

Le web collaboratif dans les services d'archives publics : un pari sur l'intelligence et la motivation des publics

Édouard BOUYÉ

Néologismes et mots-valises

Le travail collaboratif est consubstantiel à la République des lettres. Les grandes entreprises savantes, depuis le Moyen Âge mais surtout l'époque moderne, sont collectives. Les abbayes mauristes et les académies depuis le XVIII^e siècle, les sociétés savantes et les établissements d'enseignement supérieurs depuis le XIX^e siècle sont les lieux privilégiés de cette collaboration, dont les archivistes sont traditionnellement les animateurs, ou du moins les médiateurs. L'avènement du numérique, utilisé de manière privilégié par notre profession pour rendre accessibles des millions de pages d'archives, permet aussi de donner un souffle nouveau aux entreprises collaboratives. La nouveauté réside dans l'ampleur des acteurs concernés et dans l'ampleur du travail qui est mis en commun. Il ne s'agit plus seulement, en effet, d'un cénacle limité de savants œuvrant à une entreprise érudite commune débouchant sur la publication, à un petit nombre d'exemplaires, de volumes imposants. L'archiviste qui anime le web 2.0 doit se déprendre d'un modèle « top-down », où l'utilisateur est tributaire de l'inventaire d'archives imprimé « sub specie aeternitatis ». Pour définir le modèle nouveau de coproduction du savoir, les Anglo-saxons ont forgé des néologismes en forme de mots-valises, dont l'étrangeté même dit la nouveauté des réalités qu'ils recouvrent. La « folksonomie, ou indexation personnelle, est un système de classification collaborative décentralisée spontanée, basé sur une indexation effectuée par des non-spécialistes »¹. La « folksonomie est, en quelque sorte, un exemple de « crowdsourcing (en français, externalisation ouverte), qui est l'un des

¹ <http://fr.wikipedia.org/wiki/Folksonomie> (8 février 2012).

domaines émergents du management de la connaissance : c'est le fait d'utiliser la créativité, l'intelligence et le savoir-faire d'un grand nombre de personnes (des internautes en général), en sous-traitance, pour réaliser certaines tâches traditionnellement effectuées par un employé ou un entrepreneur. Ceci se fait par un appel ciblé (quand un niveau minimal d'expertise est nécessaire) ou par un appel ouvert à d'autres acteurs (...). Le terme crowdsourcing est un néologisme sémantiquement calqué sur l'outsourcing (externalisation). La traduction littérale de crowdsourcing est approvisionnement par la foule, ou par un grand nombre [de personnes] », mais l'expression ne reflète pas vraiment le sens anglo-saxon du terme ; « Impartition à grande échelle » ou encore « externalisation distribuée à grande échelle » sont d'autres traductions plus précises »¹. Ces deux définitions sont elles-mêmes tirées du site internet qui pratique le « crowdsourcing » à une échelle mondiale ; appliquées au domaine des archives, sources primaires de la connaissance, elles montrent le changement de la relation entre l'archiviste et l'utilisateur. Ce dernier n'est plus seulement un consommateur, mais un acteur ; ou, pour mieux dire, en utilisant un mot-valise français cette fois, un « consommateur » de la connaissance, qui enrichit volontairement l'instrument de recherche ou le document d'archives numérisé lui-même.

Le degré de collaboration proposé à l'internaute par l'archiviste est défini par ce dernier. Si l'indexation collaborative n'implique pas de division fixe des tâches (chacun étant libre de poser des signets où il l'entend), l'indexation coopérative implique un partage des tâches (l'archiviste distribuant des lots d'images à indexer selon un protocole et dans un délai fixé au départ). L'indexation est plus ou moins riche, suivant que le masque de saisie est simple ou complexe : quand l'indexation de l'état-civil ne comprend que la possibilité d'indexer une année, un nom et un prénom (à l'exclusion de mention de filiation, de sexe, d'âge, de métier), on parle d'indexation pauvre, ou d'annotation collaborative. Le résultat n'est pas une base de données nominative au sens où l'entend la CNIL permettant d'identifier à coup sûr un individu, mais des points de repères chronologiques (années) et un réservoir de noms formant un outil d'aide à la recherche historique.

État des lieux

¹ <http://fr.wikipedia.org/wiki/Crowdsourcing> (8 février 2012)

Qui, quoi, avec qui ?

