

**ASSISTANT TERRITORIAL DE CONSERVATION DU PATRIMOINE ET DES
BIBLIOTHEQUES PRINCIPAL DE 2^{ème} CLASSE**

EXAMEN PROFESSIONNEL DE PROMOTION INTERNE

SESSION 2017

ÉPREUVE D'ADMISSIBILITÉ :

Rédaction d'une note, à l'aide des éléments contenus dans un dossier portant sur la spécialité dans laquelle le candidat se présente.

Durée : 3 heures

Coefficient : 2

SPECIALITE : DOCUMENTATION

À LIRE ATTENTIVEMENT AVANT DE TRAITER LE SUJET :

- ♦ Vous ne devez faire apparaître aucun signe distinctif dans votre copie, ni votre nom ou un nom fictif, ni initiales, ni votre numéro de convocation, ni le nom de votre collectivité employeur, de la commune où vous résidez ou du lieu de la salle d'examen où vous composez, ni signature ou paraphe.
- ♦ Sauf consignes particulières figurant dans le sujet, vous devez impérativement utiliser une seule et même couleur non effaçable pour écrire et/ou souligner. Seule l'encre noire ou l'encre bleue est autorisée. L'utilisation de plus d'une couleur, d'une couleur non autorisée, d'un surligneur pourra être considérée comme un signe distinctif.
- ♦ Le non-respect des règles ci-dessus peut entraîner l'annulation de la copie par le jury.
- ♦ Les feuilles de brouillon ne sont en aucun cas prises en compte.

Ce sujet comprend 24 pages

**Il appartient au candidat de vérifier que le document comprend
le nombre de pages indiqué**

S'il est incomplet, en avertir le surveillant

Vous êtes assistant de conservation du patrimoine et des bibliothèques principal de 2^{ème} classe en charge de la base de données patrimoniales au sein de l'établissement public de coopération culturelle de CULTUREVILLE. La responsable du pôle documentaire vous demande de rédiger à son attention, exclusivement à l'aide des documents joints, une note relative à l'indexation collaborative.

Liste des documents :

- Document 1 :** « Folksonomies. Les usagers indexent le web » - O. Le Deuff - *bbf.enssib.fr* - bbf n° 4 - Juillet 2006 - 5 pages
- Document 2 :** « Le projet Europeana 1914-1918 : un exemple d'utilisation du crowdsourcing au sein d'une bibliothèque numérique » - C. Jullien - *masterabd.hypotheses.org* - 19 janvier 2016 - 2 pages
- Document 3 :** « Indexation collaborative et folksonomies » - E. Francis et O. Quesnel - *cairn.info* - Documentaliste-Sciences de l'information vol. 44 n° 1 - 2007 - 4 pages
- Document 4 :** « Le crowdsourcing appliqué aux archives numériques : concepts, pratiques et enjeux » - Mémoire de recherche - A. Neroulidis - *enssib.fr* - Juin 2015 - 4 pages
- Document 5 :** « Crowdsourcing : tous chercheurs ! » - L. Cailloce - *lejournel.cnrs.fr* - 16 septembre 2014 - 2 pages
- Document 6 :** « Panorama des nouveaux usages des archives audiovisuelles » - M. Dupeyrat et C. Malherbe - *ina-expert.com* - Mars 2014 - 5 pages

Documents reproduits avec l'autorisation du C.F.C.

Certains documents peuvent comporter des renvois à des notes ou à des documents non fournis car non indispensables à la compréhension du sujet.

Folksonomies

Les usagers indexent le web

Le terme de *folksonomie* est apparu récemment sur le web pour désigner le phénomène d'indexation des documents numériques par l'utilisateur. On rencontre également fréquemment le mot *tag* qui désigne en quelque sorte un mot-clé.

Le terme de *folkologie* est aussi employé, mais plus rarement. L'usage du mot *folksonomie* semble donc plus opportun. L'architecte de l'information Thomas Vander Wal a forgé ce terme en combinant la *taxinomie* (règles de classification, *taxonomy* en anglais) et les usagers (*folk*).

Ce phénomène ne cesse de prendre de l'ampleur avec l'avènement des nouvelles technologies du web, dites « web 2.0 », qui donnent plus de possibilités d'expression à l'internaute. Les articles et les blogs sur le sujet ne cessent de croître mais l'essentiel de ce mouvement s'exprime en langue anglaise. Cependant, les usagers français pratiquent déjà les techniques de tags via le site *Technorati* qui recense les blogs par mots-clés. Ces pratiques suscitent plusieurs questions.

Le phénomène est-il durable ou n'est-ce qu'un effet de mode? Les systèmes de classification et les thésaurus sont-ils menacés par ces systèmes? Débats et interrogations sur ce thème sont actuellement en cours sur le web, notamment depuis l'article un peu provocateur du consultant et spécialiste en réseaux sociaux Clay Shirky [6]. Il est intéressant de noter que ce dernier et Thomas Vander Wal n'hésitent pas à répondre aux critiques émises sur les folksonomies, ce qui crée une dynamique intéressante à suivre.

Caractéristiques

Les folksonomies constituent la possibilité pour l'utilisateur d'indexer des documents afin qu'il puisse plus

aisément les retrouver grâce à un système de mots-clés. Le concept est lié à l'accroissement et à l'accélération de la production d'informations. Le système diffère donc nettement en cela des systèmes classificatoires classiques, comme la classification décimale universelle ou la classification Dewey, qui s'inscrivent dans des processus plus longs et dont le but est d'obtenir un classement cohérent de documents physiques dont le contenu est inscrit dans la durée. Les folksonomies, au contraire, ne reposent sur aucun thésaurus, ce qui confère à l'utilisateur une liberté totale quant au choix des mots-clés. Les folksonomies sont donc initialement centrées sur l'utilisateur. Ce dernier les utilise dans un but personnel, afin d'organiser son propre système d'information.

Le paradoxe vient sans doute du fait qu'un esprit collaboratif marque le fonctionnement des systèmes dits de tags comme le fameux *Technorati*, l'un des plus importants moteurs d'indexation de blogs. Chaque utilisateur s'organise en fonction de ses besoins propres d'information et n'est plus soumis à une organisation établie par des professionnels, qui s'imposerait à lui. Tout peut alors se trouver indexé par l'utilisateur, depuis ses favoris jusqu'à

Titulaire d'un DEA en information-communication et du Capes de documentation, **Olivier Le Deuff** est enseignant-documentaliste au collège Léon Jozeau Marigné d'Isigny-le-Buat dans la Manche. Webmaster du Guide des égarés, il a contribué aux actes du colloque Le collège unique, outil de promotion collective (IUT de Saint-Denis, 2003) et a publié un article sur l'information literacy dans L'Œil de l'ADBS (n° 6, mars 2006).

ses photos, en passant par ses messages sur son blog. L'esprit collaboratif permet également d'indexer les documents produits par les autres selon ses besoins.

L'ouverture d'esprit caractérise la folksonomie, comme le web 2.0. L'un et l'autre se rapprochent des idéaux premiers d'Internet. Les folksonomies sont donc de ce fait plus proches du monde des logiciels libres que des systèmes propriétaires. C'est pourquoi la notion de partage est souvent évoquée. Les services de photos en ligne comme Flickr, Ryia, Yahoo Photos permettent de partager ses photos, qui peuvent être indexées par d'autres usagers et réutilisées pour d'autres fonctionnalités. Les sites de stockage en ligne comme Box.net ou Overdrive vont également dans cette direction en permettant de stocker des documents personnels qui seront indexés mais qui peuvent également être partagés.

La force des folksonomies réside aussi dans le fait qu'elles ne nécessitent aucun consensus, à l'inverse des taxinomies. Il ne peut leur être reproché a priori des visions politiques ou idéologiques. Un blogueur prétend même¹ : « *Il n'y a pas de politique dans les folksonomies car le système ne comprend pas de niveau méta qui autorise les communautés d'usagers de tags à discuter de la pertinence ou non de taxinomies émergentes. Il y a seulement l'acte de taguer et le résultat cumulé et amplifié de ces tags.* »

Cette affirmation est discutable : l'emploi de mots-clés n'est pas neu-

tre, il est au contraire révélateur. En outre, il faudrait également se demander à qui bénéficie le développement de l'usage des tags.

Deux types de folksonomies

Thomas Vander Wal [8] distingue en effet deux types de folksonomies, les « étroites » (*narrow folksonomies*) et les « générales » (*broad folksonomies*). Il est possible de trouver le terme de *personomie*, qu'il faut alors rapprocher des *narrow folksonomies*.

Cette distinction est importante, car les premières sont surtout utilisées dans un objectif individuel tandis que les secondes privilégient l'aspect collectif et collaboratif du partage d'information. Ainsi les sites de partages de favoris, comme *del.icio.us* ou *Connotea*, sont plutôt des folksonomies générales puisqu'un même site peut être partagé par plusieurs utilisateurs et recevoir le même tag. Ce type de folksonomies s'appuyant sur des réseaux sociaux ne fait pas que classer de l'information et la partager. Il met en relation des usagers qui partagent les mêmes centres d'intérêts.

L'utilisateur indexeur devient à son tour un peu indexé et mis en relation à la fois avec d'autres mots-clés, d'autres sites et d'autres usagers.

Qu'est-ce qu'un tag ?

Le terme qui revient le plus souvent dans les folksonomies est celui de tag qui désigne en fait un mot-clé, une catégorie ou une métadonnée. Le mot tag signifiant en anglais : étiquette de balisage, étiquetage, fléchage, marquage, voire traçage, il est possible de rencontrer des traductions comme *tagage collaboratif*. Voici la définition qu'en donnent Guy Marieke et Emma Tonkin [2] :

« *Que sont vraiment les tags ? Une définition simple serait de dire que les tags sont des mots-clés, des catégories de noms, ou des méta-*

données. Essentiellement, un tag est simplement un jeu de mots-clés librement choisi. Cependant, du fait que les tags ne sont pas créés par des spécialistes de l'information, ils ne suivent aucune indication formelle. Cela signifie que ces items peuvent être catégorisés avec n'importe quel mot définissant une relation entre la ressource en ligne et un concept issu de l'esprit de l'utilisateur. Un nombre infini de mots peut être choisi, dont quelques-uns sont issus de représentations évidentes tandis que d'autres ont peu de signification en dehors du contexte de l'auteur du tag. »

Le tag peut alors prendre toutes les formes possibles, selon le désir de l'internaute et surtout selon sa culture et sa maîtrise de la langue. Le système ne reposant sur aucun thésaurus, des mots absents du diction-

2005 : chronologie d'un succès

Janvier 2005	Technorati utilise le tracking de tags sur les blogs.
Mars 2005	- Le système de tags est un des thèmes d'actualité au sommet d'architecture de l'information à Montréal. - Article de Clay Shirky : « <i>Ontology is overrated</i> ». Dans cet entretien provocateur, il prédit la fin des hiérarchies et des classifications traditionnelles. - Yahoo achète Flickr qui permet de taguer des photos. C'est une des premières acquisitions par un grand groupe d'un site dit Web 2.0.
Août 2005	25 millions de billets seraient tagués selon Dave Sifry*, le créateur de Technorati.
Novembre 2005	Amazon accepte l'usage des tags et Google sort Google base qui permet également l'indexation d'articles par tags.
Décembre 2005	- Yahoo achète del.icio.us pour près de 30 millions de dollars. - Les folksonomies sont saluées comme une des meilleures idées de l'année 2005 par le <i>New York Times</i>.

