chipublib.org/cdm4/>>.

4. <<http://calypso.bib.umontreal.ca/>>.

**Tableau 2**  
Collections archivistiques

Archives civiles et judiciaires de BANQ <sup>a,b</sup>	Archives des notaires du Québec, des origines à 1931, ainsi que les registres de l'État civil du Québec, des origines à 1911
Canadiana.org <sup>c</sup>	La BN recense une collection numérique d'archives canadiennes (contient également des collections numériques bibliothéconomiques et muséales).
The National Archives DocumentsOnline <sup>d</sup>	Collection archivistique britannique
Expositions virtuelles, service d'archives de l'Université de Montréal <sup>e</sup>	Exemple représentant diverses initiatives à échelle variable

a. <[http://www.banq.qc.ca/collections/collection\\_numerique/index.html?categorie=2](http://www.banq.qc.ca/collections/collection_numerique/index.html?categorie=2)>.

b. D'autres documents d'archives sont également disponibles sur le site de BANQ.

c. <<http://www.canadiana.ca/fr/accueil>>.

d. <<http://www.nationalarchives.gov.uk/documentsonline/default.asp>>.

e. <<http://www.archiv.umontreal.ca/exposition/exposition.html>>.

**Tableau 3**  
Collections bibliothéconomiques

ARTICLES DE PÉRIODIQUES	
Érudit <sup>a</sup>	Revue savantes et culturelles, etc. (Québec)
Persée <sup>b</sup>	Revue en sciences humaines et sociales (France)
Repère <sup>c</sup>	Périodiques de langue française
arXiv <sup>d</sup>	Prépublications en physique et sciences connexes
LIVRES	
Les Classiques des sciences sociales <sup>e</sup>	Ouvrages en français en sciences sociales et humaines
Gallica <sup>f</sup>	Partie des collections numérisées de la Bibliothèque nationale de France
Google Books <sup>g</sup>	Livres numérisés disséminés par Google
Le projet Gutenberg <sup>h</sup>	Livres numérisés, du domaine public
HathiTrust <sup>i</sup>	Livres numérisés de bibliothèques universitaires

a. <<http://www.erudit.org>>.

b. <<http://www.persee.fr/web/guest/home>>.

c. <<http://reperes.sdm.qc.ca>>.

d. <<http://arxiv.org>>.

e. <<http://classiques.uqac.ca>>.

f. <<http://gallica.bnf.fr>>.

g. <<http://books.google.com>>.

h. <<http://www.gutenberg.org>>.

i. <[http://www.hathitrust.org/digital\\_library](http://www.hathitrust.org/digital_library)>.

## Types de collections et exemples

Toute présentation de ce type ne sera qu'éphémère, les réalisations futures rendant la liste qui suit obsolète. Elle sera tout de même utile pour caractériser les différents types de BN que l'on a vu émerger ces dernières années. Pour chaque type de collection, une série d'exemples de BN existantes est présentée.

Dans les sections qui suivent seront identifiées également des métadonnées (ou clés de description) typiques pour chaque type de collection, au-delà des éléments universels que sont le titre, l'auteur (ou créateur), la date de création et la description (par mots-clés, descripteurs ou résumé). Un cas particulier est présenté par les droits de diffusion, qui sont pertinents pour toutes les ressources répertoriées dans une BN, mais dont le respect est un sujet contentieux (voir une discussion dans Maurel, 2008). L'ouvrage de Jacquesson et Rivier (2005) explore plus en détails les caractéristiques des documents et collections numériques, et pourra être une lecture additionnelle utile.

## *Collections archivistiques*

Les collections archivistiques sont montées en suivant les principes, méthodes et outils définis par la discipline. Elles peuvent être soumises à des réglementations locales ou nationales. On peut relever, comme propriétés particulières de ces collections, deux métadonnées typiques pour leurs documents : le genre ou type (procès-verbal, rapport, etc.), qui déterminera leur traitement dans le temps, et le contexte de création. Quelques exemples sont présentés dans le tableau 2.

## *Collections bibliothéconomiques*

Les collections bibliothéconomiques représentent la forme prototypique des BN, qu'il s'agisse d'articles de périodiques ou journaux, ou encore de livres. Le tableau 3 présente des exemples de collections d'articles et de livres. Les métadonnées utiles pour les décrire sont celles développées par la tradition bibliothéconomique.

**Tableau 4**  
Collections muséales

Europeana <sup>a</sup>	Cœuvres numérisées (livres, tableaux, films, objets muséaux et document d'archives) de tous les pays d'Europe
Musée virtuel du Canada <sup>b</sup>	Guichet unique en ligne sur les musées canadiens et leurs collections, donnant accès notamment à différentes expositions virtuelles
Musée virtuel de la Nouvelle-France <sup>c</sup>	Visant à dresser un portrait fidèle de la Nouvelle-France dans sa totalité (Acadie, Louisiane, Canada, Pays d'en haut) en présentant un large éventail des facettes de l'expérience coloniale française
State Hermitage Museum <sup>d</sup>	Collection numérique qui donne un aperçu du contenu du célèbre musée physique du même nom

a. <<http://europeana.org/portal>>.

b. <<http://www.museevirtuel-virtualmuseum.ca/index-fra.jsp>>.

c. <<http://www.civilisations.ca/musee-virtuel-de-la-nouvelle-france/introduction/>>.

d. <[http://www.hermitagemuseum.org/html\\_En/index.html](http://www.hermitagemuseum.org/html_En/index.html)>.

