

Calames, et après ?

→ UN CATALOGUE DE MANUSCRITS, QUATRE LEÇONS ET MILLE MASHUPS

YANN NICOLAS

Agence bibliographique de l'enseignement supérieur, Abes
nicolas@abes.fr

Yann Nicolas est conservateur à l'Abes où il est responsable du département Études et projets. Agrégé et titulaire d'un DEA, il était auparavant professeur de philosophie. Il est l'auteur d'un article consacré aux FRBR (spécifications fonctionnelles des notices bibliographiques), paru dans *Cataloguing and Classification Quarterly* (2004) et de « Le partenariat entre la Bibliothèque nationale de Nouvelle-Zélande et les Maoris », paru dans le BBF, 2004, n° 3.

Calames¹ est le catalogue en ligne des archives et des manuscrits de l'enseignement supérieur. Il a été conçu et développé par l'Abes en collaboration étroite avec un groupe d'experts². Il recense aujourd'hui les collections d'une quarantaine d'institutions. C'est un catalogue collectif vivant qui continuera de s'enrichir de nouvelles bibliothèques, de nouvelles notices³ et de nouvelles fonctionnalités.

Le cœur des données de Calames provient de la numérisation en EAD⁴ du CGM, Catalogue général des manuscrits des bibliothèques publiques de France⁵. Calames, c'est donc l'histoire d'une série de catalogues impré-

més qui passe du papier au numérique (océrisation), puis du numérique aux métadonnées (structuration fine en EAD), puis des métadonnées au web (mise en ligne de Calames en décembre 2007). Cet article ne vise pas à retracer l'histoire de cette numérisation, ni à décrire en détail les fonctionnalités de Calames⁶. Il s'efforce plutôt de tirer quelques leçons de l'expérience Calames et de suggérer quelques pistes pour demain, pour Calames et d'autres catalogues analogues. Calames n'est qu'une pièce de ce puzzle qu'est le web patrimonial.

Leçon 1. Une notice de catalogue, ça ne se cherche pas, ça se trouve

On le sait désormais, le web est devenu un marché de l'attention, sur lequel une infinité de contenus se disputent une quantité finie d'attention. Les moteurs de recherche sont désormais les places de marché incontournables où se rencontrent les sites et les lecteurs. Google en est la principale. Calames se devait d'y être présent.

Mais que signifie « être présent dans Google ? ». Il ne suffit pas que le site Calames en tant que tel apparaisse dans la liste des résultats d'une requête comme « catalogue manuscrits ». Il faut aller plus loin : une requête comme « mazarine ms 364 » ou encore « Livre de prières de l'abbaye du Mont-Cassin » affiche un lien qui pointe directement vers la notice de

1. <http://www.calames.abes.fr>

2. Bibliothèque centrale du Muséum national d'histoire naturelle, bibliothèque de l'Institut de France, bibliothèque interuniversitaire de Montpellier (section Médecine), bibliothèque interuniversitaire de la Sorbonne, bibliothèque Mazarine, Bibliothèque nationale et universitaire de Strasbourg (BNUS), bibliothèque Sainte-Geneviève, Sous-direction des bibliothèques et de l'information scientifique (SDBIS).

3. Tout comme le Sudoc, Calames est une interface publique de consultation, mais aussi une interface de catalogage mise à la disposition d'un réseau d'établissements. Une dizaine d'établissements travaillent actuellement à enrichir les notices du CGM et à en créer de nouvelles. Une dizaine de nouveaux établissements les rejoindront en 2009 après une période consacrée au rodage de l'outil et à la rédaction des règles de catalogage dans Calames, en coopération avec les bibliothèques relevant du ministère de la Culture.

4. EAD (Encoded Archival Description) est un format pour encoder en XML la description de fonds d'archives et de manuscrits.

5. Seules les données CGM des bibliothèques de l'Enseignement supérieur et de la Recherche ont été versées dans Calames. L'ensemble du CGM numérisé sera interrogeable par le CCfR (www.ccf.fr.bnf.fr).