Sont délibérément exclus de cette étude les systèmes collaboratifs ne reposant pas exclusivement sur internet, tel que l'intégration de bases générées sur un tableur (Ain, Vendée), l'indexation faite par ou pour le service public (Besançon, Bas-Rhin) ou la mise à la disposition sur les sites web des Archives de dépouillements réalisés par des associations de généalogistes (Lyon, Hautes-Alpes). Le travail collaboratif sur internet implique l'écriture ubiquitaire, démonétisée et asynchrone, par l'internaute, dans les bases encodées par le service public. C'est là que réside le pari, qui fait l'objet de débats dans les services et qui est encadré par des garanties scientifiques et techniques inversement proportionnelles à l'audace de l'archiviste, lequel a le choix entre trois solutions :

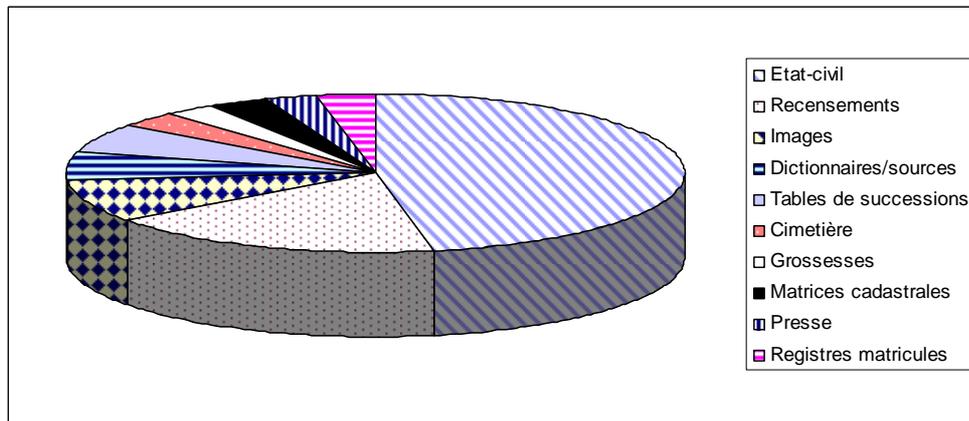
- Liberté totale laissée à l'internaute dispensé d'inscription préalable (Cantal, Corrèze, Eure-et-Loir, Loire-Atlantique, Nièvre, Var) ;
- Liberté totale laissée à l'internaute après inscription obligatoire et/ou ouverture d'un compte (Martinique, Hautes-Alpes, Rhône, Yvelines) ;
- Encadrement de l'internaute qui doit s'inscrire, passer des tests de paléographie et travailler sur des lots choisis par les Archives (Rennes, Aube, Mayenne, Puy-de-Dôme).

La liberté entraîne des risques de malveillance (signets injurieux, délibérément erronés) et d'erreurs. Elle oblige à un système de régulation par le service public. La vérification de chaque signet implique une charge de travail que tous les services ne peuvent se permettre. Un système plus léger, qui part du principe qu'une erreur d'indexation n'est vraiment gênante que lorsqu'elle est démasquée, consiste à autoriser aux internautes, outre la pose de signets, à envoyer un message d'erreur sans l'autoriser à corriger l'erreur : l'agent qui reçoit le message d'erreur arbitre en faveur de l'un ou l'autre des index. Dans le Cantal arrivent une moyenne de 5 messages d'erreur par jour, sur une moyenne de 740 signets posés chaque jour (soit environ 0,7 % d'erreur). Trois agents sont préposés à ce travail ; il s'en trouve toujours un présent pour effectuer les corrections en flux tendu (c'est-à-dire le premier jour ouvrable suivant l'envoi du message d'erreur). Ce travail de régulation représente moins de 0,1 ETP pour le service.

Trente services d'archives publiques proposent du web collaboratif, dont dix ont mis des images sur Flickr ou d'autres systèmes analogues (pour les partager

et, surtout, les identifier¹). Si l'on restreint notre comptage aux systèmes développés par les services, on obtient deux services d'archives communales (Nantes et Rennes) et dix-neuf services d'archives départementales (Ain, Hautes-Alpes, Alpes-de-Haute-Provence, Alpes-Maritimes, Aube, Cantal, Corrèze, Eure-et-Loir, Hérault, Loire-Atlantique, Lot-et-Garonne, Mayenne, Nièvre, Puy-de-Dôme, Rhône, Yvelines, Var, Vendée et Martinique).

