

# Calames (1), le catalogue

*Calames est le tout récent catalogue de l'enseignement supérieur dédié aux archives et aux manuscrits.*

*L'article qui suit porte sur Calames en tant qu'interface publique de recherche et de consultation.*

*Une prochaine livraison d'Arabesques donnera à lire un article sur Calames en tant qu'outil de catalogage en EAD.*

## Les archives et les manuscrits

Constituent un patrimoine précieux et discret des bibliothèques d'enseignement supérieur. Pour certaines bibliothèques de recherche, comme la BDIC, la Mazarine ou la Bibliothèque littéraire Jacques-Doucet, il représente une part importante des collections et de l'activité de communication. Ailleurs, dans les bibliothèques universitaires classiques, ces documents ne peuvent occuper le cœur de la politique documentaire, mais étant par définition uniques et irremplaçables, ils méritent d'être connus et mis en valeur.

Dans le passé récent, certaines initiatives, locales ou nationales, ont contribué à mettre ces documents en lumière, au bénéfice des chercheurs et d'un public plus large. Le projet national de numérisation du Catalogue général des manuscrits des bibliothèques publiques de France (CGM) donne une nouvelle dimension à ces initiatives et fournit la clé de voûte pour d'autres projets. De septembre 2007 à janvier 2008, 182 000 notices de manuscrits structurées selon le format XML EAD seront livrées par le prestataire de la rétroconversion, dont environ un cinquième pour les bibliothèques de l'Enseignement supérieur et de la Recherche. Grâce au travail de nos collègues spécialistes des manuscrits qui sont intervenus manuellement pour améliorer la qualité de l'encodage automatisé, la communauté universitaire disposera d'une masse considérable de données de qualité. Cette rétroconversion n'est pas seulement l'aboutissement d'un travail de description qui a commencé sous la Monarchie de Juillet, c'est aussi un nouveau départ.

Dès la livraison de ces fichiers, les bibliothèques devront avoir en mains les outils nécessaires pour mettre à jour, enrichir et mettre en ligne ces données numériques. Au printemps 2006, la sous-direction

des bibliothèques a confié à l'ABES la mission de mettre en place ces outils.

## Home made Catalog

Le projet Calames n'a été sous-traité ni dans sa phase de conception, ni dans sa phase de développement. Après avoir étudié différents scénarios, l'ABES et sa tutelle ministérielle ont décidé de développer Calames en interne. L'architecture a été conçue et le noyau développé à l'été 2006.

Parallèlement s'est constitué un groupe de travail composé de spécialistes des manuscrits ; pour la plupart, ces experts travaillaient déjà depuis longtemps à la rétroconversion du CGM. L'ABES a choisi d'impliquer ce groupe de travail très en amont du projet.\* La souplesse de l'architecture technique offrait une grande marge de liberté pour imaginer ensemble les fonctionnalités de Calames. Durant ce processus de concertation (juillet 2006 - janvier 2007), les principales contraintes furent les spécificités du format EAD, des délais assez courts (15 mois) et le manque de données brutes sur lesquelles s'appuyer.

Développement interne à l'ABES et élaboration collective au sein du réseau sont apparus comme étant deux dimensions inséparables : les fonctionnalités ont été imaginées, évaluées, testées et implémentées avec plus de souplesse que ne le permet en général l'externalisation du cahier des charges et des développements. De plus, ce mode de fonctionnement permettra à l'avenir des ajustements plus rapides en fonction de la satisfaction des utilisateurs et des enrichissements réguliers au fur et à mesure des besoins et des opportunités.

Parmi les enjeux sensibles de la conception d'un tel catalogue en ligne, il y a la nécessité de faire cohabiter plusieurs modes d'utilisation – un mode simple pour le lecteur pressé et un mode avancé.

## Le lecteur pressé

Le lecteur pressé peut se contenter d'une utilisation *a minima*, proche du fonctionnement d'un moteur de recherche généraliste : recherche de mots clés dans une zone unique, liste de résultats, consultation de chaque résultat.