6. Pour une présentation plus complète, voir les numéros 48 et 49 de la revue *Arabesques*, publiée par l'Abes (www.abes.fr/abes/page,386,arabesques.html).



Figure 1



Figure 2
Un résultat Calames dans Yahoo aujourd'hui



Figure 3
Un résultat Calames dans Yahoo demain, grâce à SearchMonkey

Calames décrivant ce manuscrit, et ce dans les dix premiers résultats. D'une certaine manière, Google devient l'interface de recherche rapide de Calames (figure 1).

Mais il ne suffit pas de figurer dans la liste de Google, même dans les dix premiers résultats. Il faut encore que Google affiche les informations pertinentes, qui décident l'internaute à cliquer sur le lien qui mène à Calames – ou à ne pas cliquer. En effet, une (trop) bonne indexation dans Google peut aussi être source de spam pour certains utilisateurs. Pensons (exemple attesté!) au malheureux fétychiste qui, après avoir saisi « grosse sein », voit s'afficher le ms 2 900 de la BNU, intitulé « *Das grosse Sein. Sténographié* ⁷. Pour éviter ces déconvenues, il faudrait qu'on puisse maîtriser la manière dont s'affichent dans

un moteur de recherche les liens vers Calames. C'est ce que Yahoo vient de proposer, sous le nom de SearchMonkey⁸. Yahoo nous permet désormais d'enrichir en métadonnées de qualité les pages que ce moteur de recherche exploite et d'utiliser ces métadonnées pour améliorer l'affichage des notices Calames dans sa liste de résultats (figures 2 et 3).

On l'a vu, les moteurs de recherche ne sont pas loin de faire office d'interfaces de recherche rapide dans Calames. Avec SearchMonkey de Yahoo, le moteur s'occupe aussi de l'affichage court des résultats. Demain, pour peu qu'ils aient accès aux métadonnées, les moteurs permettront aussi d'effectuer une recherche avancée. Ce qu'il restera, peut-être, aux catalogues, c'est l'affichage détaillé d'une notice et sur-

tout le mode de navigation dans le corpus de métadonnées.

Leçon 2. Un catalogue de bibliothèques n'est pas si différent d'un catalogue de chaussures

Le défi d'aujourd'hui pour un catalogue émanant du monde des bibliothèques est de proposer une interface de recherche et de consultation qui trouve le bon compromis entre la richesse des fonctionnalités et des contenus, d'une part, et la simplicité d'utilisation d'un moteur de recherche, d'autre part.

D'un côté, un site comme Calames doit proposer aux chercheurs des fonctionnalités avancées efficaces qui soient à la hauteur de tout l'investissement qui a été mis dans le catalogage, c'est-à-dire dans la structuration fine de l'information.

D'un autre côté, il faut veiller à ne pas décourager le lecteur en lui imposant une interface dont la complexité ne serait que le reflet mécanique et souvent stérile de la complexité des données elles-mêmes.

Comment approcher un tel compromis? Bien avant les catalogues patrimoniaux, il existe d'autres sites qui sont à la recherche de modes de recherche et de consultation assez fins et efficaces pour satisfaire l'internaute aux habitudes et aux exigences alignées sur sa pratique de Google. Il s'agit des catalogues commerciaux, des sites de vente. Même s'ils ne relèvent ni du même registre ni de la même stratégie que les catalogues de bibliothèques, certains de ces sites commer-

8. Le blog de Calames évoque cette question : <http://calames.wordpress.com/2008/07/22/calames-yahoo-rdf>

7. www.calames.abes.fr/pub/ms/D47A16287

Résultats
Pour les critères suivants : [Mot(s) clé(s) = Lyon]
Résultats 1 - 41 sur un total de 41

