

La coopération autour des programmes de numérisation :
Les projets fondés sur l'OAI avec le réseau des bibliothèques pôles associés

Présentation du projet réalisé avec le Conservatoire national
des arts et métiers autour des expositions universelles
et
Le protocole OAI et la bibliothèque numérique du CNAM

par Geneviève DEBLOCK et Pierre CUBAUD - CNAM

La Bibliothèque du CNAM conserve une collection très spécifique d'ouvrages et de périodiques concernant les techniques et les sciences appliquées à l'industrie. Dans cet ensemble, le fonds du XIX^e siècle est riche, fragile, et très consulté. Il correspond à une des périodes fastes du CNAM, et reflète la participation active de ses membres à l'innovation en matière de technologie, et à l'industrialisation. Ceux-ci sont d'ailleurs particulièrement bien représentés dans les expositions nationales, internationales et universelles qui ont jalonné le siècle.

En 2000, a été inaugurée une bibliothèque numérique spécialisée en histoire des sciences et des techniques, le *conservatoire numérique* (CNUM, <http://cnum.cnam.fr>), fruit d'un partenariat entre la Bibliothèque du CNAM, le Centre d'histoire des techniques et de l'environnement (CDHTE), et le Centre d'études et de recherche en informatique du CNAM (CEDRIC) [1, 2, 3]. Cinq cent volumes sont actuellement numérisés (200 000 pages), suivant un programme qui s'établit selon trois axes : l'histoire des sciences et des techniques, l'histoire des institutions et la vulgarisation scientifique et technique.

Il était important pour le CNUM de mettre en valeur le fonds concernant l'abondante littérature parue autour des expositions universelles. En 2005, le CNAM est devenu pôle associé de la BNF. Le premier travail entrepris par la Bibliothèque pour inaugurer cette collaboration a été l'établissement d'un corpus des documents officiels (catalogues des expositions, rapports des jurys internationaux, rapports administratifs, congrès, comptes-rendus des délégations ouvrières, périodiques, musées rétrospectifs), en vue d'une numérisation partagée avec Gallica.

Cette collaboration de contenu prend un véritable sens avec l'installation d'une passerelle OAI-PMH. Celle-ci permet, dès maintenant depuis Gallica, de consulter l'ensemble de la collection numérisée tout en préservant pour chaque site partenaire sa liberté d'évoluer selon un développement personnalisé. Dans la continuité de ce travail, nous pensons à une possible collaboration avec le Musée des arts et métiers, par l'intégration de notices Dublin Core renvoyant à des instruments et des machines entrés au Musée à l'occasion des expositions universelles.

Les travaux réalisés pour l'installation de l'OAI

La construction d'une passerelle OAI ne concerne pour le moment que les notices bibliographiques, mais il est possible d'établir une passerelle incluant aussi d'autres métadonnées, comme les tables des matières, les tables des illustrations et les index.

Un client et un serveur OAI ont été installés. Nous nous sommes appuyés, pour effectuer notre choix, sur le recensement effectué par le site de l'OAI. Le moissonneur retenu est OAIHarvester2, un logiciel Java source-ouverte d'OCLC¹. Le serveur de l'Université d'Oldenburg² intègre le protocole OAIv2. Il est implémenté en Php. Un script de conversion prend les informations de la base de données des notices bibliographiques Marc puis effectue une conversion à la volée des métadonnées Dublin Core.

Les documents numérisés sur le CNUM sont divisés en trois groupes, et donc trois sortes de notices bibliographiques :

- Les « petites monographies », qui comportent un ou plusieurs volumes, mais un seul titre (exemple : Nollet, *Leçons de physique expérimentale*, en 6 volumes),

- Les « périodiques », qui ont une notice chapeau et une notice de dépouillement par volume (exemple : *Les Annales du CNAM*),

- Les « Grosses monographies », monographies en plusieurs volumes comportant un titre général et des titres secondaires (exemples : *Les rapports des jurys internationaux de l'exposition universelle* de 1900, ou bien *La Description des arts et métiers*), qui sont traitées avec une notice chapeau, et des notices de dépouillements sur le CNUM, alors que les normes de catalogage du Dublin Core ne prennent pas en compte ce niveau de granularité.

Pour les deux premières sortes de documents, la conversion a été très simple. Pour les périodiques, seule, la notice chapeau est actuellement moissonnée. Il y aura à revoir les notices Dublin Core lorsque seront établies des notices d'articles.

Pour les « grosses monographies », seule la notice chapeau est moissonnée actuellement. Nous continuerons à établir des notices Dublin Core pour les notices chapeau. Nous devons ajouter dans le script de conversion de chaque notice de dépouillement le titre général dans un champs « relation », ainsi que certains champs de la notice chapeau (les champs sont répétables), car chaque notice du Dublin Core doit être autosuffisante, c'est là certainement sa plus grande faiblesse.