Pour qu'un tel mode de recherche soit satisfaisant, encore faut-il que la liste des résultats soit pertinente. Dans le cas d'un corpus de données hiérarchiques, cela suppose de faire en sorte qu'une requête (portant sur « Balzac » par exemple) ne renvoie pas comme résultats à la fois la notice du fonds et la notice de chacune des pièces qui composent ce fonds et qui ont hérité individuellement de l'indexation « Balzac », qui s'applique collectivement à l'ensemble de ce fonds. Cette question de l'héritage complique grandement le traitement de ce type de corpus hiérarchiques. On ne la trouve pas dans le cas d'un catalogue bibliographique classique, qui indexe des notices séparément les unes des autres.

Le lecteur pressé pourra ignorer toutes les icônes qui donnent accès aux fonctionnalités avancées et permettent une recherche plus précise, une exploration du corpus plus systématique et un affichage des inventaires à géométrie variable.

## Les fonctionnalités avancées

Actuellement, Calames offre plusieurs « fonctionnalités avancées ».

- **Recherche multicritère.** Cette fonctionnalité classique permet de croiser les critères de recherche.

- **Feuilletage des index *Personne physique, Collectivité, Famille, Lieu, Titre d'œuvre et Sujet*.** Les entrées d'index apparaissent au fur et à mesure de la saisie des lettres (autocomplétion). Ces entrées ont soit la forme des entrées du catalogue imprimé, soit la forme normalisée selon les autorités du Sudoc.

Après avoir sélectionné une entrée, le lecteur peut préciser encore le rôle d'une personne (ex. : Cuvier *en tant que destinataire de lettres*) ou l'entrée secondaire d'un sujet (ex. : France – *Flore*). Ces possibilités exploitent tout le travail d'encodage fin des index effectué au cours de la rétroconversion. Loin d'être gratuite, cette finesse permet, par exemple, de rechercher la relation épistolaire entre une personne A (dans le rôle d'auteur) et une personne B (dans le rôle de destinataire).

#### • Filtres *a posteriori* pour affiner la liste des résultats.

Si le lecteur n'utilise pas les différents critères de la recherche avancée, il peut le faire *a posteriori*, en utilisant ces filtres (ou facettes) pour réduire progressivement la liste des résultats. Ainsi, le lecteur peut commencer par une recherche large sur « France », obtenir des centaines de résultats, puis restreindre ce nombre en utilisant la facette « Établissement », puis la facette « Provenance », puis la facette « Langue ». À chaque étape, il est sûr d'obtenir au moins un résultat.

• **Navigation dans l'arborescence des inventaires.** Les catalogues de manuscrits ou d'archives sont des structures de données arborescentes. Les explorer revient à parcourir un arbre, comme on parcourt l'arborescence des dossiers d'un ordinateur. Dans Calames, ces arbres contiennent des milliers d'éléments, parfois situés sur un même niveau. En effet, dans notre contexte, une arborescence EAD ne reflète pas un fonds particulier, mais la totalité de la collection d'une bibliothèque. Il a donc fallu trouver des solutions d'affichage qui conviennent à cet énorme corpus hiérarchique.

• **Affichage des notices à géométrie variable.** Quand on affiche le détail d'une notice, on peut aussi être intéressé par les sous-notices qui décrivent un sous-fonds, un dossier ou un folio. Dans Calames, c'est le lecteur qui compose sa page, en décidant de la quantité d'informations qu'il souhaite voir imprimée ou affichée à l'écran.



1-599  
Manuscrits de la bibliothèque municipale de Caen  
1-599 Breviarium, ad usum ecclesiae Bajocensis.  
20 Calendrier avec mention de décès des abbés de Saint-Etienne

Article 409  
Prise de la ville de Cherbourg par les Anglais  
Copie d'un manuscrit portant le no 124 dans le Catalogue méthodique de la Bibliothèque communale de Cherbourg, publié par G. Arriot, Cherbourg, 1885, 1 vol. in-8°. En tête du manuscrit de Cherbourg on lit la note suivante : « Ce journal est la copie d'un mémoire plus ancien dressé, peu de temps après l'évacuation de Cherbourg, par le frère Zacharie, religieux de la communauté de M. de La Salle... spectateur judiciaire et témoin oculaire de la plupart des faits qui sont rapportés. A Caen, le 24 décembre 1773. Vittel. »

Description physique : Papier, 59 pages, 310 x 200 mm. Broché.  
Date : XIXe siècle.  
Provenance : Cette copie du manuscrit de Cherbourg a été donnée par M. Adrien Pluquet à M. Gaston Lavaury, au moment où celui-ci préparait pour l'Annuaire du Calvados de 1882, p. 177-214, la publication de cette relation d'après le manuscrit 413 (85) de la Bibliothèque de Caen.

