



Indexation collective des images et partage de collections patrimoniales en ligne, l'expérience de Flickr Commons

Date de publication : 28/11/2010

Ce billet est paru dans la revue Bibliothèque(s) sur le thème Nouveaux patrimoines (octobre 2010, n°52) sous le titre « Indexation collective des images et partage de collections patrimoniales en ligne, l'expérience de Flickr Commons », dans le cadre de la rubrique Le billets des hybrides. Il est à mettre en liaison avec le billet publié sur le journal Telemme Infos n°45 de mai 2009, diffusé sur les Carnets de la phonothèque.

Le photochrome représente une route qui longe la mer, en pleine ville. Quelques maisons dominent la côte rocheuse, leurs façades sont claires. Pas très loin du bord de la côte, se devinent des îles à la roche blanche qui reflètent le soleil. Cette image dormait dans les collections de la Bibliothèque du Congrès, (Library of Congress, Loc) sans titre ni localisation, avec une date de prise de vue approximative et quelques informations sur sa technique de production. Sa diffusion sur le site Flickr Commons a suscité une douzaine de commentaires qui ont rapidement permis aux bibliothécaires de compléter leurs métadonnées. Les internautes ont repéré Marseille et sa corniche et, à l'aide d'annotations, appliquées directement sur l'image, précisé les toponymes.. En décidant d'utiliser un outil ancré dans les pratiques d'utilisateurs rodés au Web 2.0, les responsables des institutions participant à Flickr Commons ont ainsi réussi plusieurs défis dont certains ont de quoi faire rêver les conservateurs de bibliothèques : ouvrir leurs collections à un public immense et faire participer ce public à la documentation de leurs fonds.

L'aventure a commencé au sein d'une petite équipe du bureau des initiatives stratégiques de la Bibliothèque du Congrès alors à la recherche de nouvelles audiences et curieuse d'explorer les nouveaux usages de partage collaboratif du Web 2.0. Ils ont travaillé en collaboration avec le département de l'image de la Bibliothèque du Congrès qui commençait à se décourager du manque d'intérêt du public pour les millions d'images numérisées pourtant librement accessibles en ligne. Ils cherchaient un outil de diffusion qui puisse atteindre instantanément des millions de personnes et utilise les normes et les formats en usage pour la photographie¹. L'équipe souhaitait aussi que le projet démarre rapidement. Flickr s'est imposé. Ce site web de partage de photographies et de vidéos² identifié comme un pionnier du Web 2.0 était déjà utilisé par des millions d'amateurs et de professionnels de l'image à travers le monde. Le projet de Flickr Commons a été lancé le 16 janvier 2008 et aussitôt de grandes institutions à travers le monde y ont été associées pour que l'impact patrimonial puisse être démultiplié³.

Le seul défaut de Flickr était son système d'expression des droits développé en direction de photographes présentant leur oeuvres en ligne. Cela ne pouvait correspondre au statut juridique de documents conservés dans les bibliothèques, les centres d'archives ou les musées qui, la plupart du temps, ne détiennent pas eux-même la totalité des droits sur les images mises à disposition du public. La formule « Pas de restriction de droit connue » (No known copyright restriction)⁴ a ainsi été ajoutée.

Les objectifs étaient clairs. Il s'agissait de mettre à la disposition d'internautes qui a priori ne connaissaient pas la LOC, une partie de ses collections d'images photographiques libres de droits. L'intérêt était aussi de faire bénéficier la bibliothèque des avantages des réseaux sociaux. L'entrée dans Flickr permettait d'atteindre

de façon instantanée une communauté habituée au système de tags, de notes et de commentaires. En exploitant le puissant réseau social de Flickr l'idée était de faire travailler les foules en améliorant l'identification des images (sur le web, cela s'appelle le crowdsourcing⁵) et de diffuser les collections bien au-delà des murs de la bibliothèque. Enfin, l'association à d'autres grands fonds patrimoniaux à travers le monde créait un vaste catalogue d'images inédit en permettant une interrogation mutualisée des collections. Le tout devait pouvoir se dérouler sans un investissement trop important en terme de financement comme de temps de travail.

Pour mettre en place le projet, la Bibliothèque du Congrès, en collaboration avec l'équipe de programmation de Flickr, a travaillé sur la structuration des différentes bases de données. Ils ont créé une méthode de dérivation automatique du catalogage des bibliothèques (format MARC) vers les champs de la base Flickr et associé les fichiers images à leur description. La capacité de téléchargement a également été portée à 1500 photographies. L'ensemble des informations technique est décrit avec précision sur un rapport en ligne⁶ qui indique comme coût (hors temps de travail⁷) celui de l'abonnement annuel à Flickr, soit 24.95\$!

