

Pour des collections numériques durables

# **La diffusion des documents numériques**

**Alain Boucher**

*Directeur de la numérisation*

***Bibliothèque et Archives nationales du Québec***

Stage pratique sur la numérisation

Dakar, 24 janvier -- 1<sup>er</sup> février 2011

- Il dépasse le cadre du présent stage de couvrir en profondeur les questions liées à la diffusion des documents numériques.
- Toutefois, comme la diffusion la plus large possible est généralement l'un des objectifs qu'on poursuit en numérisant des documents, nous allons donner quelques pistes sur trois sujets importants:
  - La question des métadonnées pour les documents numériques
  - La diffusion sur CD-ROM ou sur DVD
  - La diffusion sur Internet

# Les métadonnées

- Pour rendre possible l'accès aux documents numériques, il faut leur associer des renseignements à leur sujet, renseignements qui pourront être utilisés pour effectuer des recherches. On appelle ces renseignements des métadonnées:
  - Métadonnées descriptives (par ex., auteur, titre, sujet, date)
  - Métadonnées techniques ou administratives (par ex., taille et format du fichier, restrictions en matière de droit d'auteur, etc.)
- Les banques de données déjà constituées pour les documents de bibliothèque et les documents d'archives servent généralement de source principale pour les métadonnées.

# Les métadonnées

- Dans une perspective d'échange et de partage, il est toujours préférable de se conformer aux normes internationales pour les métadonnées, au moins pour l'essentiel:
  - *RDA (Resource Description and Access)* pour les documents de bibliothèque  
Voir [www.rdaonline.org](http://www.rdaonline.org)
  - *ISAD (International Standard Archival Description)* et *EAD (Encoded Archival Description)* pour les documents d'archives  
Voir [http://fr.wikipedia.org/wiki/ISAD\(G\)](http://fr.wikipedia.org/wiki/ISAD(G))
- Sans doute va-t-on observer de plus en plus à un rapprochement entre les deux mondes traditionnellement bien distincts des bibliothèques et des archives. Dans l'univers numérique, les différences tendent à s'atténuer.

# Les métadonnées

- Dans le monde numérique, la nécessité de donner accès à une multitude de collections disparates a aussi donné lieu à la définition d'un ensemble minimal de métadonnées, le *Dublin Core*. Voir [www.dublincore.org](http://www.dublincore.org).
- De la même façon, on a aussi défini un protocole pour l'échange des métadonnées, l'*OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting)*. Voir [www.openarchives.org](http://www.openarchives.org).

## Un exemple d'enregistrement OAI-PMH de la Bibliothèque nationale de France (métadonnées Dublin Core) pour un numéro du journal *La Croix* incorporé au RFN

- <record>
- <header>
- <identifier>oai:bnf.fr:gallica/ark:/12148/bpt6k503572j</identifier>
- <datestamp>2008-02-02</datestamp>
- <setSpec>gallica:periodiques:fascicules</setSpec>
- </header>
- <metadata>
- <oai\_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:oai\_dc="http://www.openarchives.org/OAI/2.0/oai\_dc/">
- <dc:identifier>http://gallica.bnf.fr/ark:/12148/bpt6k503572j</dc:identifier>
- <dc:date>1880-04-01</dc:date>
- <dc:description>1880/04/01 (Numéro 1)-1880/04/30. Note : p. 17-18: pages détachées.</dc:description>
- <dc:title>La Croix</dc:title>
- <dc:creator>Bayard</dc:creator>
- <dc:publisher>La Croix (Paris)</dc:publisher>
- <dc:type xml:lang="fr">texte</dc:type>
- <dc:type xml:lang="en">text</dc:type>
- <dc:type xml:lang="fr">publication en série imprimée</dc:type>
- <dc:type xml:lang="en">printed serial</dc:type>
- <dc:language>fr</dc:language>
- <dc:rights xml:lang="fr">domaine public</dc:rights>
- <dc:rights xml:lang="en">public domain</dc:rights>
- <dc:relation>Notice du catalogue : <http://catalogue.bnf.fr/ark:/12148/cb343631418/description></dc:relation>
- <dc:source>Bibliothèque nationale de France</dc:source>
- <dc:relation><http://gallica.bnf.fr/ark:/12148/cb343631418/date></dc:relation><dc:format>application/pdf</dc:format>
- </oai\_dc:dc>
- </metadata>
- </record>

# La diffusion sur CD-ROM ou DVD

- Malgré tout l'intérêt d'Internet, il arrive qu'il soit avantageux de diffuser des documents numériques sur support optique (CD-ROM ou DVD).
- Sous sa forme la plus simple, l'accès aux documents se fait par une arborescence de répertoires (par ex., année – mois – jour pour un journal).
- Il existe aussi des logiciels pour rendre possibles des recherches élaborées de documents sur CD-ROM ou DVD.
- Un logiciel multilingue libre et gratuit qui mérite d'être mentionné: *Greenstone*, réalisé par l'Université de Wakaito en Nouvelle-Zélande et soutenu par l'Unesco. Il peut aussi servir pour la diffusion sur Internet. Voir [www.greenstone.org](http://www.greenstone.org).

# La diffusion sur Internet

- Diffuser des documents numériques sur Internet entraîne forcément la mise en place d'une infrastructure qui peut être complexe et coûteuse.
- Dans le monde des bibliothèques et des archives, l'accès aux documents numériques s'intègre généralement aux catalogues et autres banques de données bibliographiques déjà disponibles sur Internet. On incorpore simplement l'adresse des fichiers numériques dans les notices des documents concernés.
- Il existe des logiciels libres pour la gestion des bibliothèques. Par ex., *Koha*. Voir <http://www.koha-fr.org/>



# La diffusion sur Internet

- Si l'on veut rendre accessibles des collections numériques avec un logiciel qui leur soit propre, on dispose d'une vaste gamme de produits commerciaux (par exemple, *CONTENTdm* d'OCLC) et de logiciels libres (par ex., *DSpace*, développé par le MIT et HP Labs) (voir [www.dspace.org](http://www.dspace.org))
- BAnQ utilise *DSpace* pour la gestion des collections numériques du portail du *Réseau francophone numériqueales numériques (RFN)*.
- Dans le cadre du RFN, la Bibliothèque nationale de France et BAnQ acceptent d'héberger et de diffuser les collections numériques d'autres partenaires du réseau, pour autant qu'elle répondent aux critères d'admissibilité des documents.

# La diffusion sur Internet

- La diffusion des collections numériques des documents de bibliothèque et des documents d'archives est un dossier qui progresse continuellement. Les défis sont nombreux. Pour n'en citer que quelques-uns:
  - L'adressage permanent des documents (faire en sorte que les adresses des documents numériques soient en quelque sorte éternelles)
  - Une juste représentation des diverses langues et cultures
  - Des moyens toujours plus conviviaux d'accès aux ressources, tenant compte par exemple des demandes confuses ou mal exprimées.