Auteur : Zacharie (Frère)  
Oeuvre : Mémoire sur la prise de Cherbourg  
Propriétaire privé : Lavaury (Gaston)  
Pluquet (Adrien)

30 articles associés

I « Mémoires de la maison d'Alençon », par Mannoury.  
Auteur : Zacharie (Frère)  
Oeuvre : Mémoire sur la prise de Cherbourg  
Mémoires de la maison d'Alençon, par Mannoury

III « Les fiefs des vicomtes d'Argentan et Exmes dépendant du domaine desdits Argentan et Exmes, tirés sur le compte de la recette dudit domaine par feu Me Jean Hermon, vivant receveur ancien dudit domaine, année 1536. »

• **Des URL pour mémoriser les requêtes ou les notices.** Au lieu de gérer un panier de requêtes et de résultats et de forcer le lecteur à s'authentifier, Calames ne s'appuie que sur l'un des principes élémentaires du web : les URL (hyperliens). Chaque requête et chaque notice sont associées à une URL durable (pas encore pérenne<sup>1</sup>) que l'on peut garder en favori, citer dans un document, etc.

• **Nuage de mots-clés.** Un nuage de mots-clés donne à voir une liste de termes selon deux dimensions : l'ordre alphabétique et le nombre d'occurrences. Pour chaque bibliothèque, Calames propose le nuage des sujets et le nuage des personnes citées, ce qui esquisse une vue d'ensemble sur les manuscrits de cette bibliothèque.

## Un catalogue ouvert sur le web

Calames est un catalogue en ligne. Sur le web. Dans le web. Il doit pouvoir interagir avec d'autres applications du web.

Chacune des notices de Calames sera indexée par les moteurs de recherche généralistes, qui permettront de rebondir au

cœur même des inventaires, et pas seulement sur la page d'accueil de Calames. Ainsi, en cherchant sur Google ou un autre moteur, on aura la chance de tomber sur telle lettre de René Char (Bibliothèque littéraire Jacques-Doucet) ou tel « Recueil d'aquarelles et dessins originaux de vers et mollusques » (Muséum national d'histoire naturelle).

Les moteurs de recherche fédérée pourront interroger Calames de manière structurée grâce au protocole SRU, qui est le successeur de Z39.50. Ce sera notamment le cas du nouveau CCFr que la BNF met en place. L'existence d'URL persistantes est une autre manière pour Calames d'être accessible depuis n'importe quel point du web, sans

obliger le lecteur à passer par la page d'accueil et une recherche initiale. Chaque bibliothèque de Calames, chaque notice ou chaque entrée d'index peut devenir un point d'entrée de Calames.

Une autre manière de faire de Calames un catalogue ouvert, c'est d'y agréger des informations venues d'ailleurs pour compléter les métadonnées catalographiques EAD. La priorité est de mettre en regard des notices une version numérisée des manuscrits décrits, lorsqu'il en existe. C'est ce que nous avons commencé à faire avec les enluminures de Liber Floridus, mais il faudra continuer avec d'autres images numériques dont les bibliothèques de Calames disposent déjà. Au-delà des images, il existe aussi des transcriptions, des descriptions iconographiques (comme dans Liber Floridus) ou des références bibliographiques associées (présentes dans le Sudoc ou ailleurs).

## Les technologies derrière Calames

Calames s'appuie essentiellement sur les technologies standard du web.

Les données elles-mêmes sont structurées en XML. Elles sont stockées sous forme native dans une base de données (MS SQL Server 2005) qui allie la puissance et les performances du modèle relationnel et de SQL avec les nouvelles possibilités du monde XML (stockage XML natif, langage de requête XQuery). Le mapping entre l'XML et le relationnel est effectué en traitant conceptuellement chaque composant <c> d'un fichier EAD comme une classe d'objet (au sens de la programmation à objet) héritant de certaines propriétés.

L'interface de consultation tire sa réactivité et sa malléabilité de l'utilisation de la technologie AJAX (*Asynchronous Javascript And XML*). Avec AJAX, c'est le navigateur qui travaille, autant que le serveur interrogé. Le navigateur reçoit des données en XML (essentiellement de l'EAD) pour les afficher grâce à Javascript et XSLT. À chaque étape, la page de Calames est rechargée partiellement, et non dans sa totalité, ce qui donne de la réactivité. Globalement, c'est un peu comme si l'interface de Calames était entièrement construite dans le navigateur (côté client) en interrogeant un serveur SRU.