- « Extrait d'un livre in-8o, intitulé : Ordonnances et privilèges des foires de Lyon... A Lyo...
PARIS. SORBONNE-BIU Centrale
- Mémoires des intendants sur l'état des généralités de la France, dressés pour l'instruction ...
PARIS. SORBONNE-BIU Centrale MS 1519 154 feuillets
- « Partage de l'hostel de la monoye de Lyon pour le logement des officiers d'icelle, le 10e 8...
PARIS. SORBONNE-BIU Centrale
- Isidore J.-M. Marthoud. Observatoire météorologique de la basilique de Fourvière, Lyon. Résu...
PARIS. SORBONNE-BIU Centrale MS 1642 36 feuillets XXe siècle
- Histoire littéraire de la ville de Lyon, seu Lugdunum sacro-prophanum, par Bullioud S. J.
MONTPELLIER. BU Médecine H 256 Sept volumes. Pas de deuxième volume ; les premier et deuxième tomes sont réunis. Manque l'index 8, ou tome VIII XVIIe siècle
- Missel à l'usage du diocèse de Lyon
PARIS. Bibliothèque Mazarine Ms 418 296 feuillets, plus sept préliminaires ; 2 colonnes 1404

Boîte à outils
Trier les résultats par :
Pertinence
Filtrer les résultats par :
Bibliothèques
SORBONNE-BIU Centrale (27)
MONTPELLIER. BU Médecine (6)
PARIS. Museum
Langues
Sujet
Propriétaire préc.

Figure 4
Un résultat dans Calames

ciaux sont des sources d'inspiration majeures, tant sur le plan de l'ergonomie que sur le plan de l'exploitation avancée des données disponibles. Ceci ne vaut pas seulement pour les sites du secteur culturel marchand, comme Amazon, mais bien au-delà, quelle que soit la nature des articles. Ainsi, le site Endless a pu servir de référence pour le développement de l'interface de Calames. Ce site de vente de chaussures est une des plus belles bibliothèques numériques qui soit. Ironie de l'histoire : nous avons découvert en chemin que ce site était une filiale de... Amazon.

Une des fonctionnalités qu'on a vue migrer du web commercial vers le web patrimonial est la « recherche à facettes ». Il s'agit d'un mode de recherche qui permet de réduire progressivement la liste des résultats, en apposant des filtres successifs à partir d'une recherche initiale très générale. Ainsi, dans Calames, le lecteur peut commencer par une recherche large sur « Lyon », obtenir des centaines de résultats, puis restreindre ce nombre en utilisant la facette Bibliothèque, puis la facette Langue, puis la facette Propriétaire précédent. À chaque étape, il est sûr d'obtenir au moins un résultat. C'est une manière efficace d'exploiter

les index spécialisés de la recherche avancée, systématiquement délaissée par les utilisateurs (figure 4).

Désormais, les bibliothèques ne sont plus un continent à part sur le web. Leurs technologies, leurs formats de données, et surtout leurs utilisateurs sont les mêmes que ceux de n'importe quel site commercial ou grand public. Mais le web ordinaire ne se limite pas à l'expérience minimaliste de Google ni au bric-à-brac stroboscopique des grandes surfaces virtuelles. Le web marchand (Amazon, Endless, iTunes d'Apple) et le web coopératif (Flickr, Delicious, GovTrack.us) regorgent de solutions créatives et diversifiées aux défis du web patrimonial. Pour valoriser et diffuser leur contenu original, les catalogues doivent se banaliser.

Leçon 3. Un catalogue est une base de connaissance, par nature incomplète et faillible

Comme tout autre catalogue, Calames contient des erreurs, c'est-à-dire des affirmations fausses. En creux, ce truisme nous rappelle qu'un catalo-

gue est une base de connaissance(s), au moins au sens ordinaire du terme : une collection d'affirmations qui prétendent être vraies. On raisonne trop souvent en termes de notices, ces paquets de données qu'on s'échange entre bibliothèques, qu'on charge dans nos systèmes, qu'on affiche aux utilisateurs. Au contraire, il faudra de plus en plus arracher la ficelle et l'enveloppe qui font tenir ensemble les données de la notice, et raisonner au niveau plus élémentaire de chaque affirmation de base :

Ce manuscrit a pour titre « xxx »
Ce manuscrit a pour auteur YYY
YYY a pour nom « Blaise Pascal »
...

