



**HAL**  
open science

## L'expérience de numérisation à la Bibliothèque Sainte-Geneviève

Mathieu Andro

► **To cite this version:**

Mathieu Andro. L'expérience de numérisation à la Bibliothèque Sainte-Geneviève. Manuel de constitution de bibliothèques numériques, Cercle de la Librairie, pp.66-68, 2013, 978-2765414131. hal-01094605

**HAL Id: hal-01094605**

**<https://hal.science/hal-01094605>**

Submitted on 21 Mar 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## L'expérience de numérisation à la Bibliothèque Sainte-Geneviève

Mathieu Andro, Chef des projets numérisation à la Bibliothèque Inter-Universitaire Sainte-Geneviève

La Bibliothèque Sainte-Geneviève est l'une des plus anciennes, des plus grandes et des plus prestigieuses bibliothèques universitaires de France. Pourtant, au-delà de la mise à disposition de ses manuscrits enluminés dans le cadre d'une campagne de numérisation d'enluminures conduite par l'Institut de Recherche et d'Histoire des Textes (IRHT) à la fin des années 1990, l'expérience de la Bibliothèque Sainte-Geneviève, en matière de pilotage de programmes de numérisation est assez récente. En effet, ce n'est qu'en 2009 qu'un premier appel d'offres est lancé, et remporté par la société Diadéis, afin de numériser 321 incunables, 263 récits de voyages nordiques et 348 livres rares et difficilement accessibles du 19<sup>e</sup> siècle. Il sera rapidement suivi par un autre appel d'offres, remporté par la société Aurexus, pour la numérisation de 861 factums des 17<sup>e</sup> et 18<sup>e</sup> siècles, dans le cadre d'un appel à initiatives lancé par la Bibliothèque nationale de France et la Bibliothèque Cujas. Au total, ce sont donc près de 2000 imprimés qui auront été numérisés entre 2009 et 2012, une quantité relativement importante.

La chaîne de traitement du document numérisé comprend de nombreuses étapes pour lesquelles des professionnels divers, des services et des départements multiples vont avoir à intervenir pour la sélection, la description matérielle, le constat d'état, la numérisation, le contrôle qualité des livraisons, l'archivage pérenne au Centre Informatique National de l'Enseignement Supérieur (Cines), la mise en ligne, l'indexation, le signalement dans le SUDOC, et la facturation. Il était donc nécessaire de trouver un mode de fonctionnement adapté à cette multiplicité d'étapes et d'acteurs sollicités. Des *Google Documents* ont été utilisés afin de gérer la complexité de ce *workflow*. Ils ont permis à plusieurs personnes d'intervenir simultanément sur un tableur unique, tout en permettant au prestataire de voir, en temps réel, sur le web, le résultat des contrôles. Les procédures de travail ont également été rédigées de manière itérative, « agile », et participative sous la forme de *Google Documents*. Certaines ont même donné lieu à des démonstrations vidéo *screencasting*<sup>1</sup>.

