

Mutualiser pour diffuser

Genèse d'un projet de plateforme mutualisée pour la diffusion des documents numérisés

Une partie importante¹ des documents numérisés par les bibliothèques en France n'est pas diffusée en ligne et repose sur des DVD et disques durs, dont la durée de vie est d'ailleurs limitée².

En effet, les bibliothèques renoncent souvent à développer des plateformes de diffusion trop complexes³ et trop coûteuses en ressources financières et humaines⁴, d'autant que le résultat n'est pas toujours à la hauteur des attentes. Par ailleurs, Gallica, la plateforme de la Bibliothèque nationale de France, ne peut, à ce jour, offrir un débouché à ces bibliothèques et héberger les documents qu'elles ont numérisés, à cause d'un « workflow » adossé au seul catalogue de la BNF⁵.

Dans ces conditions, la mutualisation d'une plateforme semble pertinente ; elle permettrait, en effet, d'accroître la visibilité des projets de numérisation grâce à un volume cumulé de documents numériques⁶ et de limiter leur balkanisation. Par ailleurs, le caractère collectif de la plateforme serait susceptible d'en améliorer la qualité et d'en garantir la pérennité et ce, sans nuire à l'identité de chaque institution qui pourrait bénéficier de sa propre vitrine, de son nom de domaine, de son graphisme et de ses statistiques de consultation. Enfin, les coûts de développement et de maintenance pourraient ainsi être partagés. Un tel projet de mutualisation a tout d'abord été envisagé, à l'initiative de la bibliothèque Sainte-Geneviève, dans le cadre d'un groupe de travail informel de bibliothèques universitaires réunies pour conduire un « benchmarking », organiser des présentations de solutions logicielles et partager des synthèses en intranet. Un premier cahier des charges a ainsi été produit. Par la suite, le PRES, pôle de recherche et d'enseignement supérieur, Sorbonne-Paris-Cité⁷ a manifesté son intérêt pour le projet. Le cahier des charges a donc été revu et corrigé, en collaboration avec la bibliothèque de Sciences Po. La plateforme devrait voir le jour à l'été 2012⁸. Elle pourra d'ailleurs être progressivement ouverte au delà des bibliothèques du PRES.

Les fonctionnalités suivantes, en particulier, sont envisagées : encodage TEI (Text Encoding Initiative), correction participative du texte « océrisé », numérisation à la demande, vente d'impressions brochées à la demande... En attendant de conduire le développement de cette plateforme, la bibliothèque Sainte-Geneviève a décidé d'archiver ses fichiers de conservation sur les serveurs du CINES, le Centre informatique national de l'enseignement supérieur, et de diffuser les documents⁹ qu'elle numérise sur archive.org.

Cette solution de diffusion présente, en effet, les avantages suivants : bon référencement par les moteurs de recherche, forte visibilité, envergure internationale, site public, gratuité, EPUB (electronic publication) à la volée, liens pérennes ARK (Archival Resource Key), compatibilité RSS (Really Simple Syndication), Zotero, et pour finir, moissonnage par Gallica.



Globe céleste

Bibliothèque Sainte-Geneviève

Mathieu Andro

¹ « [...] seulement 42 % des fonds numérisés par les bibliothèques municipales étaient en ligne [...] ». Dutremée, Régis, « La mise en ligne des documents numérisés », BBF, 2007, n° 5, p. 100-101 [en ligne] <http://bbf.enssib.fr/> Consulté le 24 mars 2011.

² CD-R : 1 an, disque dur : 5 ans. D'après l'ouvrage Longévité de l'information numérique : les données que nous voulons garder vont-elles s'effacer ? : rapport du groupe PSN (pérennité des supports numériques) commun à l'Académie des sciences et à l'Académie des technologies / Erich Spitz, Jean-Charles Hourcade, Franck Laloë - EDP sciences, 2010.

³ J'ai publié un article sur les solutions logicielles pour diffuser les documents numérisés : http://www.bibliopedia.fr/index.php/Solutions_logicielles_pour_bibliothèques_numériques.

⁴ Plus de 100 k € et du personnel qualifié pour maintenir les serveurs et administrer les SI.

⁵ Gallica pourra, par contre, importer (« moissonner ») les références de documents numérisés par les bibliothèques et accroître leur visibilité, mais il est néanmoins nécessaire que ces documents aient été préalablement mis en ligne ; or c'est précisément là que se situe la difficulté pour les bibliothèques ; les projets « Gallica tiers-archivage » et « Gallica marque blanche » pourraient néanmoins permettre prochainement aux bibliothèques de disposer d'une vitrine Gallica à leurs couleurs.

⁶ Les algorithmes de référencement des moteurs de recherche tiennent largement compte du nombre de liens qui pointent vers les sites web.

⁷ Il regroupe les universités Sorbonne-Nouvelle, Paris-Descartes, Paris-Diderot, Paris-XIII, Sciences Po, l'EHESP, l'INALCO et l'IPGP.

⁸ Le projet est annoncé sur le site du PRES <http://www.sorbonne-paris-cite.fr/index.php/fr/projets-en-cours/autres-projets/84-plateforme-de-diffusion-de-contenus-numeriques-de-sorbonne-paris-cite>); le nom de domaine bibliotheque-numerique.fr a déjà été acheté.

⁹ Plus de 2 000 documents (280 incunables, 285 livres de voyages nordiques, 600 livres rares et difficilement accessibles du XIX^e siècle, 900 factums des XVII^e et XVIII^e siècles) sont en cours de numérisation à la bibliothèque Sainte-Geneviève par la société Diadésis ; ils sont progressivement diffusés sur <http://www.archive.org/details/bibliothequesaintegeneviève>.

Bibliothèque Sainte-Geneviève

<http://www-bsg.univ-paris1.fr>

Programme de numérisation ☎ 01 44 41 97 53

Mathieu Andro mathieu.andro@univ-paris1.fr