Cette façon de penser nos données⁹ explicite le sujet de nos catalogues : de quoi parlent-ils ? Quel est le sujet de ces petites phrases ? Une fois qu'on a répondu à cette question, il devient évident que nos catalogues n'ont pas le monopole du discours sur les objets qu'ils décrivent : Calames n'est pas le seul à parler de tel manuscrit, de telle personne, de telle famille, de tel sujet. D'autres catalogues, d'autres

9. Qui est l'essence du modèle RDF, langage de base du web sémantique.

sites, des ouvrages, des encyclopédies, des annuaires, des anthologies... parlent des mêmes choses, et ce, de manière complémentaire ou redondante, parfois contradictoire.

Désormais, la mission d'un catalogue comme Calames n'est plus de présenter son propre corpus de données sur un mode autarcique. Il doit s'ouvrir aux compléments d'information diffusés par d'autres sources fiables disponibles sur le web. Pour Calames, dans un premier cercle, il s'agit de se connecter étroitement aux informations complémentaires présentes dans le catalogue Sudoc¹⁰, grâce au mécanisme des autorités¹¹. Pour le lecteur, il peut être précieux de savoir immédiatement que l'auteur de tel manuscrit est aussi l'auteur de telle monographie ou le sujet de telle thèse. Au-delà du Sudoc, Calames devrait pouvoir « se brancher » sur les bases d'articles (HAL, Persée, Revues.org¹²) ou les ressources pédagogiques (cours en ligne).

Le chantier est encore long qui permettra cette vraie interopérabilité, qui va à la racine de l'information, au lieu d'en rester au niveau de l'échange de notices ou de simples hyperliens. C'est dans cette perspective d'une connaissance partagée et distribuée sur les manuscrits ou les œuvres que les fonctionnalités web 2.0 prennent sens. Si Calames doit très bientôt offrir à ses lecteurs la possibilité d'annoter les notices de manuscrits, c'est bien pour faire remonter à la surface du web certaines connaissances sur ces mêmes manuscrits, émanant de chercheurs ou d'autres institutions patrimoniales. Mais ces connaissances ne doivent pas rester enfouies dans Calames. D'autres sources doivent pouvoir réutiliser les notices et les annotations

10. Système universitaire de documentation, www.sudoc.abes.fr

11. Dans Calames, les catalogueurs peuvent normaliser leurs points d'accès en les rattachant aux autorités du Sudoc. Ce lien a pour premier objectif d'améliorer la qualité des données Calames, mais, effet secondaire positif, il garantit aussi que le Goethe du Sudoc est bien la même personne que le Goethe de Calames.

12. <http://hal.archives-ouvertes.fr>
www.persee.fr
www.revues.org

MS 767 Recueil sur Port-Royal

Fol. 311 « Lettre de Cl. de S^{te} Marthe à M. de Péréfixe. 1664 »

Auteur :	Sainte-Marthe, Claude de (1620-1690)
Destinataire :	Péréfixe, Hardouin de Beaumont de (1606-1671)

Figure 5

savantes de Calames, dans un autre contexte. Cela suppose de mettre en place de nouvelles solutions techniques, qu'il n'est hélas pas possible d'expliquer ni même de résumer ici.

Leçon 4. Un catalogue est aussi un annuaire inversé

Truisme suivant: un catalogue de manuscrits parle de manuscrits. Ce n'est pas faux, mais c'est incomplet. Calames parle également, entre autres entités, de personnes physiques, de collectivités ou de familles. Quand Calames affirme que:

« *Le ms 594 du Château de Chantilly a-pour-auteur*

Arnaud Vidal de Castelnaudary»,

il dit aussi bien que:

« *Arnaud Vidal de Castelnaudary est-auteur-de*

Le ms 594 du Château de Chantilly ».