* Dave Sifry, « *State of the Blogosphere, August 2005, Part 3: Tags and tagging* », <http://www.technorati.com/weblog/2005/08/37.html>

1. La plupart des citations étant en anglais, nous avons effectué la traduction en français. La citation originale se trouve ici : <http://www.test.org.uk/archives/002380.html>

naire ou des néologismes peuvent devenir des tags. De même, certains sites permettent de lier le tag à un nom de personne. C'est le cas du site Tagalag.

Le mouvement ne cesse de s'accroître mais il n'est guère aisé de chiffrer le nombre de billets de même qu'il est difficile d'évaluer le nombre de blogs.

Inconvénients et avantages

Adam Mathes [1], dans son article paru en décembre 2004, résume le problème en une phrase : « *Une folksonomie représente en même temps ce qu'il y a de meilleur et de pire dans l'organisation de l'information.* » Mathes montre ainsi que la force des folksonomies constitue également sa faiblesse. Sa souplesse d'utilisation et son adaptabilité infinie font que l'organisation de l'indexation par l'utilisateur est chaotique.

Les folksonomies suscitent par conséquent de nombreuses critiques, notamment de professionnels de l'information qui y voient parfois des similitudes avec l'emploi des métadonnées. Les folksonomies constituent également un moyen pour attirer des visiteurs sur un site par l'emploi de mots-clés détournés.

Elles peuvent dès lors être source d'*info-pollution*, pour reprendre l'expression d'Éric Sutter [7], vidant de leur sens de nombreux mots-clés. Les photos recevant les mots-clés de *vacances* sont tellement innombrables, que cela n'a plus grand intérêt si ce n'est pour celui qui les a indexés ainsi. Le fait qu'elles ne reposent sur aucun thésaurus fiable introduit de nombreuses confusions et des polysémies dérangeantes. Il en résulte une Babel sémantique qui peut finalement aboutir à un résultat inefficace et à une grande perte de temps.

Le nouveau danger est que certaines entreprises tentent de faire croire, via les systèmes de recommandation par les internautes, à la fiabilité de leurs produits. Or il n'est pas si dif-

ficile de créer de faux usagers. Cette pratique est qualifiée d'*astroturfing*.

Critiques et débats

Beaucoup de spécialistes de l'information, comme l'architecte de l'information sur le web Louis Rosenfeld [5], se montrent très critiques à l'égard des folksonomies : « *Il est facile de dire que les usagers des réseaux sociaux sont parvenus à réaliser ce que les bibliothécaires n'ont pas réussi : un moyen de faire fonctionner des métadonnées dans des collections de contenu largement distribuées et non connectées les unes aux autres. Facile mais faux : les folksonomies sont clairement convaincantes en favorisant une forme de navigation basée sur la sérendipité² qui peut être tout à fait utile. Mais elles ne favorisent pas la recherche et les autres types de navigation de manière aussi précise que des tags émanant de vocabulaire contrôlé par des professionnels.* »

Clay Shirky lui répond sur un autre blog³ : « *Le problème de l'analyse de Rosenfeld provient de son manque total de sens économique. [...] L'avantage des folksonomies n'est pas qu'elles sont meilleures que les vocabulaires contrôlés, mais qu'elles sont meilleures que rien, car les vocabulaires contrôlés ne sont pas adaptables dans la majorité des cas où un tag est requis. Construire, maintenir et améliorer un vocabulaire contrôlé revient très cher par rapport aux folksonomies, à la fois en temps de développement et pour l'utilisateur, notamment l'utilisateur amateur dans l'utilisation d'un système normé.* »

2. De l'anglais *serendipity* (1754) : « *Pure luck in discovering things you were not looking for* » (dict. Webster), en français l'art ou la faculté de trouver ce qu'on ne cherche pas.

3. La totalité de la réponse de Shirky se trouve sur ce blog : http://many.corante.com/archives/2005/01/07/folksonomies_controlled_vocabularies.php

Des moyens de veille et de mesure

Cependant les folksonomies présentent un intérêt notable en ce qui concerne la traque de termes précis (*tracking*), notamment pour faire de la veille. C'est un élément intéressant pour des veilles spécialisées. Par

Le nouveau danger
des folksonomies est
que certaines entreprises
tentent de faire croire,
via les systèmes
de recommandation
par les internautes,
à la fiabilité
de leurs produits

exemple, nous utilisons le *tracking* de Technorati concernant le mot-clé *information literacy*. Nous pouvons ainsi consulter de nombreux blogs où ce terme est employé.

Les folksonomies sont aussi à la base des mesures de popularité. En effet, des sites comme Digg ou l'équivalent francophone Fuzz permettent de mesurer la popularité d'un message sur un blog et affichent sur leur page d'accueil les billets de blogs les plus vus du moment. C'est aussi un moyen de voir comment et à quelle vitesse se propage une information. Il existe pour cela plusieurs sites tels des agrégateurs de tags comme Guten Tag, qui affichent la popularité des tags utilisés.

Mais il ne faut pas confondre popularité et pertinence, qui sont parfois des notions opposées. De même que pour les métadonnées, ce type d'usages ne garantit pas la fiabilité de l'information et est source de manipulations évidentes

pour figurer « en haut de l'affiche ». En effet, ces usages reposent aussi sur l'ego des blogueurs, qui cherchent à être lus et référencés. D'ailleurs, le site Techno-rati oblige les blogueurs qui veulent référencer leurs derniers billets à effectuer de régulières mises à jour de leur site par un *ping* (Packet Internet Groper)⁴ sur le site.

Les nuages de tags permettent d'afficher l'essentiel des thèmes abordés de manière visuelle sur un blog, ce qui est pratique pour voir en un seul coup d'œil l'intérêt du site. Autre avantage de ce système, l'aspect linguistique. En effet, cela permet de voir quels sont les champs lexicaux prédominants ainsi que les langues les plus employées. Au vu du succès des tags en anglais, nous pouvons nous demander si les usagers ne vont pas être d'ailleurs tentés de n'employer que des termes anglais, voire de n'écrire qu'en anglais afin d'être lus par le plus grand nombre, ce qui n'est pas anodin en termes de rentabilité publicitaire du blog.

Néanmoins ces usages constituent avec les flux RSS une alternative intéressante à la seule recherche via les moteurs de recherche. D'ailleurs, les firmes comme Google et Yahoo surveillent de près les folksonomies et n'hésitent pas à acheter les plus populaires, comme ce fut le cas pour del.icio.us racheté par Yahoo.

Le web sémantique du pauvre ?

Le succès des folksonomies vient sans doute du fait qu'il s'agit, en quel-que sorte, d'un web sémantique du pauvre. Il s'agit aussi d'un système qui change les méthodes de navigation et de recherche, puisqu'il permet d'accéder à des ressources pertinentes qu'un moteur de recherche aurait pu ignorer. Malgré des limites manifestes, les folksonomies mettent

en évidence la possibilité de développer un système plus élaboré, auquel les internautes pourront être tentés de participer. Elles ont aussi le mérite d'exister face aux projets de web sémantique dont les applications concrètes sont encore limitées.

Il est donc probable que nous ne sommes qu'à une étape intermédiaire du système. Les micro-formats, avec X-folk⁵, qui permet de définir des tags dans n'importe quelle page HTML, ou bien encore l'extension pour Mozilla Firefox Tagsy⁶, sont parmi les projets les plus attendus.

Comment éduquer à l'indexation ou la Tag literacy

Une des solutions pour améliorer le système serait de former les utilisateurs à l'indexation par tag. Pour cela, il faudrait aussi que les sites permettent l'usage de plusieurs mots-clés pour définir un document. Mais il faut d'abord comprendre quels sont les mauvais tags.

Nous pouvons établir une liste de mots-clés mal indexés. Nous nous sommes inspirés des travaux du consultant en systèmes d'apprentissage Ulises Mejias [3] pour établir une liste de tags inefficaces : les tags mal orthographiés, les tags mal conçus, notamment les groupes de mots collés ensemble ou séparés sans utiliser l'*underscore*, ceux qui ne respectent pas les règles d'usage ou les conventions, les tags personnels n'ayant aucun intérêt collectif ou encore un tag unique qui n'apparaît qu'une seule fois dans une base de données.

Des règles de bonne indexation par tags ont été édictées :

- l'utilisateur doit penser collectivement : les tags sont certes personnels mais peuvent également être utilisés par d'autres ;
- employer le pluriel pour définir

des catégories. Le pluriel est plus approprié car la catégorie peut contenir différentes variations ;

- ne pas employer de majuscules, à moins que le mot ne puisse être compris sans ;
- utiliser l'*underscore* pour définir un groupe de mots ;
- inclure des synonymes afin d'éviter les confusions ;
- observer et utiliser les conventions d'indexation des sites et des réseaux sociaux utilisés ;
- contribuer à ce que les efforts d'indexation soient efficaces en collaborant et en ajoutant des tags à d'autres ressources.

Outre ces règles, des sites comme Tagyu tentent de mettre de l'ordre, en conseillant l'utilisateur sur les tags à utiliser, par analyse des mots-clés utilisés par d'autres.

**Il s'agit aussi
d'un système
qui change les méthodes
de navigation
et de recherche,
puisque'il permet
d'accéder à des ressources
pertinentes qu'un moteur
de recherche
aurait pu ignorer**

Cet ensemble de règles peut être utile. Toutefois, contribuer à une indexation efficace implique non seulement des compétences mais aussi du temps. C'est ce que souligne Ian Davis, spécialiste en bibliothéconomie⁷. Selon lui, l'indexation par les usagers semble moins coûteuse en

5. Vous pouvez retrouver plus d'informations sur le projet : http://thecomunityengine.com/home/archives/2005/05/xfolk_entry_04.html
6. <http://tagsy.com/>

7. Ian Davis, « Why tagging is expensive », in *Silkworm blog*, http://silkworm.talis.com/blog/archives/2005/09/why_tagging_is.html

4. Outil permettant de vérifier si un serveur répond, et donc s'il est accessible par le réseau.

temps, mais c'est le temps passé à retrouver l'information qui s'accroît, contrairement aux systèmes d'informations hiérarchisées, où le coût d'indexation par les professionnels est plus élevé, mais la recherche d'informations facilitée pour l'utilisateur. Partant de l'adage « *le temps, c'est de l'argent* », il en conclut que le système de tag devient dès lors onéreux.