Dans Flickr Commons (tout comme dans Flickr), l'utilisateur dispose de trois fonctions pour réagir aux images : les tags, les commentaires de forme libre et les notes directement appliqués sur la photographie. L'équipe à l'origine du projet s'est très vite rendue compte de la force de tous ces mots apportés au fil de leur consultation par les internautes : indexés par la plupart des moteurs généralistes, ils sont aussi une des causes de la réussite fulgurante des Commons.

Au moment du téléchargement des photographies chaque image est systématiquement caractérisée par trois **tags**⁸. Tous les autres sont ajoutés par la communauté. Flickr en limite le nombre à 75 par photographies ; la moyenne des ajouts sur Commons est de 23 tags par image. Les gestionnaires du projet ont choisi de n'effectuer qu'un minimum de corrections, même si il y a des erreurs d'orthographe ou si le terme semble n'intéresser que celui qui l'indique. Très peu de tags qui auraient pu être considérés offensants ont été enlevés⁹ et il arrive quelquefois que les termes des internautes soient ajoutés dans la base originelle à ceux de l'indexation des bibliothécaires, faisant ainsi évoluer leur thésaurus¹⁰. Les **commentaires** ont permis d'ajouter des descriptions à d'anciennes photographies souvent muettes. De véritable « détectives historiques » contribuent à signaler des informations. Des discussions naissent souvent entre membres de Flickr sur ces photos sur des thèmes comme la guerre, le travail des femmes, les pratiques des grand-parents... Les **notes** sont des annotations laissées directement sur l'image. Elle permettent de faire un focus sur une partie de l'image, par exemple commenter un acte technique, identifier un lieu ou indiquer l'identité d'une personne. Ce sont des sortes de messages graffitis qui peuvent aussi indiquer une humeur, faire des blagues. Leur prolifération pourrait saturer l'image mais ils n'apparaissent que lorsque la souris passe dessus.

Les images mises en ligne sont extrêmement variées : couleur et noir et blanc, lieux sujets et périodes différentes. L'objectif est toujours de toucher des audiences multiples et montrer la diversité des collections des institutions¹¹. Toutes sont téléchargeables librement en haute définition, autorisant ainsi une très large gamme d'usages et une nouvelle vie éditoriale... Elles racontent notre histoire depuis la fin du 19e siècle, avec une dominante de photographies anciennes. Vie quotidienne, monde du travail, évolution des technologies, photographies de voyages ou tout simplement de vacances, portraits d'hommes et de femmes connus ou inconnus, paysages, monuments, événements historiques ou militaires, intérieurs d'habitations... Aujourd'hui The Commons réunit une petite cinquantaine d'institutions mais a dû faire une pause¹² pour intégrer de nouveaux sites, victime de son succès. Vivement que d'autres institutions françaises rejoignent ce projet ! L'intégration pourrait être plus rapide car elles bénéficieront des nouvelles expériences en train de voir le jour¹³ autour de l'image, du rassemblement des collections et de leur partage.

Institutions participant à Flickr Commons en septembre 2010

Australie

- Australian National Maritime Museum (participe au projet depuis 2009)
- Australia State Library of Queensland (participe au projet depuis 2009)
- Australian War Memorial (participe au projet depuis 2008)
- Powerhouse Museum Collection (participe au projet depuis 2008)
- State Library of New South Wales collection (participe au projet depuis 2008)

Canada

- Musée McCord Museum (participe au projet depuis 2008)

Etats-Unis d'Amérique

- Brooklyn Museum (participe au projet depuis 2008 mais a une collection flickr depuis 2006)
- Center for Jewish History NYC (participe au projet depuis 2009)
- Cornell University LibraryCornell (participe au projet depuis 2009)
- DC Public Library (participe au projet depuis 2010)
- The Field Museum Library (participe au projet depuis 2009)
- Galt Museum & Archives (participe au projet depuis 2009)
- George Eastman House (participe au projet depuis 2008)
- Getty Research Institute (participe au projet depuis 2009)
- Jewish Historical Society of the Upper Midwest 2010
- Jewish Women Archives (participe au projet depuis 2009)
- Keene and Cheshire County (NH) Historical Photos (participe au projet depuis 2008)
- The Library of Congress (créateur du projet en janvier 2008)
- The Library of Virginia (participe au projet depuis 2008)
- The Nantucket Historical Association Research Library (participe au projet depuis 2009)New York Public Library (participe au projet depuis 2008)
- Oregon State University Archives (participe au projet depuis 2009)
- Smithsonian Institution (participe au projet depuis 2008)
- SMU Central university library (participe au projet depuis 2009)
- Texas State Archives (participe au projet depuis 2008)
- State Library and Archives of Florida (participe au projet depuis 2008)
- University of Washington Libraries – Image collection (a une collection sur Flickr depuis 2007 mais participe au projet depuis 2010)
- Upper Arlington History UA Archives (participe au projet depuis 2009)
- The U.S. National Archives (participe au projet depuis 2009)