**Tableau 5**  
Dépôts institutionnels

Papyrus <sup>a</sup>	Université de Montréal
PolyPublie <sup>b</sup>	École Polytechnique de Montréal
Archipel <sup>c</sup>	Université du Québec à Montréal
Bibliothèque numérique du CRDI <sup>d</sup>	Dépôt du Centre de recherches pour le développement international
Archives des publications du CNRC <sup>e</sup>	Dépôt du Centre national de recherche du Canada

a. <<http://papyrus.bib.umontreal.ca/jspui/>>.

b. <<http://publications.polymtl.ca/>>.

c. <<http://www.archipel.uqam.ca>>.

d. <<http://idl-bnc.idrc.ca/dspace/>>.

e. <<http://nparc.cisti-icist.nrc-cnrc.gc.ca/npsi/ctrl?lang=fr>>.

### *Collections muséales*

Les collections muséales ont, du point de vue social ou culturel, l'objectif de faire rayonner le patrimoine du musée en dehors de ses murs. Du point de vue technologique, force est de constater que les ressources numériques sont souvent un pauvre substitut à l'objet original : elles peuvent se réduire à des photographies d'œuvres bi- ou tridimensionnelles. Par contre, l'interactivité permise par les environnements informatiques a rendu possible la création d'expositions virtuelles dynamiques, notamment par le Musée virtuel du Canada ; le visiteur évolue dans un univers en ligne où peut se dérouler un scénario mettant en scène les objets muséaux (voir des exemples au tableau 4). Des métadonnées typiques pour les objets muséaux sont le type d'œuvre, le format (dimensions), les matériaux des objets, la propriété, les modes d'accès, l'état des objets, etc.

### *Dépôts institutionnels*

Les dépôts institutionnels sont des sites développés et maintenus par des institutions de recherche, qui rassemblent diverses publications (textes, données brutes, matériel d'enseignement, rapports, etc.) d'auteurs affiliés à l'institution. Les publications y sont diffusées librement, et représentent souvent une copie addition-

nelle d'un document publié sous une forme différente par un éditeur commercial. Le tableau 5 en donne un aperçu au niveau québécois et canadien. Une métadonnée utile pour ces publications « dupliquées » pourra être un lien vers des publications externes.

### *Collections d'images*

Les images peuvent regrouper des genres variés (photo, diapositive, peinture, dessin, graphique, affiche, plan, carte, menu, programme, image d'un original textuel, etc.). Notons que les cartes et les plans combinent l'image et le texte et peuvent poser, de ce fait, des défis particuliers au traitement et à la description. Pour décrire ces ressources, des métadonnées utiles seront les objets ou personnes représentées, le format de l'original, le format de l'image numérique, le genre ; pour les peintures, on pourra indiquer le courant ou la période historique et, plus spécifiquement, pour l'art abstrait, les couleurs, formes et textures de l'œuvre ; etc. Le tableau 6 présente un petit nombre d'exemples de ces collections, qui se retrouvent très nombreuses sur le Web.

### *Collections audiovisuelles*

Les collections audiovisuelles contiennent des films ou des vidéos. Il peut s'agir de productions professionnelles artistiques ou journalistiques, d'enregistrements



**Tableau 6**  
Collections d'images

Collection numérique de cartes et plans, BAnQ <sup>a</sup>	Documents cartographiques du Québec à différentes périodes
Images, BAnQ <sup>b</sup>	Une partie des collections iconographiques numériques de BAnQ
Images dans les fonds d'archives de BAC <sup>c</sup>	Cartes, dessins, photos, timbres-poste...
Getty Images <sup>d</sup>	Collection d'images protégées
Flickr <sup>e</sup>	Photos

a. <<http://services.banq.qc.ca/sdx/cep/accueil.xsp>>.

b. <<http://www.banq.qc.ca/collections/images/>>.

c. <<http://www.collectionscanada.gc.ca/lac-bac/recherche/images>>.

d. <<http://www.gettyimages.ca/>>.

e. <<http://www.flickr.com>>.

**Tableau 7**  
Collections audiovisuelles (vidéo)

Collection de l'INA <sup>a</sup>	Archive des productions audiovisuelles françaises
ONF <sup>b</sup>	Collection de films de l'Office national du film canadien
YouTube <sup>c</sup>	Vidéos diffusées librement
Canal U <sup>d</sup>	Vidéotheque numérique de l'enseignement supérieur (France)
Documents audiovisuels de la Bibliothèque nationale suisse <sup>e</sup>	Documents audio, cinématographiques et vidéo d'auteurs suisses des quatre langues nationales (Archives littéraires suisses à Berne, Archives Max Frisch à Zurich, archives des chaînes de télévision et de radio)

a. <<http://www.ina.fr/>>.

b. <<http://www.onf.ca/explorer-tous-les-films/>>.

c. <<http://www.youtube.com>>.

d. <<http://www.canal-u.education.fr/>>.

e. <<http://www.nb.admin.ch/sla/03114/03557/index.html?lang=fr>>.