Il émet même l'hypothèse que, dans cinq ans, les folksonomies seront petit à petit abandonnées pour un retour à la médiation des bibliothécaires et de Google (*sic*). On peut trouver cette vision tout aussi exagérée que celle de Shirky.

Conclusion

Les folksonomies s'inscrivent donc dans une « redocumentarisation » [4] qui voit l'utilisateur conférer un sens au document, à partir de l'usage qu'il souhaite en faire. Cela entraîne des mutations qui peuvent laisser la place aussi bien à une nouvelle sous-plateforme documentaire qu'à des manipulations.

Comme l'écrit Roger T. Pédaque [4], « *les bouleversements induits par les nouveaux usages du web affectent autant la valeur attribuée aux contenus (crédit, autorité, représentativité) que les modes de médiation eux-mêmes (conditions spatio-temporelles de l'interaction, brouillage des rôles et des sphères « public/privé », camouflage des identités, rupture dans les genres, les discours et les usages, etc.) On se trouve désormais confronté à un jeu de rôles dont les règles changeraient en permanence, à l'insu*

même des acteurs qui tirent profit de cette permissivité ou subissent les désagréments des transgressions. »

On peut sans doute parier sur la coexistence des différents systèmes d'indexation, par les utilisateurs et par des professionnels : les deux opérations ne répondent pas toujours aux mêmes objectifs. Les folksonomies présentent de forts avantages pour

**Les folksonomies
présentent de forts
avantages pour l'accès
à de l'information rapide,
voire dans le cas
de la veille, tandis que
l'intérêt de l'indexation
professionnelle demeure,
pour sélectionner
de l'information
à caractère plus durable**

l'accès à de l'information rapide, voire dans le cas de la veille, tandis que l'intérêt de l'indexation professionnelle demeure, pour sélectionner de l'information à caractère plus durable.

Dans ce contexte, les folksonomies peuvent nous permettre d'envisager sur la Toile des systèmes de catalogues hybrides, permettant les deux types d'indexation.

RÉFÉRENCES BIBLIOGRAPHIQUES

1. **MATHES**, Adam. – *Folksonomies. Cooperative Classification and Communication Through Shared Metadata*.
<http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
2. **MARIEKE**, Guy, **TONKIN**, Emma. – « Folksonomies. Tidying up Tags? ». – *Dlib Magazine* Volume 12, Number 1, 2004
<http://www.dlib.org/dlib/january06/guy/01guy.html>
3. **MEJIAS**, Ulises Ali. – *Tag literacy*. – 2005
http://ideant.typepad.com/ulises_mejias.html
4. **PÉDAUQUE**, Roger T. – *Document et modernités*. – 2006
<http://rtp-doc.enssib.fr/IMG/pdf/Pedauque3-V4.pdf>
5. **ROSENFELD**, Louis. – *Folksonomies. How about metadata ecologies*. – 2005
http://louisrosenfeld.com/home/bloug_archive/000330.html
6. **SHIRKY**, Clay. – *Ontology is Overrated; Categories, Links and Tags*. – 2005
http://www.shirky.com/writings/ontology_oversrated.html
7. **SUTTER**, Éric. – « Pour une écologie de l'information ». – *Documentaliste - Sciences de l'information*, vol. 35, n° 2, 1998, p. 83-86.
8. **VANDER WAL**, Thomas. – *Explaining and Showing Broad and Narrow Folksonomies*. – 2005
<http://www.vanderwal.net/random/entrysel.php?blog=1635>

Le projet Europeana 1914 – 1918 : un exemple d'utilisation du crowdsourcing au sein d'une bibliothèque numérique.

Pour Rose Holley, le crowdsourcing est une notion qui se développe au sein des bibliothèques numériques anglophones, elle reste novatrice en France, et ce malgré les travaux et études qui se multiplient sur le sujet. Elle définit la notion comme le fait de confier des tâches normalement effectuées par un employé à un groupe de personnes. Ramené au domaine de la culture, le crowdsourcing s'applique à des projets collaboratifs qui appellent les utilisateurs à intervenir dans le traitement des archives afin de les enrichir.

La définition de Kate Theimer permet de compléter ces informations : les archives participatives peuvent se définir comme « un organisme, un site ou une collection auxquels des personnes qui ne sont pas des professionnels des archives apportent leur connaissance ou ajoutent des contenus, généralement dans un contexte numérique en ligne. Il en résulte une meilleure compréhension des documents d'archives ». Les bibliothèques et centres d'archives sont aujourd'hui des instances privilégiées d'utilisation du crowdsourcing parce qu'elles constituent, en un sens, des laboratoires permettant d'innover dans ce domaine. L'utilisation toujours plus importante du numérique dans l'exercice du métier pousse donc les professionnels à expérimenter dans la gestion du numérique et dans l'accessibilité des contenus au public. Le crowdsourcing permet de répondre à ces nouvelles pratiques. Les initiatives de crowdsourcing au sein du milieu des bibliothèques et archives numériques sont nombreuses et variées. En effet, ce sont des lieux où l'intervention plus ou moins directe des utilisateurs dans le contenu du site (sous la forme de commentaires, de partage ou d'ajout de documents ou d'informations) est acquise. Cet article présente une initiative européenne de crowdsourcing qui incite les utilisateurs à ajouter, commenter et enrichir un fonds en fournissant directement des documents personnels.

Présentation du projet.

La bibliothèque numérique Europeana constitue aujourd'hui la bibliothèque du patrimoine de l'Europe, en rendant accessible en ligne près de 3000 collections de bibliothèques, d'archives, musées et autres. La bibliothèque se place donc comme un véritable outil de diffusion par le numérique du patrimoine. Elle lance en 2011 le projet Europeana 1914 – 1918. Le portail d'Europeana 1914 – 1918 permet d'avoir accès à trois types de documents : des documents déjà détenus par des bibliothèques et des centres d'archives autour du monde, des documents audiovisuels ainsi que des archives privées de familles et personnes relevant de la Première Guerre Mondiale. Ces archives sont notamment désignées sous le terme d'histoire du public et se distinguent des archives dites des collections de bibliothèques et de musées.

Le but de ce projet était, à sa création, de collecter le plus grand nombre de documents sur la Première Guerre Mondiale. Il s'agissait aussi de permettre à des familles de raconter et de témoigner de leur histoire par le biais de leurs documents. Les échanges et les enrichissements des autres collaborateurs et des usagers (chercheurs, passionnés d'histoire ou simple visiteur) étaient également favorisés. L'accès gratuit et libre de ces documents sur internet a donc été rendu possible par leur numérisation.

De grandes journées de collecte ont été lancées en France. Porté par la BnF ce travail préalable incitait les personnes en possession de quelques documents portant sur la guerre (journal de guerre, lettres, photographies, cartes postales ou tout autre objet) à se rendre sur les lieux pour partager leurs documents. Contrairement à des bibliothèques ou des centres d'archives classiques, Europeana ne conserve pas les documents numérisés, qui sont remis à leur propriétaire après la numérisation. Il est surtout possible de déposer soi-même des documents par le biais d'une simple

inscription. Les techniques du crowdsourcing sont donc bien mises en évidence ici et utilisées : l'utilisateur est un « contributeur ». Il peut numériser ses propres documents et les déposer sur le portail en remplissant un simple formulaire qui doit préciser le contexte historique du document. Le crowdsourcing lui permet donc d'identifier, de décrire un document, de laisser des commentaires, et même de le localiser. L'internaute peut donc déposer du contenu scientifique.

Les utilisateurs peuvent y déposer leurs documents en fonction du genre, des sujets (la mémoire, la vie dans les tranchées ou encore les prisonniers de guerre par exemple) ou encore de la zone géographique concernée. Les utilisateurs sont également incités à identifier les documents déposés dans la base et à les décrire. Les documents s'accompagnent donc d'un court texte de présentation rédigé par le contributeur ce qui permet de raconter son contexte historique. Ainsi, le projet Europeana 1914 – 1918 mobilise plusieurs opérations de crowdsourcing qui se complètent telles que la description des documents ou leur identification par les usagers. Le projet fut d'autant plus propulsé sur le devant de la scène en 2013 grâce aux commémorations du centenaire de la 1ère Guerre Mondiale en France, ce qui a ainsi permis aux participants du projet Europeana 1914 – 1918 d'être de véritables acteurs de ces commémorations et d'une mémoire collective.

Réussite du crowdsourcing ?

Le crowdsourcing constitue un outil novateur pour la bibliothèque numérique d'Europeana et pour son projet relatif à la Première Guerre Mondiale. Aujourd'hui, elle compte une importante collaboration des internautes de plusieurs pays qui ont enrichi la bibliothèque de plus de 14 000 histoires et d'environ 181 000 fichiers personnels. Le projet est donc un succès et la participation des usagers est conséquente. Ceci doit être noté car le crowdsourcing reste encore relativement peu utilisé en France, mais demeure présent dans le monde anglo-saxon ou, dans notre exemple, européen. Il convient donc de se demander pourquoi le projet Europeana 1914 – 1918 est une réussite d'utilisation des techniques du crowdsourcing dans une bibliothèque numérique. Rose Holley notait dans son article *Crowdsourcing: How and Why Should Libraries Do It ?* des éléments à respecter afin de bénéficier du crowdsourcing dans les meilleures conditions, autant sur le contenu que sur les modalités de participation des utilisateurs.

Le crowdsourcing doit être interactif. Il place l'utilisateur au centre du projet en l'impliquant au maximum. Les modalités de niveau requis pour participer sont variables. Pour le projet Europeana 1914 – 1918, cela se présente de manière très concrète : une interface simple, sobre et l'absence de prérequis pour y participer (des compétences scientifiques ou historiques par exemple, comme c'est le cas pour d'autres initiatives de crowdsourcing). L'internaute dispose d'une certaine liberté dans ses choix, même si un contrôle est effectué avec l'inscription obligatoire. Il faut noter que la validité et la véracité scientifiques ainsi que historique de ce contenu sont vérifiées par une tierce personne avant la publication. Les enjeux de la fiabilité du contenu du site, et de surcroît de la bibliothèque, sont bien évidemment à prendre en compte.