Quelques exemples usuels de requêtes :

- pour obtenir la description du service OAI :
<http://cnum.cnam.fr/oai/oai2.php?verb=Identify>

- pour obtenir la liste des SETS retenus à ce jours :
<http://cnum.cnam.fr/oai/oai2.php?verb=ListSets>

- pour obtenir la liste des notices :
http://cnum.cnam.fr/oai/oai2.php?verb=ListRecords&metadataPrefix=oai_dc

Le moissonnage a ensuite été effectué par la BNF et les notices du CNUM sont ainsi interrogeables pour les utilisateurs de Gallica (fig. 1). La consultation des statistiques d'accès du CNUM montre que pour 23 294 visites effectuées en novembre 2006, il y a eu 2310 accès

¹ <http://www.oclc.org/research/software/oai/harvester2.htm>

² <http://physnet.uni-oldenburg.de/oai/>

depuis les pages de recherche de Gallica, soit un rapport de 10%. Ces chiffres sont encourageants, mais il est encore trop tôt pour conclure à un gain d'audience. Nous poursuivrons ces mesures dans les mois à venir.

Les réflexions autour de l'installation d'une passerelle OAI :

Nos réflexions sont surtout d'ordre organisationnel. Elles débouchent, nous semble t'il, sur le besoin pour chaque équipe de travailler en collaboration et en cohérence, en conservant cependant ses spécificités.

Cinq réflexions, mais d'abord deux questions pratiques : si un ouvrage est supprimé, faut-il supprimer la notice, ou bien signaler dans la notice sa disparition ? Serait-il possible de penser à une organisation telle que chacun des partenaires lancerait lui – même le moissonnage de ses dernières notices ?

1) La visualisation des notices sur un petit site très localisé, spécialisé comme le CNUM, est structurée, organisée. Le moissonnage par Gallica transforme le catalogue du CNUM en une simple liste non structurée. Y a t'il un ordre de moissonnage et de présentation des métadonnées ? Pourrait-on décrire l'organisation même du catalogue dans le cadre du protocole OAI ?

2) Pour rassembler et interroger les notices d'un fonds reconstitué, par exemple sur les expositions universelles, l'utilisation d'une indexation Rameau n'est pas suffisante. Une possibilité est alors la construction d'un « Set » par chaque partenaire. Mais la création d'un « Set » demande une réflexion commune, une démarche commune, une définition précise du contenu (« documents officiels parus à l'occasion des Expositions universelles », ou bien « ensemble de la littérature parue à l'occasion des expositions universelles », ou bien...), une capacité et une volonté de gérer cet ensemble sur le long terme.

3) Le moissonnage par le CNUM des notices Gallica et de celles produites par les autres bibliothèques numériques : quel intérêt, dans quel but, avons-nous les forces pour maintenir une veille? Il s'agirait alors de rassembler les notices correspondant à la littérature technique numérisée, ainsi qu'à certains ouvrages généraux comme les dictionnaires.

4) Notre attitude face à Google print, sans omettre de reconnaître ses défauts et ses qualités : des titres d'ouvrages déjà numérisés sur nos sites paraissent dans Google Print (exemple : l'abbé Nollet, Borgnis, Sigaud de la Fond) et bénéficient d'une bien meilleure visibilité, alors que le travail a souvent été bâclé par Google. Ne serait-il pas préférable de permettre le moissonnage de nos sites et de donner à voir aussi ces documents numérisés de façon plus satisfaisante, tout en continuant en toute liberté à évoluer dans le sens que nous choisissons ?

5) L'intérêt du protocole OAI réside enfin dans la possibilité de téléchargement en bloc des notices descriptives. Même si les notices peuvent être en grand nombre, les volumes d'informations échangées sont en fait de taille comparable à des vidéos, par ex. Un outil de visualisation et d'exploration adapté au grand public permettrait à celui-ci de mieux comprendre le contenu des collections qu'il découvre [4,5].

Remerciements

Les auteurs remercient Niels Henner, ingénieur stagiaire, qui a mis en place la passerelle au CEDRIC durant l'été 2005. Merci également à Joel Berthelin, Françoise Chevalier, Jérôme Dupire et Rodrigo Almeida.

Références

- [1] P. Cubaud et A. Topol. A WWW-based digital library for antiquarian collections . Rapport de recherche du CEDRIC n°99-09, 1999 (en ligne sur <http://cedric.cnam.fr>)
- [2] G. Deblock, B. Rozet et P. Cubaud. Le Conservatoire Numérique des Arts et Métiers : une création patrimoniale. *Bulletin des Bibliothèques de France*, 46(4), 2001.
- [3] P. Cubaud et G. Deblock. Le conservatoire numérique des arts et métiers. Historique du projet et organisation du site. Actes du colloque *La numérisation des textes et des images : techniques et réalisations*. Univ. Lille 3. Janvier 2003.
- [4] P. Cubaud, C. Thiria et A. Topol. Experimenting a 3D Interface for the Access to a Digital Library . *Proc. of the third ACM conf. on Digital Libraries*, June 23-26, Pittsburgh, USA, 1998. pp. 281-2.
- [5] T. Ong, T. et J.J. Leggett, J. Collection understanding for OAI-PMH compliant repositories. *Proc. of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries*, June 07-11, Denver, USA, 2005, pp. 258-9

Geneviève Deblock	Pierre Cubaud
Bibliothèque centrale	CEDRIC
deblock@cnam.fr	cubaud@cnam.fr
Conservatoire national des arts et métiers	
292 rue Saint Martin F-75141 Paris cedex 3	