**Tableau 8**  
Collections musicales

Collection d'enregistrements sonores de BAnQ <sup>a</sup>	Disques 78 tours de musique (1900-1952)
Le Gramophone virtuel de BAC <sup>b</sup>	Divers documents, dont des enregistrements sonores, donnant un aperçu de l'époque des 78 tours au Canada
Last.fm <sup>c</sup>	Webradio et site Internet proposant des recommandations de musique
Memory Project (Library of Congress) <sup>d,e</sup>	Enregistrements musicaux (sous Performing Arts, Music)

a. <[http://bibnum2.banq.qc.ca/bna/musique\\_78trs/accueil.htm](http://bibnum2.banq.qc.ca/bna/musique_78trs/accueil.htm)>.

b. <<http://www.collectionscanada.gc.ca/gramophone/index-f.html>>.

c. <<http://www.last.fm/>>.

d. <<http://memory.loc.gov/ammem/browse/ListSome.php>>.

e. Ce projet inclut également des collections de musique en feuille (partitions) et d'autres types de documents tels que photos, correspondance, etc.

créés dans des contextes de travail, ou bien de clips personnels tournés par des amateurs (voir le tableau 7). Les métadonnées typiques incluent le format d'encodage, la durée, le genre (film, nouvelles télévisées, entrevue, clip, enregistrement de réunion, formation, témoignage personnel, etc.).

### *Collections musicales*

L'histoire juridique des 20 dernières années a mis en vedette, jusqu'à un certain point, les collections de ressources musicales. Sans entrer dans ce débat juridique (encore une fois lié au droit d'auteur), on peut examiner les caractéristiques des collections musicales. Nous visons ici les collections d'enregistrements sonores

musicaux, réalisés dans divers environnements naturels ou de production commerciale, dont on voit des représentants dans le tableau 8<sup>5</sup>. Pour décrire les ressources musicales, on peut utiliser des métadonnées portant sur le compositeur de la musique, le(s) interprète(s) et chef d'orchestre (si pertinent), le genre (musique instrumentale, chant, gospel, bluegrass, cajun, nouvel âge, country, jazz, blues, hip hop, folklore, etc.) ou la forme musicale (opéra, opérette, sonate, fugue...).

5. On pourrait aussi considérer les BN de musique sous forme symbolique (partitions musicales), comme par exemple la collection numérique de livres et partitions musicales de BAnQ. (<[http://www.banq.qc.ca/collections/collection\\_numerique/index.html?categorie=8](http://www.banq.qc.ca/collections/collection_numerique/index.html?categorie=8)>). Celles-ci présentent des caractéristiques différentes de la musique enregistrée et les besoins de descriptions seront conséquemment différents.

**Tableau 9**  
Collections audio

Émissions radiophoniques de la SRC, CBC, BBC... <sup>a</sup>	Émissions de différents genres (reportage, actualités, entrevues, etc.)
Collection ArchiviaNet de BAC <sup>b</sup>	Émissions de radio, discours, etc.
Bibliothèque numérique de l'INCA <sup>c</sup>	Livres, magazines, quotidiens — audio
Collection audio de l'INA <sup>d</sup>	Archive des productions audiovisuelles françaises
Early Sound Recording Collection and Sound Recovery Project (Smithsonian Institute) <sup>e</sup>	Les tout premiers enregistrements sonores, sur divers supports
The Library of Congress American Memory <sup>f</sup>	Divers enregistrements ayant une importance historique <sup>g</sup>
Macaulay Library <sup>h</sup>	Collection d'enregistrements de chants d'oiseaux, de bruits animaliers

a. <<http://www.radio-canada.ca/radio/>> , <<http://www.cbc.ca/radio/>> , <<http://www.bbc.co.uk/radio/>>

b. <[http://www.collectionscanada.gc.ca/archivianet/020114\\_f.html](http://www.collectionscanada.gc.ca/archivianet/020114_f.html)> - enregistrements non consultables en date de mars 2013.

c. <<http://www.cnib.ca/fr/services/bibliotheque/>>

d. <<http://www.ina.fr/>>

e. <<http://newsdesk.si.edu/factsheets/early-sound-recording-collection-and-sound-recovery-project>>

f. <<http://memory.loc.gov/ammem/index.html>>

g. Chercher « sound tape » pour éviter la musique avec « sound recording ».

h. <<http://macaulaylibrary.org/>>

**Tableau 10**  
Collections d'objets d'apprentissage

Projet Eurêka <sup>a</sup>	Initiative de la Vitrine Technologie-Éducation (projet de coopération Québec-Wallonie-Bruxelles)
RESPEL (RESources PEdagogiques en Ligne) <sup>b</sup>	Banque belge de ressources pédagogiques en ligne (tous niveaux, toutes disciplines)
Center for International Education Learning Object Collections <sup>c</sup>	Répertoire multinational de collections variées d'objets d'apprentissage

a. <<http://eureka.ntic.org/>>

b. <<http://www.enseignement.be/index.php?page=25937>>

c. <[http://www4.uwm.edu/cie/learning\\_objects.cfm?gid=37](http://www4.uwm.edu/cie/learning_objects.cfm?gid=37)>

### *Collections audio*

On trouve dans les collections audio des enregistrements sonores variés : émissions radiophoniques, discours, livres parlés, etc., tel qu'illustré au tableau 9. Les métadonnées pour les décrire incluent les noms des locuteurs, le responsable de la diffusion, la durée, la langue d'expression et le genre (discours, cours, bulletin de nouvelles, entrevue, livre parlé...).