L'ABES s'est appuyée sur des technologies standard et des outils éprouvés pour développer l'outil de consultation et l'outil de production de Calames. Pour être une application ouverte, l'utilisation systématique d'outils *Open Source* n'est une condition ni nécessaire ni suffisante. Ce qui compte, c'est le respect des standards de structuration, de transport et d'affichage des données.

## Un peu plus qu'un catalogue...

Calames n'est donc pas un catalogue en vase clos : c'est aussi un point de départ et un point d'arrivée. Il tire profit des technologies web les plus récentes pour interagir avec d'autres applications web, agréger des données extérieures. Grâce à ces choix d'architecture, Calames pourra devenir plus qu'un simple catalogue, en se développant dans différentes directions.

**Visualisation des données.** Dans la continuité des nuages de mots-clés, Calames expérimentera d'autres modes de visualisation de l'information, qui permettent d'appréhender et d'explorer des corpus sous différents angles. Nous avons déjà travaillé à représenter sous formes de graphes les relations épistolaires à travers Calames.

**Enrichissement du contenu.** Autour des notices, Calames pourrait faire graviter non seulement de l'information déjà présente sur le web (autres catalogues, entrées d'encyclopédie, images, textes...) mais aussi de l'information créée par les utilisateurs de Calames. Au-delà du principe, il y a différentes manières de le réaliser. Entre l'ouverture au grand large qui autoriserait tout un chacun à « poster » librement des mots-

clés et des commentaires et un contrôle éditorial plus sélectif qui n'ouvrirait des droits d'écriture dans Calames qu'à des chercheurs autorisés, il existe bien des degrés intermédiaires, ce qui devrait nous dispenser de verser dans les querelles du type « Pour ou contre le web 2.0 ? ». Sur ce point comme sur d'autres, il faut expérimenter et tirer des leçons.

**Web sémantique.** La forte structuration et la normalisation des données EAD de Calames se prêtent bien à leur intégration à cet écosystème de données structurées en développement qu'on appelle le web sémantique. Cette orientation faciliterait l'agrégation des données de Calames à d'autres données qui ne sont pas sous le contrôle des établissements d'enseignement supérieur. Elle permettrait aussi d'effectuer des requêtes complexes et d'obtenir, en appliquant des capacités de raisonnement sur différents corpus, dont Calames, des réponses inédites. La première étape sera d'exprimer les données EAD selon le modèle RDF.

Une fois encore, qu'il s'agisse des innovations en matière de visualisation de l'information, du web 2.0 ou du web sémantique, l'heure n'est plus à la rêverie ou au dénigrement *a priori*, mais à l'expérimentation et à l'évaluation. Un des atouts de Calames sera de se décliner en versions

locales : les établissements qui le souhaitent pourront bénéficier d'une interface personnalisée, revêtue d'un habillage conforme à l'identité visuelle de la bibliothèque mais aussi, pourquoi pas, de fonctionnalités nouvelles. Ces innovations pourraient être expérimentées et évaluées localement avant d'être éventuellement généralisées.

Yann Nicolas et alii  
✉ [manuscripts@abes.fr](mailto:manuscripts@abes.fr)

### \* Calames

Catalogue en ligne des archives et des manuscrits de l'enseignement supérieur  
✉ <http://www.calames.abes.fr>

Le groupe de travail :  
Jacqueline Artier (Bibliothèque de la Sorbonne), Pascale Heurtel (MNHN), Patrick Latour (Mazarine), Gérard Littler (BNUS), Max Naudi (SDBIS), Fabienne Queyroux (Institut de France), Nathalie Rollet (Sainte-Geneviève), Mireille Vial (BIU de Montpellier).

À l'ABES : Christophe Bonnefond (Chef de projet informatique), Aurélien Charot (Responsable de l'interface web), Agnès Manneheut (Responsable des données), Yann Nicolas (Chef de projet bibliothéconomique).

1 Parallèlement, l'ABES travaille à mettre en place un système d'identifiants pérennes pour les données et les documents qui sont sous sa responsabilité.