Ces pratiques participatives mobilisent des souvenirs, des histoires de familles et permettent aussi de présenter des documents inédits, non exploités par la recherche scientifique. Les participants y voient donc un intérêt personnel mais aussi collectif, la satisfaction de contribuer à un projet majeur par exemple, comme celui d'une participation à l'élaboration d'une mémoire collective. Pour Europeana, l'utilisation des méthodes du crowdsourcing permet un enrichissement sans précédent de ses fonds et avec une rapidité importante, un enrichissement qui sans doute n'aurait pas été possible sans la participation massive des utilisateurs non professionnels.

Ce projet est toujours sollicité puisque l'année 2016 sera marquée par le deuxième temps des commémorations de la Grande Guerre en France, qui sera axé sur les batailles militaires de 1916 notamment.

INDEXATION COLLABORATIVE ET FOLKSONOMIES

L'homme a toujours utilisé des systèmes de classification dans les divers domaines de la connaissance. C'est en traçant des frontières, en définissant des catégories d'objets, des familles de plantes ou d'animaux que nous arrivons à mieux comprendre le monde. Nous dessinons des plans, des graphes et des diagrammes pour représenter des idées. Nous avons besoin de tels outils pour trouver notre chemin dans un monde virtuel où les corpus sont de moins en moins identifiants. Ces techniques et ces outils, les professionnels de l'information-documentation les connaissent bien :

- description documentaire selon des structures normalisées ;
- indexation documentaire pour décrire et caractériser un document à l'aide de représentations des concepts contenus dans ce document, mais sans oublier que l'indexation permet d'alimenter des index et donc d'ouvrir un chemin d'accès pour parvenir aux informations relatives aux documents ;
- classifications comme moyens de mettre de l'ordre dans une collection, d'aider voire d'enrichir les recherches des utilisateurs. Les requêtes de type booléen, parce qu'elles ne sont pas contextualisées, trouvent vite leurs limites : bruits importants dans les résultats, quand ce n'est pas aberrations...

Par sa nature même, le Web ne peut exister sans ces techniques documentaires, et ceci est d'autant plus criant que le volume d'informations et de documents potentiellement accessibles croît chaque jour. [...]

1 - Méthodes d'indexation et de classification sur le Web

La classification globale

Avec la multiplication rapide du nombre de documents sur le Web, le besoin de trouver les documents intéressants a suscité le développement de deux types d'outils :

- un outil de classification de documents sous forme de répertoire : Yahoo!, qui est très vite devenu populaire grâce à son système, a profité de la fonctionnalité des liens « HyperText » pour la mise en place de pages de liens contenant des catégories et des sous-catégories. Dans chacune, il a placé des liens vers des URLs ayant un contenu qui lui correspond. Le système de navigation offert par le logiciel et par les liens permettait à l'utilisateur de passer d'une catégorie à l'autre pour trouver les informations recherchées ;
- un outil de recherche en texte intégral : plusieurs sites ont fait leur apparition sur le net pour aider l'utilisateur à trouver, à partir de quelques mots clés, les pages web souhaitées. Nous ne développerons pas ce type d'outils qui sortent du cadre de cet article.

Plusieurs éditeurs de sites ont adopté ce type de catégorisation, avec plus ou moins de succès. Cependant, ce qui est clair dans ce type de démarche, c'est l'aspect subjectif de cette catégorisation. La personne qui analyse un document (ou une URL) décide de le classer sous une ou plusieurs catégories existantes. L'ajout d'une nouvelle catégorie obéit aux mêmes règles que celles qu'appliquent les documentalistes depuis des dizaines d'années.

Au-delà de la classification automatique par les éditeurs, on trouve également ce type de classification sous la forme d'indexation collaborative. Un des premiers projets collaboratifs sur Internet qui ait vu le jour est l'Open Directory Project [voir ci-contre]. Cet annuaire s'est construit et se construit chaque jour grâce à des milliers d'utilisateurs d'Internet dans le monde ; toutefois les catégories proposées ne sont pas libres comme elles le seraient dans une folksonomie. On est ici dans une démarche documentaire plus stricte que l'on retrouve dans les modes de recherche : recherche sur l'URL, sur le titre, sur les descriptifs ou sur les catégories.

2- Propriétés des folksonomies

Les folksonomies prennent donc une place croissante dans l'internet d'aujourd'hui. Sans préjuger de ce qu'elles deviendront dans la durée, il semble intéressant de creuser un peu leurs mises en œuvre et leur efficacité.

Espace de noms plat

Une folksonomie n'est pas structurée. Un terme faisant partie de celle-ci n'a pas de liens avec les autres termes de la même folksonomie (synonymes, génériques, spécifiques, etc.). Cette propriété n'a pas empêché les sites tels que del.icio.us d'aider l'utilisateur en détectant des liens d'une manière automatique, non pas par rapport au sens des mots employés, mais en utilisant les associations existant entre les termes d'indexation. [...]

Liberté de choix des termes

Un terme employé pour indexer un document peut être n'importe quel mot reconnu et accepté par l'utilisateur. Ce mot peut ne pas exister dans le dictionnaire. Mais, du moment où il sert à faciliter la vie de l'utilisateur au niveau de l'organisation de ses pages web, ce terme sera considéré comme valide du point de vue individuel. Précisons cependant que la plupart des internautes prennent l'habitude d'employer des termes qui peuvent être utiles pour les autres personnes qui s'intéressent au même document!

Le principal inconvénient de cette liberté est la possibilité de se trouver, dans sa liste de mots clés, en présence de termes écrits de différentes manières mais ayant le même sens. Plusieurs raisons peuvent être à l'origine de cette multiplicité. Les plus fréquentes sont :

- l'utilisation ou non d'un trait d'union dans les noms composés (exemple : *Open-Source*, *Open Source* ou *OpenSource*) ;
- l'utilisation des initiales (exemple : *IE* ou *Internet Explorer*) ;
- des erreurs de frappe ;
- l'utilisation ou non du pluriel.

Liberté de description

En analysant les étiquettes les plus populaires dans del.icio.us, nous constatons que les utilisateurs choisissent des termes qui ont des fonctions différentes au niveau de l'indexation descriptive du document. [...]

Ambiguïté

Les termes utilisés par une personne peuvent être très clairs pour elle mais plus ou moins ambigus pour la communauté. L'utilisation de plusieurs termes peut réduire cette ambiguïté, et souvent l'éliminer. Une dernière limite de taille est bien sûr que chacun ne rattache pas la même réalité conceptuelle à un même descripteur...

3- Les raisons du succès de l'indexation folksonomique

Les systèmes qui deviennent populaires ne sont pas nécessairement ceux qui sont très bien préparés, analysés et étudiés pendant des années, mais ceux qui répondent le mieux aux besoins de l'utilisateur. Quels sont les éléments qui ont aidé à populariser les systèmes à indexation folksonomique ?

Facilité - Simplicité

Pour indexer un document de cette manière, l'utilisateur n'a pas besoin de comprendre la structure d'un thésaurus, le niveau de spécificité qu'il doit appliquer ni les termes associés auxquels il doit réfléchir. Tout ce qu'il a à faire, c'est de poser les termes auxquels il pense même d'une manière spontanée. L'effort à fournir est donc négligeable et son coût est proche de zéro. Du côté technique, les interfaces homme-machine proposées sont très simples et intuitives pour ne pas décourager l'utilisateur de recourir aux services du site.

Pertinence des termes utilisés

L'indexation folksonomique ne consiste pas uniquement en l'indexation d'un document par un utilisateur pour ses propres besoins. Sa force provient surtout de l'agrégation des termes proposés par les différents utilisateurs et de l'affectation d'une importance supérieure aux termes les plus fréquemment utilisés.

Assistance à l'indexation

Les sites fonctionnant sur ce principe assistent l'utilisateur au moment de l'indexation d'un document. En fait, ils lui proposent les termes les plus populaires. À ce moment-là, trois actions sont possibles.

- Accepter les termes proposés sans aucune modification. Ceci renforce les poids accordés à ces termes.
- Ajouter un ou plusieurs termes. Selon leur rang, les termes ajoutés pourront être pris en compte par le système lors d'une prochaine assistance offerte à un autre utilisateur. S'ils sont acceptés par les nouveaux arrivants, ils pourront prendre de l'importance et feront partie des termes populaires qui catégorisent le document. Sinon, et si la différence continue à augmenter entre les termes populaires et ceux-ci, ils disparaîtront des prochaines propositions, tout en restant disponibles pour l'utilisateur qui les a créés.
- Supprimer un ou plusieurs termes. Le poids des termes supprimés baissera, en comparaison avec les autres termes. Si cette opération est répétée par plusieurs personnes, ces termes finiront par disparaître de la liste des mots proposés à cause de leur faible popularité.

Raisons psychologiques

Plusieurs sentiments de nature positive peuvent pousser une personne à participer à ce type d'expériences.

- Être « vu » sur Internet dépasse le sentiment de jouer un rôle de figurant au cinéma, car c'est la personne elle-même qui décide de l'action à effectuer.
- Voir son avis « accepté » par les autres internautes donne à la personne (consciemment ou non) une confiance en soi.
- Se sentir utile en participant à l'organisation du net encourage l'utilisateur à ne pas hésiter à continuer, d'autant plus qu'il est le premier à profiter directement de la catégorisation d'un document. De la même manière que le Web a « libéré » le lecteur de l'auteur, le laissant libre de son parcours, le lecteur par empirisme a découvert le besoin d'indexation et celle-ci a vu son statut désacralisé.

Possibilité de « parcourir » et de « rechercher »

Les sites utilisant ce type d'indexation offrent les outils nécessaires pour effectuer des recherches sur les termes de la folksonomie. Ceci est surtout utile quand les documents indexés ne contiennent pas du texte (par exemple : les images sur www.flickr.com ou les vidéos sur www.youtube.com). D'un autre côté, la possibilité de « parcourir » les listes de termes en affinant ses choix ressemble à la notion de construction assistée d'une requête en choisissant les critères à la souris. Ainsi, en choisissant un premier terme, le système affiche les termes liés. En sélectionnant un autre, le système l'ajoute à ses critères en arrière-plan et filtre les documents selon les deux critères. Le nombre de niveaux d'affinage dépend surtout de l'outil logiciel utilisé par le site et du nombre de mots clés utilisés pour indexer les différents documents. Un des outils les plus novateurs dans la présentation des mots clés est celui du « nuage de mots-clés » (de l'anglais tag cloud). Il consiste à afficher les n mots clés les plus populaires, triés par ordre alphabétique. Il donne à chacun un style qui dépend de la fréquence de son utilisation pour indexer les documents. Les nuages permettent de rattraper la disparité de l'indexation au sein d'un site.