### *Collections d'objets d'apprentissage*

Les objets d'apprentissage sont définis par l'organisme IEEE Learning Technology Standards Committee (IEEE LTSC<sup>6</sup>) comme étant toute entité, numérique ou non, qui peut être utilisée, réutilisée ou référencée lors de l'apprentissage assisté par ordinateur. Ceci comprend donc des ressources pédagogiques comme des syllabus de cours, des protocoles de travaux pratiques, des examens, des grilles de corrections, des exercices, etc. Les BN du tableau 10 sont des spécimens provenant

de différents endroits. On va typiquement les décrire en précisant le niveau académique visé ainsi que le groupe d'apprenants visés, le type de document, sa langue d'expression, etc.

### *Collections de composantes logicielles*

Il existe des collections de code informatique (modules, fonctions, logiciels plus ou moins élaborés) en libre accès pour les programmeurs. Il s'agit ici de ressources qui sont des objets numériques de manière inhérente, qui n'auraient pas d'existence propre à l'extérieur du monde numérique. Deux répertoires notables sont indiqués au tableau 11. Parmi les métadonnées typiques pour les décrire, on peut relever le langage de programmation utilisé (Perl, C++, Java...), le système d'exploitation pour lequel le logiciel a été développé (Windows, MacOS, Linux), le numéro de version, l'objectif de la fonction, la forme des données en entrée et à la sortie, etc.

6. <<http://www.ieeeltsc.org>>

**Tableau 11**  
Collections de composantes logicielles

SourceForge <sup>a</sup>	Logiciel libre gratuit
Comprehensive Perl Archive Network (CPAN) <sup>b</sup>	Modules logiciels en langage Perl

a. <<http://sourceforge.net/>>.

b. <<http://www.perl.org/cpan.html>>.

**Tableau 12**  
Collections de données

Montréal — données ouvertes <sup>a</sup>	Données ouvertes de la Ville de Montréal
Data.gov.uk <sup>b</sup>	Données ouvertes du gouvernement britannique
Linked Open Data <sup>c</sup>	Données du Web sémantique en format RDF (Resource Description Framework)
Interuniversity Consortium for Political and Social Research <sup>d</sup>	Données variées résultant de projets de recherche en sciences politiques et sociales

a. <<http://donnees.ville.montreal.qc.ca/>>.

b. <<http://data.gov.uk/>>.

c. <<http://linkeddata.org/>>.

d. <<http://www.icpsr.umich.edu/icpsrweb/ICPSR/>>.

### Collections de données

Les données peuvent être des bases de données essentiellement statistiques (par exemple, données de sondage géologique). Plusieurs proviennent d'administrations publiques ; la Ville de Montréal, par exemple, a publié dès 2012 diverses données concernant sa gestion (distribution des votes lors des élections récentes, données sur les jardins communautaires, recensement du mobilier urbain, résultats de sondages, etc.). Il peut s'agir également de collections que l'on appelle « données liées » (*Linked Data*), soit des répertoires dans lesquels des objets du monde réel reçoivent des identificateurs (ou adresses) sur le Web et des données à leur sujet sont publiées en libre accès dans un format lisible par ordinateur. Pour ces collections (illustrées par les exemples du tableau 12), les métadonnées typiques incluent la portée des informations, la source, la dimensionnalité et le format d'encodage.

### Caractéristiques des BN

Il est clair, donc, que des BN très variées existent. Mais qu'ont-elles en commun ? On peut identifier trois caractéristiques communes.

Une collection de ressources qui, parce qu'elles sont en format numérique, partagent les modes d'encodage et de diffusion. Ainsi, un article, une photo, une composante logicielle et un enregistrement sonore peuvent cohabiter dans le même répertoire.

La définition de métadonnées comme « clés de description » des ressources. Elles s'avèrent nécessaires pour identifier, décrire, repérer et interpréter les ressources. Par leur expression commune (en texte, même pour les ressources non textuelles), les métadonnées

deviennent des points de jonction entre les types de ressources.

- i) Des outils de recherche et navigation dans la collection. Ceux-ci reposent essentiellement sur les métadonnées définies. Ils doivent répondre aux attentes et aux besoins des utilisateurs.
- ii) Un examen attentif des modèles de BN (présentés ci-haut) révélerait d'autres points communs, mais les trois éléments ci-dessus suffisent déjà pour comprendre le traitement uniforme qui peut être fait de toutes ces collections hétéroclites. Ces caractéristiques communes ouvrent la porte à des collaborations potentielles et des transferts d'expertise entre les diverses BN.
- iii) Un examen des enjeux de recherche et de développement pour les BN permettra de broser les orientations des travaux à venir.

### Enjeux soulevés par les bibliothèques numériques

Divers enjeux façonnent le développement des BN : des enjeux institutionnels, technologiques, juridiques, économiques, culturels, linguistiques, ainsi que des enjeux liés à l'utilisation, aux métadonnées, à la communication et à la préservation.

#### *Enjeux liés aux aspects institutionnels*

Les caractéristiques de chaque type de « milieu » (archivistique, bibliothéconomique, muséal) et leurs préoccupations traditionnelles influencent la manière dont sont élaborées les collections et les fonctionnalités d'accès à celles-ci. L'exemple le plus probant est le

*De nouveaux modèles économiques doivent être élaborés pour tirer profit de la diffusion de contenu protégé (ou non) alors que la diffusion sur Internet est largement gratuite.*



fait que le terme retenu pour décrire ces réalisations diverses est « bibliothèque », peu importe la nature de la collection. Dans les faits, on note le terme « archives numériques » et « musée virtuel » qui coexistent, mais le terme général demeure « bibliothèque numérique », qui n'est pas neutre et qui a déjà imposé (ou du moins suggéré) son type d'organisation thématique hiérarchique, souvent observé dans les BN. La convergence des institutions vers le modèle unique des BN peut également entraîner des modifications pour chacune. Ainsi, la notion d'exposition issue des musées peut être introduite pour une collection très vaste d'un autre milieu, pour faire la promotion d'un aspect particulier. Par contre, les institutions peuvent se sentir menacées et refuser certaines uniformisations ; chaque institution pourra vivre une période de questionnement et d'affirmation de ses spécificités.