Sous un certain point de vue, Calames est autant un catalogue de personnes qu'un catalogue de manuscrits. Certes, le point de vue des bibliothèques est légitimement centré sur les documents. Selon cette perspective, on ne mentionne la personne qu'à titre de propriété du document. Mais le point de vue inverse est tout aussi justifié: le manuscrit est un épiphénomène de la personne. Actuellement, en tant que catalogue de bibliothèques, Calames privilégie l'accès par la cote et l'organisation hiérarchique des fonds. On pourrait imaginer un Calames inversé où l'*index personarum* serait mis en avant, avec possibilité de naviguer d'une personne aux personnes associées.

Les catalogues de documents contiennent de précieuses informa-

tions sur les personnes, qu'il serait souvent facile et utile d'expliquer, d'extraire et de publier à destination des chercheurs. Prenons un exemple simple. La notice en figure 5 établit implicitement une relation entre deux personnes:

Claude de Sainte-Marthe

a-écrit-à

Hardouin de Beaumont de Péréfixe,

ce qui implique trivialement que:

Claude de Sainte-Marthe

a-connu

Hardouin de Beaumont de Péréfixe

et que:

Hardouin de Beaumont de Péréfixe

a-connu

Claude de Sainte-Marthe.

Ces petites vérités sont des évidences pour le lecteur du catalogue, mais, aujourd'hui, elles ne sont pas exploitables aisément par des programmes. Il pourrait être utile de constituer une base de connaissance interrogeable qui agrège toutes ces petites informations élémentaires extraites de nombreux catalogues du web patrimonial. Chaque catalogue est un *Who's who* qui s'ignore, ou plutôt un *Who knew whom* en puissance¹³.

13. En écrivant cette phrase, je découvre par hasard un site web nommé... *Who knew whom* (<http://culturalcartography.net>) qui prétend justement représenter de manière graphique les relations interpersonnelles dans le monde de la culture. On regrettera seulement que cette base semble avoir été constituée *ex nihilo*, au lieu de réutiliser les données présentes implicitement dans les catalogues. Fin 2007, l'équipe Calames avait développé une interface interactive de ce genre qui, en exploitant les informations de Calames relatives aux correspondances, affichait tous les correspondants d'une personne, puis, à la demande de l'utilisateur, les correspondants de ces correspondants, ce qui donnait à voir certaines relations intéressantes.

Patrimoine, puzzle, mashup

Le patrimoine documentaire est dispersé entre des centaines d'institutions de conservation. Chaque catalogue est une pièce de cet immense puzzle documentaire formé par l'histoire. Créer une interface de recherche fédérée sur les différents catalogues n'est pas la solution : un méta-catalogue ne fait que juxtaposer. Ce n'est pas en rassemblant toutes les pièces dans une grande boîte et en secouant le tout avec énergie et conviction qu'on reconstitue le puzzle. Le modèle à suivre est plutôt celui du *mashup*, patchwork numérique de différents corpus d'information accessibles en ligne. HousingMaps¹⁴, par exemple, connecte un site d'annonces immobilières avec Google Maps. Dans le domaine des manuscrits, le projet de bibliothèque virtuelle de Clairvaux¹⁵ a l'ambition de reconstituer en ligne la collection de la bibliothèque virtuelle de Clairvaux, aujourd'hui dispersée entre différentes institutions, dont certaines bibliothèques de Clairvaux. Un tel projet serait le candidat idéal pour un *mashup* patrimonial, à condition que cette reconstitution résulte de l'agrégation dynamique de données dispersées et non de la création d'une nouvelle base *ad hoc*.

Ainsi, c'est dans cet esprit du *mashup* que Calames et la base d'enluminures Liber Floridus¹⁶ ont été connectées : c'est en temps réel que Calames va chercher dans Liber Floridus les images correspondant à ses notices.

Le catalogue en miettes

Pour que les chercheurs, les éditeurs et les institutions patrimoniales aient une chance de reconstituer ces puzzles documentaires, il faut que chaque catalogue s'assume comme base de connaissance. Qu'on fasse l'inventaire de toutes les entités dont parle un catalogue (collections, documents, personnes physiques, collecti-

ités, concepts/sujets...). Qu'on baptise ces entités de façon à ce que n'importe quel autre site du web puisse les identifier, les nommer et ainsi partager à son tour les connaissances qu'il possède sur ces mêmes manuscrits, ces mêmes objets...

