

Biblissima : observatoire du patrimoine écrit du Moyen Âge et de la Renaissance

MATTHIEU BONICEL

Observatoire Biblissima
matthieu.bonichel
@biblissima-condorcet.fr

Matthieu Bonicel est coordinateur de l'observatoire Biblissima. Il est également conservateur responsable de l'informatique et de la numérisation au département des Manuscrits de la Bibliothèque nationale de France.

Rappel du contexte et présentation du consortium

Depuis le xvii^e siècle, un grand nombre d'érudits et de chercheurs se sont intéressés à l'histoire de la transmission des textes et des savoirs et à leur conservation dans les bibliothèques. Une longue tradition de recensement des inventaires et des catalogues anciens a conduit à la publication de plusieurs ouvrages, parmi lesquels on peut citer deux grands ouvrages fondateurs : la *Bibliotheca Belgica manuscripta* de Sanderus (1641)¹ et la *Bibliotheca bibliothecarum* de Montfaucon (1739)².

Quelques siècles plus tard, l'offre s'est considérablement enrichie et diversifiée. Les catalogues, rétroconvertis pour la plupart, sont désormais accessibles en ligne, de nombreuses bases de données spécialisées ont vu le jour, d'abord accessibles en local dans les instituts de recherche puis directement sur le web et, fait majeur, les ouvrages eux-mêmes, manuscrits et imprimés, sont désormais acces-

sibles sous leur forme numérisée. En l'espace de quelques années, l'offre de manuscrits médiévaux accessibles en ligne sur Gallica, la bibliothèque numérique de la Bibliothèque nationale de France, est passée de quelques dizaines à plusieurs milliers³. Les éditions de textes, quant à elles, sont également entrées dans un processus de dématérialisation encouragé par le développement du standard TEI⁴.

Face à cette offre pléthorique, il devenait de plus en plus difficile d'organiser sa recherche et d'être en mesure de disposer d'une information exhaustive. L'idée est donc née de construire une *Bibliotheca Bibliothecarum Novissima*, ou *Biblissima*, bibliothèque des bibliothèques du xxi^e siècle. Ce projet, doté d'un budget de 7 millions d'euros dans le cadre du programme Équipex (Équipements d'excellence, permis par le programme Investissements d'avenir⁵), s'est ouvert pour une durée de 7 ans le 1^{er} octobre 2012. L'objectif est double : augmenter de manière significative la masse de données primaires et secondaires (numérisation de manuscrits et d'imprimés anciens, enrichissement de bases de don-

1. Antonius Sanderus, *Bibliotheca Belgica manuscripta, sive Elenchus universalis codicum mss...* Insulis : ex-officinal T. Le Clercq (1641-1644).

2. Bernard de Montfaucon, *Bibliotheca bibliothecarum manuscriptorum nova, ubi quae in numeris pene manuscriptorum bibliothecis continentur...* Parisiis : Briasson (1739).

3. Au 6 septembre 2013, 6 470 manuscrits en français, italien, espagnol, allemand, grec ou latin, datant des origines à 1500, sont accessibles sur <http://gallica.bnf.fr>

4. TextEncoding Initiative. www.tei-c.org

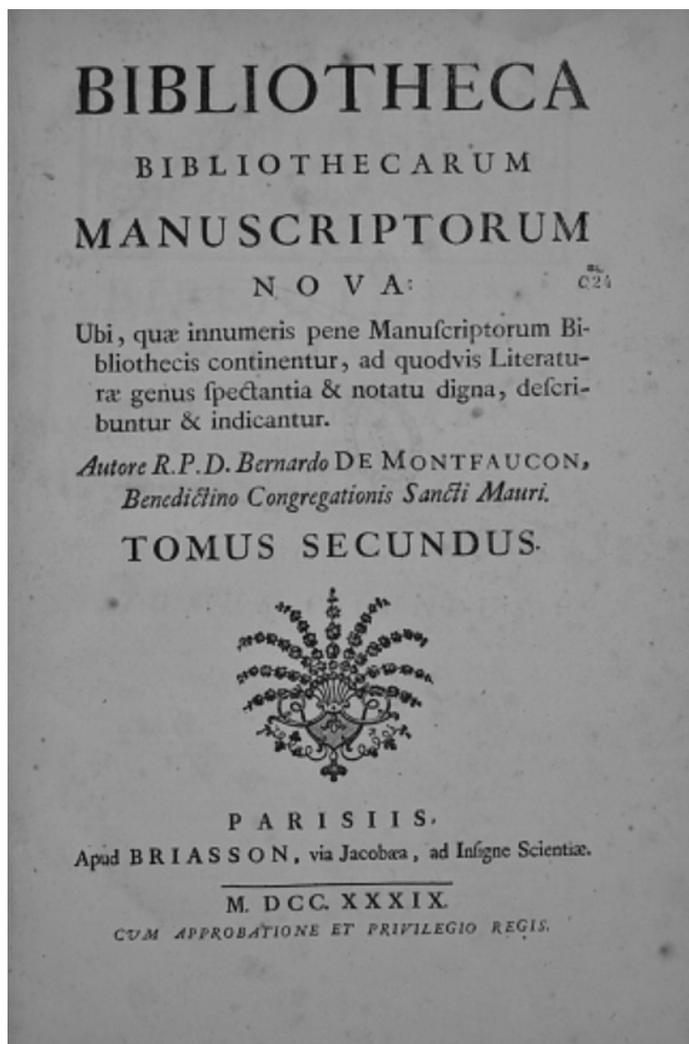
5. Pour en savoir plus sur les Équipex : www.agence-nationale-recherche.fr/investissementsdavenir/AAP-EQUIPEX-2010.html