*Enjeux technologiques*

Sans surprise, les BN soulèvent plusieurs enjeux technologiques. D'abord, la numérisation d'œuvres imprimées ou autrement physiques afin de les incorporer aux BN représente une activité importante pour celles dont les ressources originales ne sont pas numériques (c'est le cas, par exemple, de musées voulant mettre à disposition des photographies des objets de leur collection). Celle-ci mobilise des ressources humaines, matérielles et financières.

Ensuite, on voit la création d'infrastructures logicielles pour héberger les BN (Andro *et al.* 2012) et y donner accès grâce à des outils nouveaux ; on note entre autres des outils de traitement automatique de la langue pour exploiter les documents textuels (voir par exemple Da Sylva 2012 et Bernardi 2011, ainsi que l'article de Gagnon dans ce numéro thématique). La prolifération d'outils sera sans doute suivie d'efforts de consolidation et de normalisation.

Également, les équipes de mise sur pied de BN développent des expertises technologiques (Tedd et Large 2005), élaborent des « bonnes pratiques » pouvant s'établir comme normes de facto et établissant ainsi des communautés de pratiques.

Enfin se pose la question d'intégration et d'interopérabilité entre les systèmes existants : avec l'existence aujourd'hui d'un nombre croissant de BN, la question se pose à savoir si elles peuvent communiquer entre

elles, ou si les utilisateurs devront inévitablement se résigner à faire l'apprentissage de différentes interfaces, différentes manières de décrire et de rechercher les ressources.

*Enjeux juridiques*

Ces enjeux gravitent autour du flou juridique concernant le droit d'auteur, face à la diffusion d'œuvres originales protégées (documents écrits, photos, tableaux, performances musicales, objets muséaux, etc.). L'article de Vermeys dans le présent numéro explore cette dimension des BN.

*Enjeux économiques*

De nouveaux modèles économiques doivent être élaborés pour tirer profit de la diffusion de contenu protégé (ou non) alors que la diffusion sur Internet est largement gratuite (voir l'article de Paquin dans ce numéro thématique). On voit une cohabitation de modèles d'accès libre (inspirés par les pratiques de diffusion des résultats de la recherche des scientifiques à l'origine de certaines des BN) et d'accès payants (modèle habituel des bibliothèques qui, si elles donnent accès gratuitement à leurs collections, sont habituées de les acheter). La facilité de traitement des ressources numériques entraîne des modifications importantes dans les chaînes de fabrication (ou d'acquisition) et de distribution des ressources.

*Enjeux culturels*

On peut diviser la question culturelle en deux : les cultures disciplinaires d'une part, et la culture dans un sens plus conventionnel d'autre part. Le premier aspect fait référence aussi bien aux cultures des institutions fondatrices (archives, bibliothèques et musées), dont les pratiques se voient considérablement modifiées, qu'aux cultures des disciplines qui se tournent dorénavant vers les collections numériques pour soutenir leur travail. Celui-ci est, tout autant, profondément transformé. L'accès immédiat et souvent gratuit à des ressources qui auparavant pouvaient être difficiles d'accès (pensons à des manuscrits médiévaux, des tableaux de grands maîtres, de grandes collections de matériel didactique ou des bibliothèques d'images médicales ou géologiques) vient bouleverser à la fois le rythme de travail et les types d'actions possibles à partir de ces ressources.

Le deuxième aspect, celui de la culture dans son sens plus traditionnel, est touché de deux manières. Pour les institutions culturelles, l'apparition de BN produites par des institutions nationales ou régionales accroît la visibilité de celles-ci. En fait, une BN dans leur cas devient en quelque sorte une vitrine culturelle attendue : son absence serait presque suspecte, ou du moins nuisible à leur image, ratant une occa-

sion d'attirer les visiteurs pour qu'ils explorent leur collection. L'initiative de mise sur pied de la bibliothèque européenne Europeana a misé sur la promotion de la culture et du patrimoine européen. Pour les communautés productrices de contenu culturel, l'avènement des BN permet un déploiement et une visibilité inégalés auparavant ; les cultures non dominantes de tous les continents peuvent, à plus faible coût, avoir une présence sur le Web. Cela peut modifier considérablement l'évolution des communautés en question ; encore faudra-t-il tenir compte de l'impact culturel sur l'évaluation d'utilisabilité (voir notamment Barber et Badre 1998, sur le terme « culturabilité »). Tanner (2005), pour sa part, explore divers aspects de la culture face aux bibliothèques numériques, dont la démocratisation de l'accès aux œuvres artistiques de tous genres, la facilitation de la préservation de la culture et le rapprochement des diasporas.