4- L'indexation collaborative dans les organisations

Il est difficile d'imaginer l'utilisation des sites publics pour l'indexation collaborative dans les organisations. En revanche, l'implémentation d'outils spécifiques à l'enrichissement du contenu pourrait être très utile. Cette tâche consisterait à ajouter des mots clés libres, mais permettant d'accentuer certains aspects du document, ou tout simplement à indiquer, à l'aide d'un système de notation, que tel document est intéressant pour les personnes présentant le même profil que l'utilisateur. Ce type de collaboration ne pourra pas être accepté sans respecter un certain nombre d'exigences :

- l'interface doit être très simple et intuitive pour que le temps passé à l'utiliser soit négligeable ;
- comme sur Internet, l'utilisation doit être fondée sur le volontariat ;
- la sécurité des données doit être préservée à tous les niveaux. En fait, les droits d'accès aux documents par les différents services doivent être appliqués aux termes utilisés pour les catégoriser ;
- pour qu'une folksonomie reste la plus propre possible, il faut conférer l'importance nécessaire aux outils d'assistance à la saisie des mots clés ;
- il ne faut pas délaissier l'indexation professionnelle quand elle existe.

L'indexation collaborative trouvera dans ce contexte une plus grande pertinence parce que les personnes concernées appartiennent aux mêmes services et que leur vocabulaire et leurs besoins sont communs. On connaît bien sûr la difficulté aujourd'hui de mobiliser ses collaborateurs simplement pour déposer leurs productions documentaires dans un système commun, mais nous ne sommes qu'au début de ces pratiques sur Internet et tout reste ouvert. Pour les professionnels, nous rejoindrons ici certains auteurs comme Gilles Balmisse sur l'idée que l'étude de l'écart entre le langage des documentalistes et celui des utilisateurs-contributeurs permettrait d'améliorer la pertinence du service de recherche d'information, mais aussi de mieux identifier les collaborateurs susceptibles d'aider dans tel ou tel domaine et, au-delà, de localiser les experts.

L'indexation collaborative, à la croisée de la qualité et de la désinformation ?

En quelques années, nous sommes partis de pages vierges ou presque de méthadones puis, progressivement, ces méthadones se sont développées sous le contrôle des auteurs et des webmestres pour arriver aux tags attribués par les lecteurs ou les auteurs en toute liberté, et maintenant à la hiérarchisation de ces tags. Les professionnels de l'information-documentation ne seront pas surpris d'une telle évolution, même si souvent ils n'ont pas pu ou pas su faire valoir leur professionnalisme dans ce monde virtuel. Nous ne chercherons pas à trancher ici sur l'identité de ceux qui devraient construire ces folksonomies.

Les utilisateurs finals ont l'avantage d'effectuer une indexation en nombre, à un coût moindre, et d'être représentatifs de la diversité des profils qui naviguent et recherchent. En l'absence de langage documentaire, l'indexation ainsi effectuée pose toutefois des problèmes de cohérence et, si un consensus peut émerger, on est en droit de se poser la question de savoir si les « taggeurs » ne s'influencent pas mutuellement.

Nous souhaiterions toutefois conclure sur une réflexion plus globale et un peu provocatrice sur le passage de la consommation à l'« activisme » sur Internet. L'indexation par l'utilisateur répond à l'atomisation de la production d'information mais, quel que soit le mode d'indexation, il semble aujourd'hui difficile de distinguer ce qui relève de l'opinion commune de ce qui relève de la connaissance scientifique, juridique, historique, etc. Le « Je crois que » ne pourrait-il devenir un nouveau mot clé à s'approprier ? Il nous semble qu'aujourd'hui l'indexation collaborative permet autant de valider la qualité de l'information que de promouvoir la non-qualité.

***Le crowdsourcing appliqué aux archives numériques :
concepts, pratiques et enjeux***

[...] Pour analyser la politique des publics dans les services d'archives, nous pouvons nous appuyer sur l'enquête de la sociologue Laure Ciosi datant de 2012. Son rapport a été publié par le SIAF, Service Interministériel des Archives de France. Cette étude porte sur la politique des publics et ses partenariats fonctionnels dans le réseau des archives municipales, départementales et régionales en France métropolitaine. Parmi les positionnements des services d'archives, nous nous intéresserons plus particulièrement au nouveau rapport au public instauré par Internet. En effet, l'outil numérique est une révolution pour les services d'archives, comme l'atteste ce directeur d'archives départementales :

« C'est vrai que je crois beaucoup plus à l'outil Internet, non pas pour diffuser mieux ce que l'on fait, mais pour radicalement changer le modèle de diffusion des informations ; ce n'est pas faire mieux ce que l'on fait, c'est faire autre chose »

Ce point de vue est intéressant car il montre cette volonté d'exploiter les possibilités offertes par le web. Cette diffusion pose la question de la réutilisation numérique et de l'accès à l'original.

En ce qui concerne les modalités participatives, elle laisse certains archivistes perplexes, comme ce directeur d'archives départementales qui reconnaît ses réticences dans un entretien :

« Ah (les sites participatifs), je n'aime pas ça. Mais ça, mon équipe dit qu'il faudra en faire. Moi, ce que je trouve... si vous voulez, ce qui m'ennuie, c'est de cautionner des choses... où il faut beaucoup contrôler pour être sûr qu'il n'y ait pas de... que les gens n'écrivent pas n'importe quoi (...). Je veux dire que j'ai toujours peur que l'on trompe le public... le public, il est tellement naïf... »

D'autres au contraire voient l'indexation collaborative comme un moyen d'enrichir le fonds, comme ce responsable web :

« Donc on va plus loin que simplement permettre d'indexer les images avec le nom des personnes, la date de l'acte, les types d'actes. On va permettre en fait d'indexer toutes les données archivistiques qu'il y a sur le site, aussi bien les inventaires que les images, et plus loin que seulement avec une date ou un patronyme, un prénom, avec des fonctions de profession. On laissera aussi un champ libre, un champ « commentaire » pour permettre aux gens véritablement d'apporter un plus à une analyse archivistique ou à une notion d'archives participatives – image, une anecdote historique, un complément historique, enfin tout ce que l'on veut »

On constate une ambivalence des points de vue en ce qui concerne les nouveaux usages permis par le web social.

[...] En 2013, un « Etat de l'art en matière de crowdsourcing » a été produit par la BnF, en vue d'un projet de plateforme, et dresse un bilan des pratiques dans les bibliothèques numériques. Pour estimer le degré de participation des services d'archives en France dans les activités de crowdsourcing, nous prendrons comme échantillon représentatif les Archives Départementales. Nous nous appuierons pour cela sur les chiffres fournis par Edouard Bouyé, Directeur des Archives départementales du Cantal. En 2012, il publie un article intitulé « Le web collaboratif dans les services d'archives publics : un pari sur l'intelligence et la motivation des publics » où il recense les services d'archives impliqués dans la dynamique du web collaboratif, tout en dressant un bilan encourageant de ces initiatives.

En 2012, on compte 19 services d'archives départementales proposant un module d'indexation collaborative. [...]

Le Ministère de la Culture et de la Communication s'inscrit depuis 2010 dans une démarche de soutien à l'innovation numérique en encourageant le développement de nouveaux usages culturels numériques. En 2014, 60 projets ont été sélectionnés suite à l'appel à projet SNCI, « Services numériques culturels innovants ». Ces projets impliquent une grande variété

d'acteurs, dont des services d'archives. Parmi ces services innovants, on relève de nombreux projets basés sur le principe de la contribution. Nous avons sélectionné trois projets de type crowdsourcing, dans trois domaines produisant des documents d'archives. Dans le secteur des musées, notons le projet des *Muséonautes*, lancé par un regroupement de 12 musées de Normandie. Il s'agit de créer un outil de partage collaboratif et interactif entre les internautes et les institutions culturelles. En effet, la *Museobase* permet désormais de contribuer à l'enrichissement des données relatives aux images et objets des collections, organisées en parcours thématiques. Au niveau musical, la Fédération des Ensembles Vocaux et Instruments Spécialisés (FEVIS), en collaboration avec 40 ensembles musicaux, propose une ontologie musicale et indexation sémantique à partir d'OMNI, Objets Musicaux Non Identifiés. Enfin, dans le domaine du spectacle vivant, on relève le projet mené par la Scène nationale de Valenciennes : *Rekall-online*. Cette interface permettrait aux internautes d'annoter des captations vidéo de spectacle, agrégeant différents types de documents liés à l'œuvre (photos, croquis, plans, vidéos, sons). Un module de « spectacle enrichi » sur tablette permettrait de diffuser ce travail collaboratif. Ces projets qui verront le jour en 2015 sont donc la preuve d'une dynamique en faveur du crowdsourcing dans les institutions culturelles en général.

[...] En marge des institutions de type GLAM (Galleries-Librairies-Archives-Museums), d'autres projets collaboratifs émergent à travers des initiatives associatives, ministérielles ou bien individuelles.

Geneanet

L'association *Geneanet* rassemble près de deux millions de membres passionnés de généalogie, depuis sa création en 1996. Il s'agit du 6ème site mondial de généalogie. Cet outil collaboratif est gratuit, mais comporte des fonctionnalités réservées aux abonnés *Premium*. Il offre la possibilité de reconstituer l'arbre généalogique de sa famille, en interrogeant une base de données puissante -par nom, lieu, période-, avec un module comparatif. Des « projets contributifs » sont également proposés aux internautes. Pour l'heure, quatre projets sont en cours, classés par difficulté, par période et par localisation. Ils sont détaillés sur *GeneaWiki*, encyclopédie de la généalogie gratuite, éditée de manière coopérative. [...]

Mémoire des Hommes

Le projet *Mémoire des Hommes* a été lancé en novembre 2013 par le Ministère de la Défense, et plus particulièrement par la Direction de la mémoire, du patrimoine et des archives. Il propose un programme d'indexation collaborative de la base des Morts pour la France de la Première guerre mondiale. Le module d'indexation est accessible aux internautes disposant d'un espace personnel. Au niveau des règles de saisie, la visionneuse permet une indexation guidée par des info-bulles. Les champs d'indexation sont les suivants : grade, unité, lieu de naissance, bureau de recrutement, classe, matricule au recrutement, date de décès, lieu de décès (département, pays), transcription du décès (lieu, département). Aucun de ces champs d'indexation n'est obligatoire. En effet, les index sont interrogeables indépendamment les uns des autres. Selon Sandrine Aufray, responsable TIC du ministère, ce projet a été pensé comme un hommage à la mémoire de nos ancêtres, et comme une participation à la politique mémorielle et historique de notre territoire et du conflit. Le bilan est très positif : en mars 2015, 102 205 indexations ont déjà été réalisées, en complément des 47 609 indexations réalisées par le webmestre, soit 10 % de la base, qui compte 1,4 million de fiches. [...]