nées nouvelles ou existantes, constitution d'éditions électroniques...) et construire « l'observatoire Bibliissima », un outil en ligne permettant à la fois une recherche claire et structurée dans cet immense corpus et l'utilisation d'outils innovants destinés à faciliter l'exploitation de ces données par les étudiants et les chercheurs.

Les thèmes de recherche sont nombreux et volontairement très ouverts dans leurs dimensions historique, linguistique et géographique. On y retrouve par exemple les manuscrits enluminés, les textes humanistes, l'étude des fragments de manuscrits hébreux, les bibliothèques byzantines, la transmission des savoirs médicaux au Moyen Âge, l'édition des inventaires de bibliothèques médiévales comme celles du Mont-Saint-Michel ou du roi Charles V et bien d'autres encore.

Bibliissima regroupe neuf partenaires fondateurs : le campus Condorcet (Paris Aubervilliers) – qui en assure le pilotage administratif –, le département des Manuscrits de la Bibliothèque nationale de France, le Centre d'études supérieures de la Renaissance (CNRS, université François Rabelais à Tours), le laboratoire « Histoire, archéologie, littérature des mondes chrétiens et musulmans médiévaux » (CNRS, université Lumière-Lyon 2, EHESS-ENS de Lyon, université d'Avignon et des Pays de Vaucluse, université Jean Moulin Lyon 3), le Centre Jean Mabillon (École nationale des Chartes), le Centre de recherches archéologiques et historiques anciennes et médiévales (université de Caen Basse-Normandie), l'Institut de recherche et d'histoire des textes (CNRS), le Pôle du document numérique de la Maison de la recherche en sciences humaines de Caen (université de Caen Basse-Normandie) et l'Unité savoir et pratiques du Moyen Âge au XIX^e siècle (EPHE).

Fait majeur, et suffisamment rare pour être signalé, Bibliissima dispose d'une équipe propre de huit spécialistes des technologies numériques, le Pool Bibliissima, dont la mission est d'assurer la mise en interopérabilité des données et la construction de l'observatoire.



Page de titre de l'ouvrage de Bernard de Montfaucon, *Bibliotheca bibliothecarum manuscriptorum nova...*

Un vaste corpus

D'ordre, de grandeur et de type divers, les ressources qui composent le futur corpus du portail Bibliissima explorent différents aspects de l'histoire et de la transmission des textes et documents anciens⁶. On y retrouve trois grandes bibliothèques numériques : Gallica, la Bibliothèque virtuelle des manuscrits médiévaux et les Bibliothèques virtuelles humanistes. L'alliance de ces trois grands entrepôts

d'images permet de donner un panorama très représentatif des collections de documents du Moyen Âge et de la Renaissance conservés en France. Tous les documents numérisés dans le cadre du projet seront déposés dans l'un de ces trois réservoirs.

Du côté des bases de données, on y retrouve des outils fondamentaux concernant les manuscrits enluminés, comme Mandragore⁷ ou Initiales⁸, des répertoires de textes comme Jonas⁹ ou des bases centrées sur la tradition des textes de différents uni-

6. La liste complète des ressources composant le portail Bibliissima est consultable à l'adresse : www.bibliissima-condorcet.fr/fr/ressources/ressources-bibliissima