### *Enjeux linguistiques*

Un nombre croissant de collections multilingues exige donc des outils de description et de traitement différents (Oard *et al.* 1999). Le multilinguisme des collections, mais aussi la consolidation de ressources (dictionnaires, corpus) dans des langues diverses, permettent de dépasser l'offre linguistique habituelle (soit des ressources dans les langues européennes et quelques langues asiatiques). La problématique de traitement de la langue se complexifie à mesure de l'ajout de fonctionnalités supplémentaires ; par exemple, une interface multilingue doit être modifiée pour chaque langue si l'on ajoute une fonction de recherche ou d'affichage au système. De même, chaque fonctionnalité nouvelle, comme par exemple la recherche avec un thésaurus, devra potentiellement être adaptée pour chacune des langues des documents de la collection. Ceci ouvre la porte à de grands chantiers de développement de traitements linguistiques, automatiques ou non, adaptés à chaque BN.

### *Enjeux liés aux utilisateurs et à l'utilisation*

Pour améliorer l'offre des BN, les développeurs devront se pencher sur les comportements informationnels de leurs utilisateurs : leurs habitudes de travail, leurs besoins, leurs habitudes de recherche de ressources, la manière dont ils intègrent les ressources repérées dans leur chaîne de travail ou dans leurs loisirs, les modes de collaboration entre les utilisateurs en rapport avec les ressources de la BN, etc. L'ergonomie des systèmes se doit d'être étudiée soigneusement (voir notamment Dinet et Vivian 2007). C'est un aspect non technologique des BN qui revêtira sans doute une grande importance. On peut prévoir des liens importants avec le courant des « humanités numériques » (voir notamment Besser 2004 et Svensson 2010).

*Pour les communautés productrices de contenu culturel, l'avènement des BN permet un déploiement et une visibilité inégalés auparavant ; les cultures non dominantes de tous les continents peuvent, à plus faible coût, avoir une présence sur le Web.*



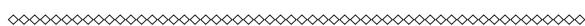
### *Enjeux liés aux métadonnées*

Les métadonnées sont au cœur des fonctionnalités incluses dans les BN pour permettre leur utilisation. Elles représentent les clés d'accès aux ressources, incluant des éléments comme le titre de la ressource, son auteur et une description de celle-ci ; elles peuvent aussi préciser des spécifications comme son format d'encodage numérique, son poids en octets, le logiciel nécessaire pour la consulter, sa date de dernière modification, etc. Les métadonnées servent à identifier les ressources, à les interpréter correctement, à les repérer dans un ensemble qui croît sans cesse (Foulonneau et Riley 2008).

La multiplication des types de BN et par conséquent de la nature de leurs ressources exige souvent que de nouvelles caractéristiques soient exprimées par des éléments de métadonnées. On observe donc la prolifération de normes ou de schémas de métadonnées, parallèlement au célèbre Dublin Core (<<http://dublincore.org/>>), largement répandus pour les ressources numériques. En voici quelques-uns : MODS (Metadata Object Description Schema) pour des ressources bibliothéconomiques, VRA Core v4.0 pour les ressources visuelles, CDWA Lite (Categories for the Description of Works of Art) utilisé en art et architecture, GEM (Gateway to Educational Materials) et IMS Learning Resource Metadata conçus pour décrire les objets d'apprentissage. Le schéma ETD-MS sert à décrire les thèses en format numérique. Le schéma DDI (Data Documentation Initiative) est utilisé pour décrire les ensembles de données issues des travaux de recherche en sciences sociales et comportementales. MIX (NISO Metadata for Images in XML) sert à la gestion d'images. Les éléments du schéma PREMIS (Preservation Metadata : Implementation Strategies Working Group) expriment les caractéristiques importantes pour la préservation. Le schéma ODRL (Open Digital Rights Language) sert à exprimer les politiques et contraintes liées aux droits d'utilisation des ressources numériques. Finalement, METS (Metadata Encoding and Transmission Standard) est une norme pour l'encodage de métadonnées en XML pour les BN, destinée particulièrement aux échanges entre institutions patrimoniales. Il y a une tension entre assurer des descriptions appropriées pour chaque BN



*D'un point de vue culturel, une BN sera attendue de toute institution culturelle : le public s'attendra à pouvoir consulter au moins une partie de ses avoirs par le biais d'une BN. L'appropriation et la démocratisation de la culture seront décuplées par le biais de cet accès facile.*



spécialisées et tendre vers une universalité qui permet aux BN de communiquer et d'échanger entre elles.

Par ailleurs, l'extraction automatique de métadonnées apparaît comme une fonctionnalité intéressante pour traiter de grands volumes de ressources numériques. Les travaux liés à l'extraction de métadonnées, l'indexation automatique ou l'annotation (terme utilisé par le Web sémantique) devraient connaître un fort déploiement dans le contexte des BN.

#### *Enjeux liés à la communication*

Le moteur de développement des BN est la diffusion rendue plus facile. Par ailleurs, de nouveaux espaces communicationnels sont créés. Plusieurs questions se posent maintenant, comme autant d'agendas de recherche. Quelles offres de services transactionnels deviennent possibles pour les organismes par le biais de BN ? Comment la technologie a-t-elle modifié les stratégies de communication avec le public ? Quels nouveaux contextes de communication ont émergé ? Les technologies collaboratives et participatives du Web 2.0 sont d'ores et déjà mises à contribution dans plusieurs BN, entraînant les utilisateurs dans une conversation plurielle asynchrone qui ajoute à la dynamique de simple consultation des ressources.