PhotosNormandie

Le projet *PhotosNormandie* est une initiative privée, qui ne dépend pas des archives départementales. Il est porté par Michel Le Querrec et Patrick Peccatte, auteur du Carnet de recherche *Déjà vu*. Ce projet collaboratif repose sur la plateforme *Flickr*, qui héberge plus de 3000 photos de la bataille de Normandie (6 juin-fin août 1944). La plupart de ces photos proviennent des Archives Nationales des Etats-Unis (NARA) et du Canada. Elles ont été téléchargées librement à partir de la base de données Archives Normandie 1939-1945, du Conseil général de Basse-Normandie, à l'occasion du sixantième anniversaire de la

libération de la région. Ces photos portent la mention « libres de droit » -toutefois, cette formule ne trouve pas d'équivalent dans le droit français-. Depuis mai 2011, un partenariat a été signé avec la Bibliothèque municipale de Cherbourg-Octeville, ce qui a permis d'enrichir le projet de nouvelles photographies, répertoriées dans un classeur intitulé Coriallo. L'idée de ce projet était d'améliorer la légende de ces photographies, qui étaient lacunaires. Depuis le 29 janvier 2007, plus de 7800 descriptions ont pu être complétées grâce au média *Flickr*. Ce projet a reçu une Mention Honorable dans la catégorie *Best use of crowdsourcing for description* du *Best Archives on the Web Awards for 2010*.

Retours d'expérience

Si nous nous plaçons à présent du point de vue des institutions qui ont déjà mis en place ce type de projet collaboratif, quels enseignements pouvons-nous tirer de leur expérience ? [...]

Pour y participer, la quasi intégralité des sites étudiés exigent une inscription préalable, à l'exception des AD d'Alpes-Maritimes -sur *Wikisource*, l'inscription n'est pas obligatoire- et des AD de l'Ain où tout internaute peut ajouter un commentaire sur les photos mystères, en indiquant simplement son e-mail. En revanche, sur le site *Mémoire des Hommes* du Ministère de la Défense, l'inscription est davantage sécurisée puisque la demande d'indexation doit être validée par l'administrateur. Si la participation n'est pas automatique, elle reste néanmoins ouverte à tous. Pour leur module d'indexation collaborative de l'état civil, les AD de l'Ain avaient mis en place un test de paléographie à l'inscription qui s'est finalement révélé inutile. En effet, ceux qui font la démarche de s'inscrire s'avèrent déjà compétents dans le domaine, par conséquent, ce test sera supprimé dans la prochaine version du site. Certains sites proposent également un formulaire pour signaler les erreurs éventuelles, comme les AD de la Vendée, ce qui permet aux contributeurs de se corriger entre eux. En revanche, peu d'institutions ont prévu un dispositif de gratification pour récompenser les contributeurs les plus actifs. La BnF a mis en place un top 6 des correcteurs sur la version bêta de sa plateforme *Correct*. Les AD du Lot-et-Garonne eux, adressent des remerciements personnels aux donateurs de cartes postales. Enfin, le site *Mémoire des Hommes*, qui permet aux contributeurs de consulter le nombre d'annotations réalisées pourrait bientôt leur permettre de se situer dans le classement.

Au niveau de l'administration de ces sites, nous avons pu nous rendre compte à travers ces entretiens que le crowdsourcing n'est pas toujours un gain de temps pour le personnel. En effet, il faut systématiquement vérifier les métadonnées saisies par les internautes et les valider « au fil de l'eau ». Pour Véronique Pouyadou, responsable du fonds iconographique des AD du Lot-et-Garonne, cette étape de vérification représente presque une charge aussi importante qu'auparavant. Dans d'autres institutions, le personnel n'est pas uniquement en charge du projet mais du site dans son ensemble, comme Valéry Vesson, responsable de la maintenance du site internet des AD de l'Ain, d'où un manque de temps. A la BnF, l'équipe dédiée à la plateforme *Correct* est composée de 6 personnes, mais seulement 1 personne est affectée au projet à temps complet. Au Ministère de la Défense, l'ensemble du projet *Mémoire des Hommes* représente l'équivalent de 1,5 ETP (équivalent temps plein). Enfin, le projet autonome semble être celui des AD d'Alpes-Maritimes, qui grâce à leur partenariat avec *Wikimedia* sont davantage encadrés. Mme Cavalié, directrice adjointe des archives ne passe que quelques heures par an, à corriger les pages transcrites. Quant au corpus numérisé, il est mis progressivement en ligne par un conservateur, en suivant les conseils de *Wikisource* pour maîtriser le corpus et achever peu à peu les textes. En parallèle, quelques professionnels souhaitant progresser sur leur temps personnel corrigent régulièrement quelques pages.

Dans l'ensemble, le bilan des institutions est relativement positif, et encourageant pour les projets qui se trouvent dans une phase expérimentale comme la plateforme *Correct* de la BnF. Les données chiffrées fournies par ces institutions confirment les progrès réalisés et l'investissement des publics. Toutefois, ces projets doivent bénéficier d'un suivi sur la longue durée. En effet, le lancement de ce type de projet suscite généralement un enthousiasme au début, dû à l'attrait de la nouveauté, puis connaît parfois un certain essoufflement par la

suite. C'est le cas de la cartothèque des AD de Lot-et-Garonne, qui la première année, en 2010, a reçu 300 cartes postales, puis a atteint une moyenne de 30 dépôts par an en 2015. La responsable du projet cherche donc un moyen de relancer l'intérêt du public, à travers des stratégies comme des concours. De même, les responsables de *SOS détective* aux AD de l'Ain nous ont présenté un bilan mitigé. En effet, sur près de 10 000 images mystères à identifier, seules 5 enquêtes ont été résolues à ce jour. Malgré le très bon référencement du site, classé premier du département, le projet n'arrive pas à trouver son public. Alors qu'il aurait à gagner à diffuser ces photographies sur des médias sociaux tels que *Flickr the Commons*, il se heurte aux questions juridiques posées par le Conseil Général. L'investigation auprès des institutions nous a donc permis de nous rendre compte de la réalité du terrain et des difficultés auxquelles ils sont confrontés : manque de temps, de visibilité du projet, et parfois une certaine frilosité due aux valeurs archivistiques traditionnelles. [...]

Crowdsourcing : tous chercheurs !

En astronomie, en biologie ou encore en écologie, des scientifiques proposent aux citoyens de contribuer à leurs recherches en effectuant des observations ou en accomplissant des tâches simples. Nul besoin d'être expert, une bonne concentration et des yeux aiguisés suffisent.

Vous avez une connexion Internet, un peu de temps devant vous et l'envie de vous investir dans un projet de nature scientifique ? Alors le crowdsourcing est fait pour vous. De fait, lorsqu'il s'agit d'interpréter les millions d'images produites par nos télescopes, de suivre l'évolution des populations d'oiseaux à l'échelle d'un pays ou de déchiffrer les milliers de morceaux d'un papyrus en grec ancien à moitié effacé, toutes les bonnes volontés sont les bienvenues.

« Grâce au Web et à l'existence de bases de données toujours plus performantes, les chercheurs ont la possibilité de mobiliser des milliers de personnes pour collecter de l'information dans le cadre de leurs travaux, explique Mokrane Bouzeghoub, directeur adjoint scientifique à l'Institut des sciences de l'information et de leurs interactions et coordinateur du Défi Mastodons, un programme de recherche sur les masses de données. Ce qui pouvait parfois représenter le travail d'une vie s'en trouve considérablement accéléré. »

De la collecte d'information à la réalisation de tâches

Depuis la fin 2012, le Muséum national d'histoire naturelle (MNHN) fait ainsi appel au grand public pour décrypter les annotations manuscrites portées sur les 6 millions de planches fraîchement numérisées de son herbier historique (date et lieu de collecte, nom du botaniste...). « Avec les seuls moyens du Muséum, nous avons calculé qu'il faudrait employer une personne pendant 500 ans pour mener à bien cette tâche, indique Romain Julliard, spécialiste du crowdsourcing au MNHN. Grâce à la participation des internautes, 100 000 planches ont déjà été indexées dans le cadre du projet Herbonautes. » En 2007, le projet Galaxy Zoo, lancé par une équipe d'astronomes de l'université d'Oxford, a permis de répertorier par forme (ronde, spiralée, allongée) la bagatelle d'un million de galaxies en seulement... deux jours ! Une tâche que ces chercheurs auraient mis des années à accomplir.

Le crowdsourcing, que d'aucuns appellent aussi « science participative », couvre en réalité deux grands champs d'activité. La collecte pure d'information, qui va de l'observation du terrain (comptage de la faune, de la flore...) à la récupération de données concernant l'individu lui-même, comme des données liées à la santé. Et la réalisation de tâches dites de calcul qu'un ordinateur ne saura pas accomplir ou mettra des années à exécuter, autrement dit, des tâches qu'un humain réussira mieux et plus vite qu'une série d'algorithmes ! « Il s'agit généralement de problèmes de combinatoire mettant en jeu de multiples paramètres », précise Mokrane Bouzeghoub.

Ainsi, contrairement aux processeurs, l'homme excelle dans la reconnaissance des formes (caractères, formes géométriques mal définies...) et saura extraire en un clin d'œil d'une photo le slogan tracé sur le mur d'une maison ; il sait également manipuler avec facilité des objets dans un espace virtuel en trois dimensions. Une aubaine pour les scientifiques, qui ont multiplié les projets de science participative ces dix dernières années.

Grâce aux internautes, 100 000 planches ont déjà été indexées dans le cadre du projet Herbonautes.

Un phénomène en plein essor

Dans l'Hexagone, le MNHN, toujours lui, se révèle particulièrement actif sur les projets liés à l'observation. Depuis 2006, pas moins de douze projets de comptage (des oiseaux, des papillons, des chauves-souris...), ont été lancés dans le cadre du programme Vigie Nature : six font appel à des naturalistes avertis (ornithologues ou entomologistes amateurs) et six font participer un public

motivé mais sans connaissances particulières, comme L'Observatoire des papillons des jardins. « Disposer de données à grande échelle, notamment dans les zones urbaines peu observées par les chercheurs, se révèle essentiel dans le cadre du changement climatique et permet d'étudier le déclin ou au contraire l'essor de nombreuses espèces », ajoute Romain Julliard, qui indique que 80 articles scientifiques ont déjà été publiés grâce à ces bases de données collaboratives.