7. <http://mandragore.bnf.fr>

8. <http://initiale.irht.cnrs.fr>

9. <http://jonas.irht.cnrs.fr>

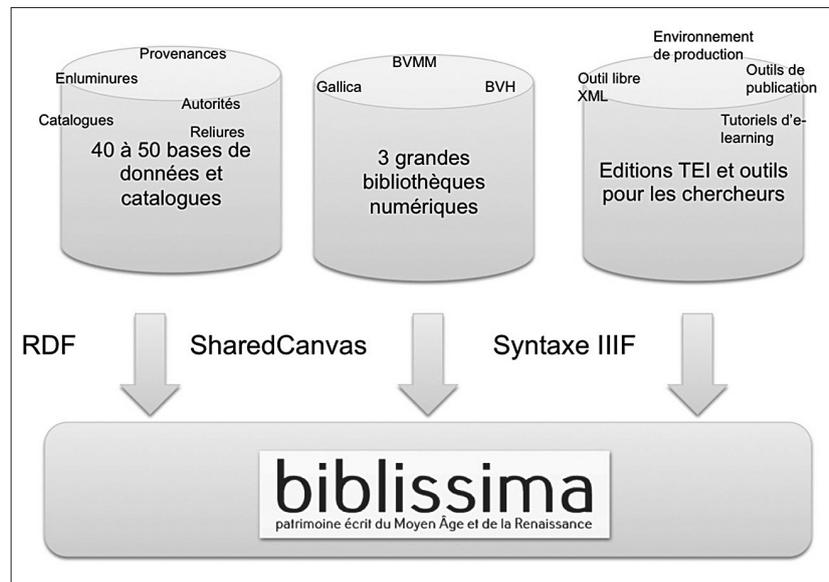
vers linguistiques (Pinakes¹⁰ pour les manuscrits grecs, E-ktobe¹¹ pour les manuscrits syriaques...).

Vers l'interopérabilité profonde

La mise en interopérabilité des contenus existants ou à venir est en effet un des enjeux majeurs du projet. Le caractère disparate des formats de données et leur volume conduisent les membres du Pool à réaliser une étude profonde de chacune des 40 à 50 composantes décrivant les données qui seront versées dans le futur observatoire. Il s'agit en effet de comprendre la structure des informations à fédérer et les relations qu'elles peuvent entretenir entre elles, par le biais d'une modélisation fonctionnelle. Cette dernière, basée sur des modèles conceptuels existants comme le CIDOC-CRM et plus spécifiquement l'initiative FRBRoo¹², devrait aboutir à la publication d'une ontologie décrivant les données Biblissima et proposant une vue schématique aussi précise que possible de la composition même de ce corpus à la fois complexe et étendu. C'est sur cette ontologie qu'il sera ensuite possible de concevoir les développements informatiques nécessaires à la réalisation de l'observatoire lui-même.

Pour que les choses soient claires, il est ici nécessaire de préciser quelque peu l'objectif de Biblissima en matière d'interopérabilité, ce que nous appelons l'« interopérabilité profonde ».

Depuis plusieurs années déjà, il est possible, principalement grâce au protocole OAI-PMH¹³, d'interroger simultanément plusieurs entrepôts de données fournissant des informations, souvent d'ordre bibliographique. Beaucoup d'entre nous utilisent ainsi



L'observatoire Biblissima, description schématique.

des portails comme OAISTER¹⁴ ou Europeana¹⁵. Dans ces deux exemples précis, lorsque l'internaute a trouvé ce qu'il est venu chercher, il est généralement renvoyé sur le site d'origine où sont hébergées les données afin de pouvoir les consulter. Un document numérisé par la BnF par exemple est repérable sur Europeana mais consultable exclusivement dans la bibliothèque numérique d'origine, Gallica.

Souhaitant fédérer de manière plus forte les données qu'il englobe, l'observatoire Biblissima ambitionne un second niveau d'interopérabilité permettant de consulter les données et documents numériques eux-mêmes dans n'importe quel environnement de consultation, pourvu qu'il soit compatible avec la norme adoptée. Ainsi, un chercheur habitué à l'interface mise en place par son laboratoire ou son université pourra directement consulter dans celle-ci plusieurs manuscrits numérisés par différentes institutions et les comparer directement avec, par exemple, une édition électronique ou une notice de base de données. Le tout dans cette même interface, comme l'illustre la figure ci-dessus.

Plusieurs entreprises allant dans ce sens sont actuellement en cours au niveau international, notamment dans le domaine des bibliothèques numériques de manuscrits médiévaux. On peut citer deux d'entre elles avec lesquelles Biblissima collabore de manière soutenue. La première, SharedCanvas¹⁶, vise à établir pour chaque document numérisé ou objet scientifique produit à partir d'un document numérisé, un fichier de référence normalisé, basé sur les standards RDF et OpenAnnotation¹⁷, précisant l'ensemble des informations et fichiers disponibles. Il est ainsi possible de connaître rapidement la totalité des éléments numériques réunis pour un manuscrit médiéval donné. La seconde, IIIF¹⁸, est un ensemble de paramètres et d'éléments de syntaxe destinés à permettre la mise en interopérabilité des visualiseurs de documents numériques. Il est ainsi possible à un visualiseur d'images installé sur un serveur A d'interroger des images numériques conservées sur