#### *Enjeux liés à la préservation*

La préservation des documents sous forme numérique, que ceux-ci aient été numérisés ou qu'ils soient nés en format numérique, pose de nombreux défis (voir notamment Gladney 2006). En particulier, le projet InterPares<sup>7</sup> est un projet international de recherche sur la préservation à long terme de l'authenticité des documents d'archives numériques, qui est une des manifestations de l'ampleur de la réflexion sur le sujet depuis une dizaine d'années. Le terme anglais *curation*, difficile à traduire exactement, exprime l'activité de préservation à long terme et de mise en valeur (évoluant dans le temps) des ressources numériques. Ce terme se retrouve aujourd'hui à toutes les tribunes dédiées aux

ressources numériques<sup>8</sup>. Un périodique entier (*International Journal of Digital Curation*) et un congrès international (International Digital Curation Conference) lui sont consacrés. Ce qui préoccupe : non seulement la conservation des supports matériels, mais aussi la préservation de l'intégrité de l'information et de la possibilité d'accès aux contenus par-delà l'évolution des technologies.

### **Avenir : évolution prévue des BN**

Les BN s'installent progressivement dans le paysage des réseaux d'information. Les avantages que procure l'accessibilité simplifiée à divers types d'information leur assurent un avenir aussi solide que celui dont le livre, la bibliothèque, les archives et le musée ont bénéficié par le passé. Il est intéressant de se plier à l'exercice de prédire comment sera façonné leur avenir. Même sans avoir recours à une boule de cristal, on peut faire des prédictions sur l'évolution des BN dans un avenir rapproché.

Du point de vue juridique, on s'attend à une meilleure définition du cadre juridique et de ses implications, sur le droit d'auteur en particulier. Du point de vue économique, on peut prévoir l'incursion massive dans le domaine par des intérêts commerciaux (on en voit déjà un aperçu par le rôle de GoogleBooks dans l'entreprise du HathiTrust) ainsi que l'émergence de modèles hybrides combinant les accès libres aux accès payants. Dans une perspective sociale, on verra sans aucun doute une utilisation croissante des BN par le grand public, ce qui suscitera davantage d'études d'utilisation et d'utilisateurs pour concevoir les interfaces et leurs fonctionnalités, d'une part, et pour mieux comprendre les impacts de l'accès à ces BN pour les utilisateurs, d'autre part.

Linguistiquement, les BN se diversifieront. Des BN nationales, régionales ou ethniques apparaîtront aux quatre coins de la terre (notamment pour des collections patrimoniales), dans autant de langues. Comme les technologies sous-jacentes reposent souvent sur des techniques de traitement automatique de la langue, le multilinguisme multipliera la demande de technologies associées au repérage des ressources. D'un point de vue culturel, une BN sera attendue de toute institution culturelle : le public s'attendra à pouvoir consulter au moins une partie de ses avoirs par le biais d'une BN. L'appropriation et la démocratisation de la culture seront décuplées par le biais de cet accès facile.

Du point de vue de la communication, il est prévisible qu'un rôle grandissant soit réservé à l'interactivité et à des aspects dynamiques, notamment pour les musées virtuels. Du point de vue éducatif, on assiste

7. <<http://www.interpares.org/>>.

8. Voir notamment le site du Digital Curation Centre, à la page Digital Curation Events (<<http://www.dcc.ac.uk/events>> ), qui liste en date du 12 avril 2013 une dizaine d'événements pour l'année qui vient.

déjà à l'intégration des BN dans les environnements numériques d'apprentissage, ce qui ne fera que croître.

Le point de vue technologique implique trois situations problématiques qui demandent des solutions : d'abord, la question de la préservation des ressources (*curation*), face à la dégénérescence des supports, l'obsolescence des logiciels et l'évolution des besoins ; ensuite, le besoin de soutenir la création de métadonnées par des méthodes automatiques, plus particulièrement par des techniques de traitement automatique de la langue ; enfin, les traitements sur architecture distribuée, devant composer avec des index et catalogues multiples et indépendants et des protocoles multiples (bases de données, systèmes de recherche d'information, logiciels de gestion de contenu).

On peut prévoir des développements technologiques issus de l'entreprise privée et d'organismes publics et parapublics qui développent des BN. On verra également un nombre de travaux scientifiques d'exploration sur les divers aspects des BN. Tout ceci se traduira par la croissance et le perfectionnement de l'expertise humaine sur les divers thèmes pertinents pour les BN.

Nous croyons que ce domaine est un terrain fertile pour des travaux de recherche et de développement concertés entre les individus des diverses disciplines et institutions impliquées, qui couvrent une grande partie du spectre d'études en sciences humaines et sociales, en plus de présenter des défis intéressants du point de vue technologique. ◉

## Sources consultées

Andro, Mathieu, Emmanuelle Asselin et Marc Maisonneuve. 2012. *Bibliothèques numériques : logiciels et plateformes*. Paris : ADBS (Sciences et techniques de l'information).

Arms, William. 2000. *Digital libraries*. Londres ; Cambridge (Mass.) : MIT Press.

Bernardi, Raffaella, Frédérique Segond et Ilya Zaihrayeu (dir.). 2011. *Advanced language technologies for digital libraries. International workshops on NLP4DL 2009, Viareggio, Italy, June 15, 2009 and AT4DL 2009, Trento, Italy, September 8, 2009*. Series : Lecture Notes in Computer Science, Vol. 6699.