Concernant la diffusion sur le web des documents numérisés, une étude a été conduite afin de comparer les diverses solutions de diffusion. Cette étude, initialement publiée sous forme de wiki<sup>2</sup> dans un esprit de mutualisation des informations a été reprise et développée dans le cadre d'une enquête *benchmarking* en collaboration avec Marc Maisonneuve et Emmanuelle Asselin (Tosca consultants) et publiée aux éditions de l'ADBS<sup>3</sup>. A l'issue de cette étude, la Bibliothèque Sainte-Geneviève a fait le choix de ne pas développer une énième petite bibliothèque numérique dont les coûts financiers et humains en développement, administration et maintenance auraient été importants pour une visibilité faible et des fonctionnalités limitées. Le choix a été fait de participer à une bibliothèque numérique mutualisée à l'échelle internationale afin de disposer d'une plus forte visibilité, de moyens accrus et d'une pérennité garantie. Dans l'impossibilité d'*uploader* le produit de sa numérisation directement sur Gallica ou sur une plateforme collective de l'Enseignement Supérieur, la Bibliothèque Sainte-Geneviève a décidé de diffuser ses livres numérisés sur la bibliothèque numérique mutualisée Internet Archive. La Bibliothèque Sainte-Geneviève est la première bibliothèque en France à participer à ce projet. Internet Archive est une organisation internationale à but non lucratif, considérée juridiquement comme une bibliothèque par l'Etat de Californie. Elle est membre de l'Open Content Alliance, dispose d'un budget annuel de 15 millions de dollars, d'un effectif de 200 personnes, de data centers dans de nombreux pays dans le monde. Avec ses près de 4,5 millions de livres, dont près de 40 000 provenant du projet Gutenberg et 300 000 du projet Microsoft Live Book Search, Internet Archive se positionne comme la 2<sup>ème</sup> plus importante bibliothèque numérique dans le monde, après Google Books et un peu devant le Hathi Trust pour ce qui concerne les livres en libre accès. Ces moyens considérables, dont la Bibliothèque Sainte-Geneviève n'aurait pu espérer bénéficier seule, lui a permis de disposer de fonctionnalités avantageuses : liens pérennes ARK, lecture audio des textes pour les malvoyants, compatibilité avec des logiciels de gestion bibliographique comme Zotero, gestion des licences juridiques, et surtout, moissonnage OAI-PMH par Gallica et Isidore, création automatique de fichiers EPUB et Kindle pour lecture sur tablettes... La possibilité de lire tous les documents numérisés sur liseuses singularise et distingue, en particulier, le projet de numérisation de la Bibliothèque Sainte-Geneviève. Proposer la lecture de livres numérisés exclusivement sur écrans d'ordinateur n'est guère adapté aux besoins et au confort des usagers.

En participant gratuitement à Internet Archive, la Bibliothèque Sainte-Geneviève n'a paradoxalement pas perdu son identité au sein d'un « grand tout ». Sa plateforme en marque blanche conserve son nom de domaine, son

<sup>1</sup> [http://archive.org/details/BSG\\_Screencast\\_Selection](http://archive.org/details/BSG_Screencast_Selection)

<sup>2</sup> [http://www.bibliopedia.fr/index.php/Solutions\\_logicielles\\_pour\\_biblioth%C3%A8ques\\_num%C3%A9riques](http://www.bibliopedia.fr/index.php/Solutions_logicielles_pour_biblioth%C3%A8ques_num%C3%A9riques)

<sup>3</sup> Mathieu Andro, Emmanuelle Asselin, Marc Maisonneuve. Bibliothèques numériques : logiciels et plateformes. ADBS, 2012, 351 pages, ISBN 978-2-84365-140-3

graphisme, son logo et ses statistiques propres<sup>4</sup>. La Bibliothèque a également largement utilisé la possibilité de créer des accès embarqués sur son propre site vers les documents qu'elle a mis en ligne sur le site [archive.org](http://archive.org)<sup>5</sup>. Mais, c'est surtout au niveau de la visibilité que la participation à Internet Archive s'est révélée être un choix judicieux. En effet, plus une bibliothèque numérique propose de documents, plus le nombre de liens pointant vers son nom de domaine est potentiellement important, plus son indice *PageRank* est élevé, plus elle sort en tête des résultats de recherche dans les moteurs de recherche et, pour finir, plus nombreux sont ses visiteurs. Ainsi, lorsqu'on saisit le titre d'un des livres numérisés par la Bibliothèque Sainte-Geneviève, il apparaît systématiquement en première position dans Google. Si la Bibliothèque avait fait le choix de développer une petite bibliothèque numérique seule dans son coin, il est fort probable que ses livres numérisés aient été relégués au-delà de la 3<sup>ème</sup> page de Google et que les statistiques de consultation aient, par conséquent, été décevantes. En effet, la majeure partie des lecteurs ne prennent pas la peine d'interroger les centaines de petites bibliothèques numériques existantes et préfèrent directement utiliser les moteurs de recherche. Une bibliothèque numérique mal référencée sur le web risque donc de n'avoir qu'une visibilité restreinte et locale, un public très limité qui ne dépasse guère celui qui fréquente déjà l'institution, surtout si ses métadonnées ne sont pas moissonnées par Gallica.