10. <http://pinakes.irht.cnrs.fr>

11. www.mss-syriaques.org

12. On peut se référer pour de plus amples informations sur Cidoc-CRM et FRBRoo au site du CIDOC-CRM : www.cidoc-crm.org/frbr_inro.html

13. Open Archive Initiative, Protocol for Metadata Harvesting : www.openarchives.org/pmh

14. OAISTER : <http://oaister.worldcat.org>

15. Europeana : www.europeana.eu

16. SharedCanvas : A Distributed Canvas Rendered from Linked Data Annotations. www.shared-canvas.org

17. OpenAnnotation : Open Annotation Collaboration. www.openannotation.org

18. IIIF : International Image Interoperability Framework. www.sul.stanford.edu/iiif/image-api

un serveur B sans avoir à télécharger l'ensemble des images du document puis à les recompiler. Le serveur A se contente de demander au serveur B le morceau d'image correspondant au niveau de détail que son internaute souhaite consulter¹⁹.

Repenser les usages du numérique dans le domaine patrimonial

Ces exemples, certes techniques, montrent que les avancées récentes dans le domaine des technologies numériques au service de la recherche vont permettre dans les années à venir de repenser complètement l'écosystème environnant les chercheurs et les étudiants qui s'intéressent aux documents patrimoniaux. Là où il était jusqu'alors nécessaire de se constituer son propre environnement numérique en compilant un grand nombre de données disparates (images numériques, notices bibliographiques, notices de bases de données, éditions électroniques...), il sera bientôt possible, pour peu que l'on réalise les développements pertinents, d'offrir à ceux dont la fonction principale reste l'étude du document, des moyens performants de se constituer un corpus d'étude à la fois efficace et étendu au

prix d'un effort technique individuel réduit.

À terme, Biblissima proposera donc sur son propre site un outil de recherche fédérée sur différents ensembles de données mais surtout un environnement de métadonnées permettant aux futurs outils compatibles de consulter directement les données sans transporter sans cesse leurs utilisateurs d'une interface de recherche à l'autre. Les paramètres de cette recherche restent à ce jour à définir de même que les différentes modalités de consultation de certaines bases de données qui peuvent connaître des restrictions, notamment en termes de droit d'accès aux images. La réalisation de l'ontologie des contenus permettra de documenter les modalités d'une telle mise en interopérabilité.

Côté documents numérisés, il sera également possible d'accéder depuis le portail Biblissima à des documents conservés dans différentes bibliothèques numériques au moyen d'un visualiseur exploitant la syntaxe IIIF. Il n'y a pas encore d'interface de ce type en production à ce jour mais un premier test du visualiseur Mirador développé par l'université de Stanford a été mis en ligne²⁰.

Les bibliothèques ont joué un rôle important dans la mise en place de ces processus. Lieux de conservation des documents originaux et numériques, elles sont à la croisée de plusieurs

enjeux de recherche, de constitution et de diffusion du savoir qui ont rapidement mis au jour le besoin fondamental d'assurer la pérennité des données numériques à leur emplacement d'origine, mais aussi tout au long de la chaîne qui les conduit jusque dans l'environnement de travail ou de découverte de l'utilisateur final. La multiplicité des formes d'usage du numérique nécessitait d'entamer aujourd'hui une réflexion sur les moyens de garantir une facilité d'accès aux données primaires, quel que soit le mode de visualisation ou d'exploitation choisi (site d'origine, plate-forme mobile, environnement numérique de travail d'une institution tierce...).

C'est pour cette raison que le département des Manuscrits de la Bibliothèque nationale de France s'est engagé avec enthousiasme dans le projet Biblissima, dont le démarrage a coïncidé avec l'intensification de la réflexion autour de l'interopérabilité dans la communauté des manuscrits médiévaux numérisés. Au-delà de sa contribution informatique avec le catalogue BnF archives et manuscrits, la base Mandragore et la nouvelle base Reliures (pilotee par la Réserve des livres rares)²¹, la BnF numérisera dans le cadre du projet 200 000 images venant enrichir Gallica dans le domaine de l'histoire des bibliothèques. ●

Septembre 2013

19. On peut également consulter sur cet aspect : Matthieu Bonicel, « Hypertexte et manuscrits », *Revue de la Bibliothèque nationale de France*, 42 (2012), p. 23-38. En ligne : <http://hal-bnf.archives-ouvertes.fr/hal-00768031>

20. <http://sul-reader-test.stanford.edu/el-camino> (cet outil étant actuellement en test, il est possible que cette URL soit parfois inaccessible).

21. <http://archivesetmanuscrits.bnf.fr>
<http://mandragore.bnf.fr>
<http://reliures.bnf.fr>