Besser, Howard. 2004. The past, present and future of digital libraries. In *A companion to digital humanities*. Sous la direction de Susan Schreibman, Ray Siemens et John Unsworth. Malden, Mass. : Blackwell, p. 557-575.

Borgman, Christine L. 2000. *From Gutenberg to the global Information infrastructure : access to information in the networked world*. Cambridge, Mass. : MIT Press.

Brossaud, Claire et Bernard Reber. 2007. *Humanités numériques*. Paris : Lavoisier.

Burnard, Lou et Syd Bauman (dir.). 2007. *TEI P5 : Guidelines for electronic text encoding and interchange*. Charlottesville, VA : TEI Consortium.

Candela, Leonardo, Donatella Castelli, Nicola Ferro, Yannis Ioannidis, Georgia Koutrika, Carlo Meghini, Pasquale Pagano, Seamus Ross, Dagobert Soergel, Maristella Agosti, Milena Dobрева, Akriki Katifori, et Heiko Schuldt. 2008. *The DELOS digital library reference model - Foundations for digital libraries. Version 0.98*. <[http://www.delos.info/files/pdf/ReferenceModel/DELOS\\_DLReferenceModel\\_0.98.pdf](http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf)> (consulté le 12 avril 2013).

- Da Sylva, Lyne. 2012. NLP and Digital Library Management. In Bandyopadhyay, Sivaji ; Naskar, Sudip Kumar ; Ekbal, Asif (réds), *Emerging applications of natural language processing : Concepts and New Research*. Hershey, PA : IGI Global, p. 265-290.
- Desvallées, André et François Mairesse. 2011. *Dictionnaire encyclopédique de muséologie*, Paris : Armand Colin.
- Dinet, Jérôme et Robin Vivian. 2007. Comportements et stratégies des utilisateurs dans les bibliothèques numériques : l'apport de l'ergonomie. In *Usages et pratiques dans les bibliothèques numériques*, sous la direction de Fabrice Papy. Paris : Hermès.
- Fox, Edward A. 1999. The Digital Libraries Initiative - Update and Discussion. *Bulletin of the America Society of Information Science* 26 (1).
- Foulonneau, Muriel et Jenn Riley. 2008. *Metadata for digital resources : implementation, systems design and interoperability*. Oxford : Chandos.
- Gladney, Henry M. 2006. Principles for digital preservation. *Communications of the ACM* 49 (2) : 111-116.
- Gonçalves, Marcos André, Edward A. Fox, Layne T. Watson et Neill A. Kipp. 2004. Streams, structures, spaces, scenarios, societies (5s) : A formal model for digital libraries. *ACM Transactions on Information Systems* 22 (2) : 270-312.
- Jacquesson, Alain. 2010. *Google livres et le futur des bibliothèques numériques : historique du projet, techniques documentaires, alternatives et controverses*. Paris : Cercle de la librairie.
- Jacquesson, Alain et Alexis Rivier. 2005. *Bibliothèques et documents numériques : concepts, composantes, techniques et enjeux*. Paris : Electre-Cercle de la librairie.
- Kahn, R. E. et V.G. Cerf. 1988. *The Digital Library Project volume I : the world of knowbots, (DRAFT) : an open architecture for a digital library system and a plan for its development*. Reston, VA : Corporation for National Research Initiatives.
- Kučera, Henry et Nelson W. Francis. 1967. *Computational analysis of present-day American English*. Providence, RI : Brown University.
- Li, Wei. 2002. *The first search engine, Archie*. <<http://people.lis.illinois.edu/~chip/projects/timeline/1990archie.htm>> (consulté le 12 avril 2013).
- Library of Congress. 2006. *Development of the Encoded Archival Description DTD*. <<http://www.loc.gov/ead/eaddev.html>> (consulté le 12 avril 2013).
- Martin, Chloé et Adeline Pécout. 2011. *Le Guide des bibliothèques numériques*. Paris : FYP.
- Maurel, Lionel. 2008. *Bibliothèques numériques : le défi du droit d'auteur*. Villeurbanne : Presses de l'Essisb.
- Nunberg, Geoffrey. Google's book search : a disaster for scholars. *The Chronicle of Higher Education*, 31 août 2009. <<http://chronicle.com/article/Googles-Book-Search-A/48245/>> (consulté le 12 avril 2013).
- Oard, Douglas, Carol Peters, Miguel Ruiz, Robert Frederking, Judith Klavans et Paraic Sheridan. 1999. Multilingual information discovery and access (MIDAS). *D-Lib Magazine* 5 (10).
- Papy, Fabrice (dir.). 2005. *Les bibliothèques numériques*. Paris : Hermès.
- Svensson, Patrik. 2010. The landscape of digital humanities. *Digital Humanities Quarterly* 4 (1). <<http://digitalhumanities.org/dhq/vol/3/3/000065/000065.html>> (consulté le 12 avril 2013).
- Tanner, Simon. 2005. *Digital libraries and culture. A report for UNESCO*. Londres : King's Consultancy Services.
- Tedd, Lucy A. et Andrew Large. 2005. *Digital libraries : principles and practice in a global environment*. München : K.G. Saur.
- Walsh, Norman. 2010. *DocBook 5 : the definitive guide. 2<sup>e</sup> éd*. Sebastopol : O'Reilly Associates.
- Witten, Ian H., Alistair Moffat et Timothy C. Bell. 1999. *Managing gigabytes : compressing and indexing documents and images*. San Francisco, CA : Morgan Kaufmann.