France

- Bibliothèque de Toulouse (participe au projet depuis 2008)

Hollande

- Nationaal Archief (participe au projet depuis 2008)

Islande

- Reykjavík Museum of Ljósmyndasafn Reykjavíkur (participe au projet depuis 2009)

Norvège

- Bergen Public Library (participe au projet depuis 2009)
- Fylkesarkivet i Sogn og Fjordane (participe au projet depuis 2009)

Nouvelle Zélande

- National Library NZ (participe au projet depuis 2008)

Portugal

- Biblioteca de Arte – Fundação Calouste Gulbenkian (participe au projet depuis 2008)

Royaume-Uni

- Imperial War Museum Collections (participe au projet depuis 2009)
- Llyfrgell Genedlaethol Cymru (LIGC) ~ National Library of Wales (participe au projet depuis 2009)
- London School of Economics and Political Science Library (participe au projet depuis 2009)
- The National Archives UK (participe au projet depuis 2008)
- National Galleries of Scotland (participe au projet depuis 2008)
- National Library of Scotland (participe au projet depuis 2008)
- National Maritime Museum (participe au projet depuis 2008 mais a une collection flickr depuis 2006)
- National Media Museum (participe au projet depuis 2008)

Suède

- Swedish National Heritage Board (participe au projet depuis 2010).

¹ Depuis 1965 la norme IPTC (International Press Telecommunications Council) domine les formats de métadonnées pour la description des images. La norme XMP (Extensible Metadata Platform) lancée par Adobe Systems – Photoshop en 2001, semble en train de la remplacer.

² Flickr est apparu sur le Web en février 2004, en 2010 le site hébergerait plus de 4 milliards de photos. Il appartient au groupe Yahoo depuis 2005.

³ Elles étaient 16 au départ et 47 en juillet 2010. Pour les dates d'entrée dans le projet et la localisation des bibliothèques participantes, voir la liste en annexe.

⁴ Cette formule est inspirée du droit anglo-saxon qui a développé des pratiques autour des oeuvres orphelines (« Orphan works »)

⁵ Ce terme a été inventé en 2006 par deux rédacteurs du magazine Wired, Jeff Howe et Mark Robinson. Il s'oppose à l'outsourcing qui renvoie vers l'externalisation des tâches. Le crowdsourcing, au contraire, souligne la contribution ouverte à tous à une œuvre collective qui n'aurait pu être menée seul.

⁶ For the common Good ; the Library of congress flickr pilot project. 30 octobre 2008. [En ligne] www.loc.gov/rr/print/flickr_report_final.pdf

⁷ En terme de temps de travail, il est indiqué dans le rapport qu'aucun membre du projet (une quinzaine de personnes de fonctions différentes) n'a été directement assigné sur ce travail, 222 heures ont été dédiées à la programmation techniques tandis que la réflexion pour la mise en route du projet et les discussions avec Flickr ont pris environ 70 heures.

8 Il s'agit du nom de la bibliothèque et de deux numéros machine d'identification.

9 C'est une angoisse récurrente chez les bibliothécaires et pourtant après un an de mise en ligne seule une petite trentaine de termes « offensifs » ont dû être supprimés. Le rapport sur les Commons signale qu'enlever les spams occupe deux heures de temps de travail par semaine.

10 Par exemple, l'expression « Rosie the Riveter » a été utilisée 73 fois en correspondance à une indexation de type « Women-employment » et « World War, 1939-1945 » il a été intégré dans la base de la LOC.

11 Il s'agit toujours d'extraits des collections, la totalité des corpus doit être consulté sur la base originelle.

12 Voir à ce sujet le billet d'Olivier Ertzscheid, sous le joli titre d'« Embouteillage dans les nuages » sur son blog Affordance, en janvier 2010

13 Je renvoie ici au développement du portail MédiHAL qui a été créé en février 2010 et affichait 2 312 images en juillet 2010