Dans l'environnement du web, ces noms propres prennent la forme d'URL (www.calames.abes.fr/pub/ms/res/MAZA10697). Mais, les cotes de manuscrits traditionnelles remplissent très bien leur fonction de nom propre global : Paris, Bibliothèque Mazarine, MS. 364. Il faut ensuite que chaque catalogue expose ses données de façon à ce qu'une autre application puisse les exploiter dans ses propres programmes. C'est la notion d'API¹⁷ ou de *web services*. Par exemple, idéalement, une URL comme www.calames.abes.fr/pub/mazarine/res/364 devrait donner la description du manuscrit sous une forme exploitable par d'autres sites. Ce n'est pas le cas aujourd'hui dans Calames ni dans d'autres catalogues. Par contre, l'URL www.calames.abes.fr/prod/EadSearch.aspx?Query=Rech&histoire=fonds%20de%20clairvaux renvoie en XML toutes les notices de manuscrits provenant de Clairvaux – de quoi alimenter un beau *mashup* patrimonial sur Clairvaux. Enfin, pour favoriser l'interopérabilité, il faut désacraliser la notion de notice, l'éclater en une collection de petites vérités élémentaires dont nous avons donné des exemples plus haut. En d'autres termes, certes allusifs, on passe des blocs d'information XML aux assertions en RDF qui constituent les atomes du web sémantique.

Le document et ses collections

Le catalogue patrimonial fait l'inventaire d'une collection bien spécifique, la collection d'une institution de conservation (ou de plusieurs, dans le cas d'un catalogue collectif comme Calames). Pourtant, les pièces de cette collection appartiennent aussi à d'autres collections, moins tangi-

bles¹⁸. Comme l'exemple de la bibliothèque virtuelle de Clairvaux le suggère, des documents physiquement dispersés peuvent être réunis virtuellement dans une collection numérique permanente. Mais, là encore, il s'agit de restaurer l'unité d'une bibliothèque physique, celle de l'abbaye de Clairvaux... en 1472. Or, le rapprochement de documents peut s'appuyer sur une infinité de critères, et pas seulement sur l'unité du lieu de production ou de conservation. Il existe des bibliothèques virtuelles qui rassemblent les documents d'un même auteur, d'une même école, d'un même sujet, d'une même époque, d'une même région... Aujourd'hui, la construction de ces bibliothèques virtuelles demande beaucoup d'efforts et, de ce fait, est menée par les institutions mêmes qui possèdent les documents. Or, il devrait être (presque) aussi simple de réaliser un *mashup* que de produire une bibliographie. Une bibliographie agrège des références dans un document ; un *mashup* patrimonial agrège des métadonnées, des textes et des images dans un nouveau site web.

Un catalogue patrimonial ne doit plus être seulement une interface de recherche et de consultation. Il doit devenir un vivier documentaire dans lequel un chercheur pourra venir puiser pour constituer et diffuser sa propre collection, son propre catalogue, sa propre bibliothèque numérique¹⁹. ●

Septembre 2008

14. www.housingmaps.com

15. www.mediathèque-aggloméro-troyes.fr/bmtroyes/projets/clairvaux.html

16. <http://liberfloridus.cines.fr>

17. Application Programming Interface = interface de programmation.

18. Pour une distinction entre deux conceptions de la notion de collection, les collections-espèces et les collections-clubs : www.enssib.fr/bibliotheque-numerique/document-987, p. 53-56.

19. D'autres communautés peuvent voir dans les *mashups* patrimoniaux une occasion de se réapproprier leurs biens culturels spoliés et dispersés. Je pense moins aux cisterciens qu'aux Maoris (www.enssib.fr/bibliotheque-numerique/document-987, p. 72-74).