Fonds faisant l'objet d'indexation collaborative en janvier 2012



Ces vingt-et-un services proposent l'indexation collaborative pour un ou plusieurs de leurs fonds, ce qui porte à trente-quatre le nombre de fonds en ligne ainsi enrichis. L'état-civil s'y taille la part du lion (seize fonds, soit près de 50 %), suivi par les listes nominatives de recensement (six fonds, soit 18 %) et les images (trois fonds, soit 9 %). Les autres fonds concernés sont les tables de successions et absences (Ain et Corrèze), les registres des inhumations (Nantes), les déclarations de grossesses (Nièvre, XVIII^e siècle), les matrices cadastrales (Var), la presse ancienne (Yvelines) et les registres d'incorporation militaire de Seine-et-Oise (Yvelines).

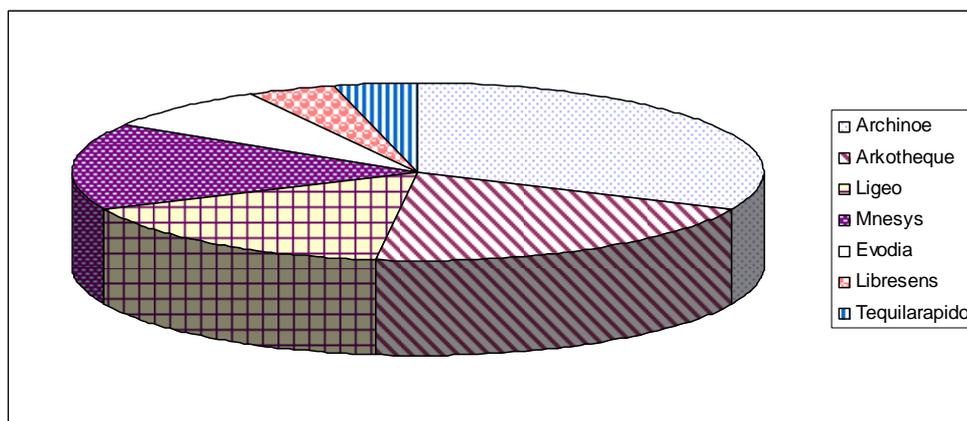
Au-delà de la simple de pose de signets (« tags »), il faut mettre à part la Vendée, avec son L@boratoire et ses dictionnaires collaboratifs², les Alpes-Maritimes, avec leurs éditions collaboratives de sources (sur Wikisource) et le Lot-et-Garonne, où les internautes sont invités à déposer des cartes postales

¹ C'est le cas, par exemple, de la Haute-Garonne, qui met en ligne sur un blog intégré à son site des photographies du grand sud-ouest et d'Espagne, ou de l'Ain, dont le module « SOS détectives » propose d'identifier des photographies de l'est de la France, du Jura à la Côte d'azur. Facebook est aussi utilisé à cette fin d'identification iconographique ; cf. la contribution de Jérémy Hallais, p. – du présent volume.

² Cf. la contribution d'Emmanuelle Roy, p. - du présent volume.

numérisées en 300 dpi, assez anciennes pour ne pas être grevées de droit d'auteur à condition qu'elles ne figurent pas encore dans les collections. Ces deux exemples ouvrent une voie prometteuse, invitant services d'archives et opérateurs à imaginer ensemble des systèmes qui automatisent l'entrée dans le web 2.0 d'enrichissements qui y sont intégrés aujourd'hui manuellement, comme c'est le cas, entre autres, dans l'Hérault ou en Gironde¹.

Systèmes d'indexation collaborative utilisés par les services d'archives en janvier 2012



Le marché du web collaboratif dans les archives est réparti entre quatre principaux opérateurs. Les vingt-et-un services utilisent vingt-cinq opérateurs (car quatre utilisent deux opérateurs couplés). Outre Evodia (société maintenant disparue), Libresens (société agenaise qui développe l'application du Lot-et-Garonne consacrée au cartes postales) et Tequilarapido (société niçoise et parisienne employée par les Alpes-Maritimes), le marché est ainsi structuré : Archinoé (société Archimaine, opérateur « historique », 8 services, soit un tiers du marché) ; Arkothèque (société 1égale2, 5 services) ; Ligeo (société V-Technologies, 4 services) ; Mnesys (société Naoned Systèmes, 4 services).