En 2012, les documents numérisés par la Bibliothèque Sainte-Geneviève ont été téléchargés 112 748 fois<sup>6</sup>. Au 1<sup>er</sup> janvier 2013, 932 livres validés avaient été mis en ligne. Ils étaient 432 au 1<sup>er</sup> janvier 2012. En moyenne, chaque document a donc été téléchargé au minimum 10,08 par mois et au maximum 21,75 fois par mois. Ce résultat n'aurait jamais été atteint si le choix d'une bibliothèque numérique locale avait été fait. Il est d'autant plus important au regard du type de documents mis en ligne et au regard d'autres projets de numérisation pour lesquels des statistiques de consultation ont été obtenues. Il est évidemment bien supérieur au nombre de consultations des documents papier. Si on considère, par exemple, l'un des ouvrages numérisés le plus téléchargé à ce jour sur cette bibliothèque numérique, *Genom Palestina, minnen från én resa våren publicé en 1891 par Erik Stave (1857-1932)*, on constate que le livre électronique a été téléchargé, dans le monde, 541 fois en 1 an (en moyenne, 45 fois par mois), alors que le livre imprimé n'a jamais été communiqué, sur place, en plus de 15 ans.

Préoccupée par le fait que la majeure partie des documents numérisés par les bibliothèques n'est pas mise en ligne et demeure sur des supports à durée de vie réduite, la Bibliothèque Sainte-Geneviève a proposé au PRES Sorbonne Paris-Cité le développement d'une plateforme mutualisée<sup>7</sup>. Le projet a ensuite été voté par la ville de Paris et doté d'un budget. La plateforme devrait ainsi enfin offrir un débouché aux bibliothèques qui renoncent encore à diffuser en ligne le produit de leurs campagnes de numérisation. Elle devrait aussi proposer des fonctionnalités innovantes, issues d'un dispositif de veille technologique, telles que la correction participative de l'OCR (*crowdsourcing*), l'annotation collaborative, l'encodage *Text Encoding Initiative* (TEI), la numérisation et l'impression à la demande. Concernant le modèle économique *crowdfunding* de la numérisation à la demande en particulier, il s'agit de permettre, dans le cadre d'une délégation de service public, à un prestataire de venir chercher les documents dont la numérisation aura été commandée, via les catalogues en ligne des bibliothèques, par des lecteurs qui, jusqu'à présent, ne disposaient pas de ce service, par des mécènes qui pourront voir leur nom apparaître sur la notice du livre numérisé, ou encore par des investisseurs qui estimeront que les coûts en numérisation pourront générer un retour sur investissement si on considère le trafic web susceptible d'être généré vers leurs sites web par tel ou tel titre. En partageant la politique d'acquisition de leurs bibliothèques numérique, les bibliothèques pourraient faire supporter par le grand public, les fondations et des investisseurs une partie du financement de leur numérisation, offrir un service de reprographie sans avoir à en supporter ni les coûts, ni la gestion comptable, et bénéficier du travail de sélection par le grand public et le privé des documents pertinents à numériser.

Au-delà de ces ambitieux projets d'envergure parisienne et nationale visant à développer une *Internet Archive* française avec des fonctionnalités scientifiques et innovantes, à la Bibliothèque Sainte-Geneviève, la réussite des projets de numérisation est avant tout le résultat d'une conduite du changement visant à décloisonner des services n'ayant pas l'habitude de travailler ensemble. La conduite du changement a aussi amené les équipes à travailler de manière plus collaborative en utilisant d'avantage les possibilités offertes par les nouvelles technologies de l'information et de la communication. Enfin, elle a amené la Bibliothèque Sainte-Geneviève à s'ouvrir sur un environnement mondial, à l'échelle du web, et sur des partenaires extérieurs publics et privés

<sup>4</sup> <http://archive.org/details/bibliothequesaintegenevieve>

<sup>5</sup> <http://www-bsg.univ-paris1.fr/ressources/bibliotheque-numerique-de-la-bsg>

<sup>6</sup> <http://indexer.us.archive.org/services/collection-downloads.php?mediatype=texts&collection=bibliothequesaintegenevieve>

<sup>7</sup> Mathieu Andro. Genèse d'un projet de plateforme mutualisée pour la diffusion des documents numérisés. Arabesque 64,

ayant une culture de travail et un mode de fonctionnement parfois très différents de celui des conservateurs de bibliothèque.