La Bibliothèque nationale de France (BNF) est, elle, bien décidée à mettre à profit notre intelligence graphique : elle s'apprête à lancer avec Orange, le 27 octobre, une ambitieuse plateforme de correction collaborative, Correct1. Objectif : améliorer les fichiers texte associés aux documents numérisés de sa bibliothèque en ligne, Gallica – soit trois millions de livres, fascicules de presse, cartes et autres documents. « Les logiciels de reconnaissance optique de caractères, aussi appelés OCR, sont loin d'être infallibles, explique Arnaud Beaufort, le directeur général adjoint de la BNF chargé du numérique. Pour certains documents, on peut obtenir jusqu'à 99,9 % de reconnaissance de mots sur les pages scannées en haute définition, mais ce taux peut chuter à 60 % pour les documents anciens dont la qualité d'impression laisse à désirer (usure des caractères d'imprimerie, mauvais encrage...). En clair, cela signifie que 40 mots sur 100 dans le fichier texte comportent des erreurs. » Par exemple, un « i » suivi d'un « n » sera transformé en « m » par l'OCR, qui aura aussi une fâcheuse tendance à prendre les « g » pour des « c », une erreur que le cerveau humain ne fera pas. Des expériences similaires ont déjà été menées avec succès par d'autres bibliothèques numériques, celle de la Bibliothèque nationale d'Australie notamment.

Quid de la propriété intellectuelle ?

Diablement séduisante, la démarche de crowdsourcing suscite pourtant le scepticisme : certains, parmi les scientifiques, doutent notamment de la fiabilité des données collectées par des néophytes... « Comme lors d'une expérimentation scientifique classique, le chercheur qui fait appel au public doit mettre en place un protocole solide qui garantira que le processus de recueil de l'information sera le même pour tous les participants, se justifie Romain Julliard. Au Muséum, nous veillons à découper les projets en tâches simples et nous misons sur l'apprentissage. Dans le projet Herbonautes, par exemple, les participants progressent étape après étape : ils retranscrivent d'abord les dates, puis les lieux, et ainsi de suite, en fonction des scores obtenus à chaque étape. »

Le « nettoyage » des données collectées se révèle tout aussi crucial pour la crédibilité des résultats. « Il convient d'être extrêmement vigilant. Certaines occurrences relevées par les chercheurs peuvent être des faux positifs et indiquer une incompréhension des participants, voire l'action d'un logiciel malveillant, avertit Mokrane Bouzeghoub. En fonction des thèmes abordés, des minorités agissantes peuvent aussi être tentées de participer à telle ou telle étude dans le but de biaiser les résultats obtenus. »

La question de la propriété intellectuelle provoque, elle aussi, des réticences. À qui appartiennent les résultats publiés ? Le cas de Foldit risque d'effaroucher plus d'un scientifique : conçu par des chercheurs de l'université de Washington, ce serious game destiné à imaginer la structure en trois dimensions de protéines traditionnellement représentées en deux dimensions a donné lieu, pour la première fois dans l'histoire de la science, à la publication dans Nature d'un article cosigné par un groupe de gamers. C'était en 2010, et ces joueurs venaient de trouver la structure d'une enzyme impliquée chez le singe dans un virus très semblable au sida. « Si la question de la propriété intellectuelle est sensible, car elle conditionne notamment l'obtention de financements, il ne faut pas oublier que ce sont les chercheurs qui posent les questions, montent les protocoles, analysent et interprètent les données », rappelle Mokrane Bouzeghoub. La participation du citoyen, si elle fait avancer la science, n'éclipse en rien le rôle fondamental du chercheur.

Panorama des nouveaux usages des archives audiovisuelles

Depuis l'arrivée d'Internet et la démocratisation des nouveaux supports de diffusion nomades (Smartphones, tablettes...), l'accès aux images d'archives s'est considérablement élargi. Ce phénomène s'est amplifié grâce à l'essor des réseaux sociaux et des sites de partage vidéo au cours de ces dix dernières années. Jadis limités aux postes physiques de consultation et réservés à un public de spécialistes — historiens, chercheurs ou encore généalogistes —, les documents du passé peuvent désormais être consultés, analysés et partagés par tout un chacun sur une multitude de supports offrant des fonctionnalités et usages très variés, allant du simple visionnage sur Internet, à la constitution de dossiers hypermédias, aux ballades virtuelles, en passant par les pratiques de *crowdsourcing*, la chrono-localisation, les archives audiovisuelles « augmentées », etc.

Comment les archives audiovisuelles sont-elles utilisées, valorisées et consultées aujourd'hui ? Quelles sont les nouvelles pratiques de médiation autour des archives audiovisuelles ? De quelle manière les usages ont-ils été modifiés avec l'émergence des technologies numériques, notamment le Web et les réseaux sociaux ? Comment ces nouveaux outils impactent les relations entre les organismes et leurs publics ? Cet article propose un tour d'horizon de ces problématiques, à l'heure où l'image, et notamment l'image d'archives, joue un rôle majeur en tant que vecteur d'information, de communication et de sens.

Des usages multiples

[...]

• Les nouveaux usages des archives audiovisuelles dans les secteurs éducatif et culturel

L'utilisation des images d'archives ne se limite pas au seul univers des médias, mais concerne aujourd'hui une multitude de secteurs, tels que la recherche, le secteur culturel et pédagogique, le tourisme et la communication. En effet, rien de plus simple aujourd'hui que d'avoir accès, via son Smartphone, à des images d'archives relatives au lieu où nous nous trouvons, de composer une « carte postale vidéo » en ligne ou encore de comparer grâce à des applications mobiles les avant/après d'un lieu à partir de photographies ou de films d'archives.

Les collectivités locales et les associations utilisent leur patrimoine audiovisuel principalement pour mettre en valeur le territoire et attirer des visiteurs. La valorisation des contenus patrimoniaux s'effectue aussi bien au travers de banques d'images numériques que de dispositifs multimédias interactifs à base d'images d'archives, comme, par exemple, des modules chronologiques, des parcours thématiques ou des fresques hypermédias. L'Ina a ainsi développé en partenariat avec certaines régions de France des fresques régionales interactives — telle « L'Ouest en mémoire » (Régions Bretagne et Pays de la Loire) — qui retracent en ligne l'histoire récente de certaines régions françaises. Les documents d'archives sont éditorialisés en fonction de grands thèmes (art et culture, économie, sport...), de façon chronologique et assortis d'outils de géolocalisation. Les utilisateurs sont invités à naviguer dans ces parcours thématiques à la découverte des richesses de leurs régions.

Penser le patrimoine non comme une ressource « figée » du passé, mais bien de façon dynamique au service d'une construction métropolitaine à venir, c'est l'ambition du projet patrimonial MétroPôle Images, initié en 2010 par la Région urbaine de Lyon (RUL) avec différents partenaires institutionnels et médias (Grand Lyon, Imaginove, Zodiak France, Ina...). Ce projet consiste à imaginer un dispositif de ressources d'archives audiovisuelles « inédites », composé notamment de films institutionnels sur la Région Rhône-Alpes et de rendre cette mémoire audiovisuelle du territoire accessible au plus grand nombre. Le

dispositif, non finalisé à ce jour, souhaite proposer sur une plateforme Web des images pour le grand public et les chercheurs en sciences humaines. Il cible aussi les professionnels du tourisme et spécialistes du marketing territorial, en leur permettant d'enrichir des supports de visites guidées dans les musées, que ce soit un discours de médiation touristique ou la possibilité par exemple de revivre la transformation de certains quartiers de la ville de Lyon sur Smartphone. Le monde de l'enseignement supérieur est également concerné, avec comme objectif d'inciter les étudiants à créer de nouvelles œuvres audiovisuelles à partir des images d'archives.

La mise en place de jeux ou de quizz autour d'images d'archives prend également de plus en plus d'ampleur, grâce au succès de réseaux sociaux comme Facebook et Twitter. Du quizz aux applications virales, le jeu permet à la fois de toucher un public nouveau, d'engager l'internaute et de « désacraliser » les contenus d'archives. Traiter les archives sous l'angle ludique implique souvent de détourner l'archive de son contexte et de sa valeur originelle. Mais l'approche différente qu'offre le jeu est aussi un moyen de toucher un public peu intéressé ou peu informé sur les archives. L'approche ludique peut permettre de revisiter une archive, de la voir sous un autre jour. Il tient aux organismes de mettre en place des dispositifs annexes permettant de poursuivre sa découverte, de recontextualiser l'archive, pour transformer la plaisir du jeu en intérêt patrimonial.

Bien que ce dispositif ne s'adresse pas spécifiquement aux archives audiovisuelles, il est intéressant ici de citer l'exemple du jeu « *Le Mystère de la Cordelière* » inventé en 2012 par les Archives départementales de l'Aube, dont l'objectif pour le joueur est de sauver un texte confisqué par les « archipages », des créatures friandes de documents d'archives. Plusieurs énigmes doivent être résolues, en fouillant les documents d'archives présentés. La résolution de l'ensemble des énigmes donne le droit de consulter le fameux texte volé.

Les vertus pédagogiques et éducatives des images d'archives sont depuis longtemps connues par les organisations patrimoniales, en plus des traditionnelles expositions et projections. Citons ici l'exemple du portail « Le Sénat junior », développé par le Sénat sur son site internet. L'objectif de ce portail est de favoriser, sous un angle ludique, l'apprentissage du système institutionnel français auprès des jeunes générations, à l'aide de documents d'archives. Une vingtaine de jeux et outils pédagogiques sont donc proposés en ligne, portant sur diverses thématiques (découverte de l'hémicycle, le Palais du Luxembourg en images, les sénateurs...). La « Machine à remonter le temps » permet aux utilisateurs de découvrir l'histoire du Sénat à travers différentes époques. Quelques photographies d'archives sont publiées, pour illustrer les biographies des présidents du Sénat.

[...]

L'ECPAD (L'Établissement de communication et de production audiovisuelle de la Défense), l'organisme français dépositaire du patrimoine audiovisuel et cinématographique de l'armée française depuis 1915, propose quant à lui diverses actions pédagogiques, culturelles et scientifiques sur mesure, à destination des scolaires, universitaires et chercheurs, telles que des visites découvertes, des ateliers d'« analyse d'images », des formations ou encore des parcours pédagogiques en ligne pour les enseignants.

La relation avec l'utilisateur transformée par le nouvel environnement numérique collaboratif

Les nouvelles plateformes numériques et les usages qui leurs sont liés ont multiplié les projets dits de *crowdsourcing* et les échanges collaboratifs entre les organismes patrimoniaux et leurs publics. Ces projets bouleversent aujourd'hui ces relations avec les usagers.

• L'utilisateur en tant qu'expert

La posture « d'expert » des professionnels des archives n'est plus exclusive. Elle est désormais partagée avec les particuliers passionnés et parfois avec le grand public, qui détient certains savoirs spécifiques. Cette collaboration avec les usagers amateurs a toujours existé grâce aux bénévoles ; mais les outils du Web permettent aujourd'hui de

réaliser ces projets collaboratifs à très grande échelle et d'optimiser (en augmentant ses chances) les retours des contributeurs.

- **L'aide à l'identification**

À titre d'exemple, les sollicitations les plus communes lancées par les organismes patrimoniaux sur le Web sont les demandes d'identification ou d'aide à la localisation d'un document ou d'une collection. « Qui est sur cette photographie ? », « Savez-vous situer cette prise de vues ? » sont les questions les plus fréquemment posées.