Résultats

Au 31 janvier 2012, 2.487.000 signets avaient été posés, dont 683.000 dans les seuls deux services d'archives communales (405.000 pour Nantes et 278.000 pour Rennes), et 1.804.000 pour les dix-neuf services d'archives départementales (l'Aube et le Cantal dépassant chacun le seuil des 400.000

¹ Cf. l'exemple de Christophe Grosvallet, p. – du présent volume.

signets posés). Si l'on veut comparer les chiffres des indexations réalisées par ou pour les services d'archives avec ceux des projets privés (associatifs ou commerciaux, collaboratifs ou non, sur le web ou non), il faut néanmoins ajouter que, par exemple, les bases de l'Ain proposent 2.200.000 noms et celles de la Vendée, 1.500.000 noms.

Lorsque l'on examine les conditions de la réussite d'un web collaboratif, on s'aperçoit que ni le choix du système, ni le caractère plus ou moins intuitif ou encadré du système, ni le degré d'anonymat de l'internaute ne semblent véritablement jouer. En revanche, il faut souligner le succès des deux projets d'archives communales, très certainement dû aux importantes médiations et médiatisations ayant entouré les projets, au caractère très « geek » de ces deux métropoles bretonnes et à l'extraordinaire vogue de la généalogie en Bretagne – y compris dans la capitale des ducs. L'exemple de Nantes et de Rennes souligne la nécessité de la médiation numérique : tournées pédagogiques et démonstration sur le territoire et au-delà (car les « émigrés » bretons et auvergnats, notamment de Paris, sont très friands de leurs racines généalogiques), création d'une émulation, réponses personnalisées aux questions des internautes indexeurs (en salle de lecture, au téléphone et par courriel, dans un climat et avec un ton cordial malgré le ton éventuellement peu amène des interpellations¹). Les projets, enfin, ne décollent véritablement qu'après la constitution d'une masse critique de signets : on a plus envie d'alimenter une base lorsque l'on y a déjà trouvé son patronyme. Ce cercle vertueux permet aussi, surtout si le système est ouvert et non point borné à un « happy few » d'experts dûment sélectionnés par un test, de surmonter le problème de l'essoufflement du projet, une fois retombé l'enthousiasme des « early users ».

Quelle « valorisation » ?

L'acception du terme de « valorisation » des archives est restée claire jusqu'à ce que les succès éclatants de la diffusion numérique du patrimoine écrit par les services français d'archives n'attisent les convoitises des chercheurs de « poule aux d'or », selon l'expression d'Hervé Lemoine. Si l'interconnexion des dépouillements est techniquement possible, est-il souhaitable ? L'un des opérateurs du marché, Archimaine, propose une plate-forme d'interrogation des bases de données cumulées de ses clients ; la base Nomina du SIAF agrège

¹ Ma formule de politesse rituelle est, en pareil cas : « Merci de votre patience, de vos encouragements et de votre compréhension. »

aussi des dépouillements de diverses provenances ; l'ambition des sociétés de généalogie commerciale est d'agréger les dépouillements achetés aux associations de généalogistes.

La question de la propriété intellectuelle des bases obtenues par la voie de l'indexation collaborative se pose. J'ai moi-même été sollicité par une société privée pour savoir à quelles conditions la base constituée par les internautes fréquentant le site du Cantal pourrait être cédée. Il m'a semblé difficile, indépendamment de la question financière, d'accéder à cette demande, puisque cette base est le fruit de l'intelligence et du travail collectifs.

La voie qui semble mieux balisée est celle du moissonnage (protocole OAI-PMH) : chaque service producteur demeure responsable de sa base, mais celle-ci peut être moissonnée par un ou plusieurs portails.

Les « permaliens », un système ouvert d'indexation collaborative

Pour les services qui ne souhaitent ou ne peuvent pas se lancer dans l'indexation collaborative, il existe une alternative plus légère : le « permalien ». Un lien permanent, ou identifiant pérenne, est une adresse URL stable et pérenne ; généré automatiquement par le système de mise en ligne, sa permanence dans le temps est garantie. Chaque cote d'archives décrite et chaque lot d'images et/ou image disponible sur les sites dotés de cette fonctionnalité sont dotés d'un lien permanent. Plusieurs systèmes existent. Le système ARK a été choisi, entre autres, par la Bibliothèque nationale de France ainsi que par les Archives départementales du Cantal, du Rhône et de la Somme¹.

Le permalien répond aux besoins de la citation en terme de stabilité et de pérennité ; dans le contexte actuel des besoins et des obligations en matière de réutilisation des archives publiques contenant des données personnelles, le permalien permet de concilier des exigences parfois vues comme contradictoires ; il est un outil privilégié d'indexation collaborative.

