

L'Aquitaine à travers une collection patrimoniale numérisée : 1886 la bibliothèque numérique de Bordeaux 3

Par : Sylvain Machefert et Grégory Miura.

Le contexte

L'Université Michel de Montaigne Bordeaux 3, pôle associé de la Bibliothèque nationale de France (BnF) au titre de ses collections sur l'Amérique latine, a connu en 2009 une évolution de l'objet de son partenariat qui a été réorienté progressivement vers la numérisation de collections patrimoniales. Motivé par l'acquisition d'un savoir-faire adaptable à différentes formes de production numérique, le Service Commun de Documentation (SCD) a saisi cette opportunité de valoriser les fonds documentaires de l'université tout en s'inscrivant dans la perspective nationale de Gallica. Intégré au projet d'établissement, ce projet de valorisation des collections avait également pour objectif d'offrir de nouveaux services numériques à la communauté universitaire, le SCD apparaissant ainsi comme un partenaire dans le cadre de projets pédagogiques ou scientifiques. Le projet, porté et validé politiquement par l'équipe présidentielle et le conseil scientifique, a associé étroitement la Direction du système d'information (DSI) et le Service communication de l'Université.

La numérisation

Dès le départ il a été décidé que la numérisation serait effectuée par un prestataire externe, le SCD ne souhaitant pas développer une compétence sur cette activité chronophage dont la pérennité n'est pas assurée au sein du service (les fonds patrimoniaux du service, bien qu'ayant un intérêt historique certain, restent limités).

La production des fichiers images et textes a donc été confiée à une société extérieure dans le cadre d'un marché renouvelable, plus adapté aux perspectives budgétaires. Ce marché à bon de commandes nous a permis de définir un certain nombre d'opérations de base : « numérisation en 300 DPI, en niveaux de gris, d'une page d'un ouvrage s'ouvrant à plat », « numérisation en 600 DPI d'une plaque de verre » ... Cette procédure nous permet d'émettre des bons de commande adaptés à notre budget et de connaître précisément et en amont le coût de chaque lot de numérisation.

Le choix des documents à numériser est réalisé par un e bibliothécaire en lien avec un conservateur, et prend en compte : l'intérêt scientifique du document, le caractère inédit d'une version numérisée du document (en particulier nous évitons les doublons avec ce qui se trouve dans Gallica) et la cohérence du projet de numérisation. À l'heure actuelle la numérisation concerne principalement des ouvrages ayant trait à l'histoire régionale et un fonds de plaques de verre dit « Fonds Brutails », contenant plus d'un millier d'originaux de Jean-Auguste Brutails, archiviste de Gironde.

La diffusion

Le choix d' Omeka

Après une étude de marché, le logiciel libre Omeka a été retenu pour diffuser les documents numérisés. Cette solution s'est développée autour du Centre d'Histoire et des Nouveaux Médias¹ (CHNM), acteur majeur de l'informatique documentaire qui développe également Zotero.

Ce choix d'une solution libre se justifie par la présence de compétences informatiques au sein du SCD auxquelles la DSI a donné la main sur un serveur de diffusion qu'elle héberge. Les réalisations existantes autour de l'outil ont permis de mesurer les possibilités d'Omeka, en particulier à Rennes

¹<http://chnm.gmu.edu/>

2² où le contexte était proche de celui de Bordeaux 3.

Enfin, en utilisant ce type d'outil, l'objectif était de rester autonome et de maîtriser les coûts récurrents liés à cette application. Bien que gratuit, la mise en œuvre du logiciel a un coût, lié au temps passé à le configurer et à suivre ses évolutions.

Mise en œuvre

Une fois la solution choisie, la répartition des tâches s'est faite naturellement : la gestion système a été confiée à la DSI (sauvegardes, mises à jour de l'OS ...), l'aspect fonctionnel a été confié à l'équipe "services numériques" du SCD. À partir de janvier 2012, Omeka fut installé et adapté aux attentes et besoins en profitant de la modularité du logiciel. Cette étape s'est traduite par le développement de quelques plugins (gestion des tables des matières³, extraction de l'OCR⁴) et l'adaptation de plugins existants (IABookReader développé par Rennes 2⁵ qui a ainsi pu enrichir son outil des fonctions que nous avons ajoutées à ce plugin).

En parallèle les documents furent chargés sur la base (PDF pour bénéficier des fonctionnalités liées aux tables des matières et à l'OCR, et JPG pour permettre une visualisation dynamique page par page qui n'oblige pas à charger tout le PDF).

Après intégration de la charte proposée par la cellule infographie de l'université et attribution d'un nom, 1886 a été lancé en décembre 2012 avec environ 200 ouvrages portant sur l'histoire locale et régionale et représentant près de 90 000 pages ainsi que plus de 5000 images, le tout représentant une centaine de gigaoctets d'espace disque.

La suite du projet

Dissémination du projet

La taille modeste du fonds numérisé conduit inévitablement à le disséminer pour que les documents trouvent leur(s) public(s). Outre l'intégration des notices dans les dépôts que sont notamment Gallica et la BnsA⁶, il est envisagé de diffuser ces contenus en créant par exemple des liens depuis les articles Wikipédia dans lesquels les ouvrages sont cités⁷.

L'objectif est d'intégrer au maximum 1886 dans le web : rebond vers les notices Mérimée⁸ des monuments illustrés⁹ ou vers les notices Wikipédia des communes. L'objectif à terme étant d'exposer 1886 sur le web de données.

Concernant les contenus, le SCD est associé à différents projets de recherche qui pourraient être valorisés sur 1886, soit en ajoutant de nouveaux corpus, soit à travers le module *Exhibition* d'Omeka qui permet d'éditorialiser les contenus¹⁰.

Au cours de la mise en œuvre du projet, l'architecture d'Omeka a été refondue et l'outil est passé en version 2.0. Cette évolution est relativement simple à mettre en œuvre et bien documentée, mais cette mise à jour n'a cependant pas été appliquée pour le moment car elle entraîne aussi des changements dans la manière de gérer les plugins. Nous devons donc mettre à jour les plugins développés en local avant de pouvoir passer à cette version 2.0.

²<http://bibnum.univ-rennes2.fr/>

³<https://github.com/symac/Plugin-PdfToc>

⁴<https://github.com/symac/Plugin-Extractocr>

⁵<https://github.com/jsicot/BookReader>

⁶Banque numérique du savoir d'Aquitaine <http://bnsa.patrimoines.aquitaine.fr/>

⁷Exemple sur l'article Victor Louis :

http://fr.wikipedia.org/w/index.php?title=Victor_Louis&oldid=87165527#Bibliographie

⁸Base de monuments gérée par le Ministère de la Culture

⁹Voir par exemple : Maison du XVIème - Maison Tenant <http://1886.u-bordeaux3.fr/items/show/3925>

¹⁰Voir par exemple le projet Mame & Fils : <http://mame.crevilles-dev.org/>