Ces sollicitations sont de plus en plus relayées sur les réseaux sociaux, et notamment sur Twitter et Facebook. Les réseaux sociaux ont l'avantage d'être par essence des plateformes d'échanges, sur lesquelles les internautes sont déjà prédisposés à dialoguer. De plus, il est plus facile de poser une question à ses abonnés sur les réseaux sociaux (en « push », qui apparaîtra sur sa *timeline*) plutôt que d'attendre leur visite sur son site.

Les archives départementales de la Manche ont lancé un dispositif sur Facebook. Ils conversent avec leurs fans pour localiser certaines photographies. Pour valider ou invalider une proposition d'un internaute, l'établissement poste une autre photographie du lieu proposé et pointe les différences ou les similarités ; une façon originale d'échanger avec leurs fans tout en valorisant leurs archives et en faisant avancer l'enquête.

Les Archives de la télévision nationale irlandaise (RTE) ont choisi de relayer leurs appels à identification via Twitter. À raison d'environ une demande par semaine, ce dispositif s'est avéré efficace, la plupart des énigmes ayant été résolues. Certains appels suscitent des échanges entre RTE, ses partenaires et ses abonnés (qui sollicitent leurs réseaux pour mieux mener l'enquête).

Certains organismes ont lancé des appels à reconnaissance de personnes sur leurs sites dans un objectif plus mémoriel, pour raviver le débat entre générations, ou solliciter des témoins âgés.

« Remember Me? » est une opération lancée par le Musée de l'holocauste des États-Unis (United States Holocaust Memorial Museum), dans le but d'identifier les enfants apparaissant sur les clichés pris par les associations humanitaires, après la seconde guerre mondiale. Outre l'identification des personnes photographiées, le projet vise à faire connaître l'histoire de ces enfants orphelins qui ont été déplacés pendant la guerre et après. Ce projet, guidé par un impératif de temps, souhaitait également donner une chance à ces enfants de renouer des liens avec d'éventuels membres éloignés de leur famille. En mettant en place une plateforme Web efficace et une communication axée sur des réseaux en ligne ciblés, le projet, initié en 2011, a su tirer profit des opportunités qu'offre le Web.

- **Le contributeur collaborateur**

Les nouvelles plateformes permettent également aux contributeurs d'aller plus loin en participant, par exemple, à l'effort d'écriture ou de réécritures de descriptions rattachées à des documents. Le projet « PhotosNormandie » repose sur la mise en ligne sur la plateforme Flickr de plus de 3 000 photos de la bataille de Normandie issues des Archives nationales des États-Unis (National Archives and Records Administration / Nara), des Archives nationales canadiennes (Bibliothèque et Archives Canada) et de la Bibliothèque de Cherbourg. L'objectif de ce projet est d'améliorer les descriptions souvent très succinctes, voire inexistantes, des photos disponibles en ligne et de relier les contenus à d'autres bases accessibles en ligne. Il est mené par une équipe indépendante de toute institution. Le travail de redocumentarisation est réalisé par les internautes.

Le *crowdsourcing* vise souvent des passionnés - « experts » prêts à apporter un savoir plus scientifique sur un sujet précis. Que ce soit via un site dédié ou un wiki, les organismes patrimoniaux mettent ainsi parfois en place des plateformes d'échanges plus « scientifiques » avec leurs usagers les plus avertis.

Les Archives de la Vendée ont créé un « L@boratoire des internautes ». Son objectif est de recréer en ligne les échanges que pratiquent chercheurs, experts et professionnels dans les salles de consultation. Les contributeurs peuvent rédiger des billets, les commenter, réagir aux questions posées par les Archives (incompréhension du contexte d'un objet ou document, termes énigmatiques inscrits sur des documents...). Le « L@boratoire »

augmente ainsi l'intérêt des archives auprès des lecteurs-internautes les plus engagés, et favorise le partage des connaissances avec les archivistes. Les Archives de la Vendée co-construisent également avec des bénévoles une base collaborative de noms : « Noms de Vendée ». Son logiciel support, initialement porté par un groupe de passionnés, a été donné aux Archives de la Vendée pour assurer une pérennité et plus d'ampleur à la base. Aujourd'hui, la base est co-administrée par Les Archives de la Vendée et ce groupe de bénévoles. « Nous avons voulu pousser au plus loin la logique de collaboration : les bénévoles ne sont pas que des fournisseurs de données, car même l'administration intellectuelle et technique de la base est collaborative », explique Thierry Heckmann, directeur des Archives de la Vendée.

Cette intégration des amateurs dans la constitution des bases n'est pas sans conséquence sur la façon dont les professionnels abordent les nouvelles plateformes d'échange et sur leur capacité à accepter ces stratégies collaboratives. S'ils sont enrichissants, utiles, innovants et parfois nécessaires, les projets collaboratifs peuvent être également déstabilisants pour un organisme. Ils sont l'engagement d'une relation nouvelle avec les usagers, où l'organisme accepte de lâcher prise, de faire confiance à des non-professionnels et de leur ouvrir ses portes et ses contenus.

- **L'utilisateur en tant que producteur de contenus**

Les usagers peuvent aussi être sollicités pour soutenir le travail des professionnels en donnant de leur temps sur des tâches simples mais chronophages. Ces projets de *crowdsourcing* participent à la tendance actuelle de rapprocher usagers et professionnels, en promouvant la même envie d'accomplissement d'une mission. Les formes les plus populaires de *crowdsourcing* dans le domaine des archives sont la transcription et le *tagging* collaboratifs. La transcription (le dictionnaire Larousse dit « reproduire un mot, un texte, à l'aide d'un système d'écriture différent », en l'occurrence ici passer de l'écriture manuscrite à l'écriture numérique) concerne principalement les documents écrits. Le *tagging* – ou identification de mots clés – s'adapte quant à lui à tous les supports : documentation écrite, images fixes ou images animées.

Les projets de *tagging* collaboratif répondent à de nouveaux besoins créés par l'essor du numérique. L'accessibilité des contenus en ligne est devenue une priorité pour de nombreux organismes patrimoniaux, or, le Web demande un niveau de granularité de l'indexation élevé. Pour que les contenus numérisés soient accessibles et remontent correctement dans les moteurs de recherche, l'indexation doit être à la fois précise et abondante. Sans indexation et sans métadonnées approfondies, les contenus n'ont pas de valeur d'usage et ne sont ni consultables, ni exploitables. Mais cette tâche prend du temps, et faire appel au public pour créer ou compléter l'indexation est un moyen d'accélérer la mise en ligne.

Le *tagging* s'inscrit, de plus, totalement dans les nouveaux usages liés au Web. En effet, le *tagging* des amis sur Facebook, des vidéos sur YouTube, des photographies sur Flickr ou les hashtags sur Twitter sont tout autant de façons de pratiquer la « folksonomie ».

Pour faciliter le *tagging*, certains organismes ont mis en place des systèmes de correction de tags générés automatiquement. BBC Service Radio Archive propose un service hybride, dans lequel le contributeur vérifie les tags créés automatiquement et peut en générer de nouveaux.

La mise en place de projets d'indexation collaborative vise également à éliminer ou réduire l'écart sémantique entre le vocabulaire des professionnels et celui des visiteurs « lambda », pour être au plus près des besoins des usagers. Différents niveaux de *tagging* peuvent également être appliqués, en fonction de l'expertise du contributeur.

Pour le projet « Your Paintings Tagger », initiative de la Public Catalogue Foundation, de la BBC et de musées britanniques, c'est lors de l'inscription qu'il est demandé au contributeur s'il a des compétences particulières en histoire de l'art. Si oui, le contributeur est invité à participer à un travail de *tagging* plus spécialisé (*tag* du mouvement artistique associé à une œuvre, etc.).

Cette expérience de *tagging* peut aussi être déclinée en jeu pour motiver les contributeurs.

En 2009, l'Institut néerlandais pour l'image et le son (Beeld En Geluid) a souhaité tester cette « gamification » en mettant en place dans le cadre d'un projet de recherche un jeu d'annotation vidéo appelé « Waisda? ». Des extraits d'émissions de télévision étaient présentés aux joueurs qui devaient proposer des *tags* adaptés à ce qu'ils voyaient. Chaque jeu fonctionnait en duo : les joueurs remportaient des points lorsque leurs *tags* correspondaient avec ceux de leur adversaire.

• **Nouvelles formes de collaboration et de valorisation autour des fonds patrimoniaux**

L'arrivée du Web a également modifié les dispositifs de collecte mis en place par les organismes patrimoniaux, notamment les cinémathèques, pour accueillir et surtout valoriser des fonds privés. De nombreux sites Web de collecte de souvenirs ou de documents d'archives issus de particuliers font ainsi surface à l'occasion d'anniversaires ou dans le cadre de projets mémoriaux. Ces plateformes Web, si elles sont souvent aussi accompagnées de campagnes de collecte hors ligne, ont l'avantage de toucher un public plus large et de permettre une valorisation quasi immédiate et accessible par tous des contenus collectés.

Les projets « Europeana 1914-1918 » et « Europeana 1989 » sont des exemples intéressants. Ils ont pour but de recueillir et de valoriser auprès du public européen des histoires et des documents de toutes provenances en lien avec la première guerre mondiale et la chute du mur de Berlin. Le dispositif s'articule entre des collectes physiques et un recueil en ligne d'objets et d'histoires.

De même, le gouvernement de Singapour, en partenariat avec de nombreux acteurs locaux, a lancé en 2011 le « Singapore Memory Project » (SMP). Ce projet de collecte — mené à l'occasion du 50^e anniversaire de l'indépendance de Singapour — a pour objectif de constituer une collection nationale de souvenirs et de moments importants des acteurs de la ville (habitants, organisations, associations, entreprises, institutions...) et de les rendre accessible au plus grand nombre afin de construire pour les générations futures une base représentative du chemin parcouru par leur nation.

Toutes ces nouvelles formes de collaboration et de valorisation autour des fonds audiovisuels et photographiques mises en place par les organismes patrimoniaux ne cessent de se renouveler. L'évolution permanente des comportements liés au Web et aux réseaux sociaux, l'apparition de nouvelles plateformes d'échanges, l'explosion du trafic sur les sites vidéo et l'arrivée de nouveaux supports de consultation offrent chaque jour de nouvelles possibilités d'exploitation des archives. Les organismes patrimoniaux et les détenteurs de fonds doivent toujours plus s'intéresser à ces nouvelles pratiques pour être au plus près des attentes de leurs publics, tout en étendant leur notoriété auprès de nouveaux publics jusqu'alors peu familiers avec les documents d'archives.