À l'indexeur, il donne des garanties sur les sources utilisées et sur le contexte où se trouve l'information. Pour le service d'archives publiques, le permalien, tout en permettant une réutilisation, garantit que l'image ne sera pas dénaturée

¹ Sur ce sujet, cf. par exemple http://fr.wikipedia.org/wiki/Archival_Resource_Key.

ou altérée par le réutilisateur, qu'elle ne voisinerait pas avec des contenus jugés inadéquats, qu'elle ne sera pas piratée au gré des transferts et des copies dans divers pays, que son accès ne sera pas rendu payant (seul l'accès aux index du réutilisateur étant éventuellement payant) ; le service d'archives ne peut être tenu co-responsable, en ayant fourni des fichiers numériques, d'une réutilisation erronée, non conforme ou délictueuse des données publiques.

L'indexeur peut quant à lui procéder à l'indexation qu'il souhaite, dans le respect des lois et règlements en vigueur et sous le contrôle de la CNIL ; il lui suffit de pointer ses index sur les liens pérennes qu'il aura au préalable copiés sur le site.

Comme l'indique Jordi Navarro dans son blog *Papiers et poussières*, en un article intitulé *Valoriser le patrimoine : entre bien public et intérêts privés*¹, publié le 6 août 2011 : « Si héberger et diffuser gratuitement les images représente effectivement un coût non supportable par l'opérateur, il peut choisir de ne pas les héberger. Les résultats de recherche dans son index pointerait alors vers les images hébergées par les services d'archives. L'investissement lié à la réalisation de ce projet s'en trouverait d'ailleurs fortement allégé. Cela réglerait également le problème du masquage et de l'anonymisation des données personnelles sensibles.

Évidemment, cela impose, en retour, aux services d'archives d'adapter techniquement leurs sites afin de permettre cette indexation extérieure. Cela passerait donc par une coopération renforcée entre l'opérateur privé et l'opérateur public. La notion de partage de la mission de service public serait renforcée en prenant appui sur un réel partenariat basé, enfin, sur un modèle gagnant/gagnant.

Cette possibilité avait été évoquée lors de la dernière journée d'études de l'Association des archivistes français (le 11 mars 2011). Je me souviens que la représentante d'Ancestry, qui développait alors le même projet que NotreFamille, avait laissé entendre que cette solution leur conviendrait. Nous n'avions pas eu la chance d'entendre l'avis de Toussaint Roze à ce sujet. Mais je suppose que cela lui conviendrait également, sauf à ce que son projet soit réellement de faire payer l'accès au patrimoine culturel. »

¹ <http://papiers.poussieres.free.fr/index.php/2011/08/06/valoriser-le-patrimoine-entre-bien-public-et-interets-prives/#more-883>, cité et commenté dans *Archimag*, 1^{er} janvier 2012, « Difficile commerce des données généalogiques », p. 20.

Projets

Au moins seize projets d'indexation collaborative sont prévus pour 2012. Douze concernent les Archives départementales :

- Alpes-Maritimes : identification de photographies, édition de corpus en wikisource
- Cantal : géolocalisation des données culturelles sur ordiphones
- Charente : état-civil, etc.
- Hérault : abandon de l'état-civil et des recensements en 2010 ; projet de reprise
- Landes
- Manche : BMS-NMD-TD
- Meurthe-et-Moselle
- Moselle : listes de combattants de 1914-18
- Nièvre : photographies
- Oise : listes nominatives
- Hautes-Pyrénées : terriers (noms propres et noms de matières), registres de délibérations (mots matière)
- Savoie

Trois concernent les communes :

- Bordeaux (Bibliothèque municipale) : indexation collaborative de gravures XVIIIe de la ville
- Romans-sur-Isère
- Saint-Denis : identification d'images ; cartographie collaborative

La direction de la mémoire, du patrimoine et des archives du ministère de la défense envisage de proposer un module d'indexation collaborative, à la fin de 2012, sur son site « Mémoire des hommes »¹.

Enfin, dans le monde des opérateurs, outre la disparition de la société Evodia, il est à noter que les progiciels Thot (société SICEM) et Pleade (société AJLSM) ont désormais (ou sont sur le point de proposer) un module d'indexation collaborative.

Un bilan et des perspectives encourageants

Des projets communaux peu nombreux mais très réussis, 20 % des départements engagés dans le web 2.0 : le bilan est d'ores et déjà flatteur, mais ce n'est qu'un début, et les perspectives sont plus flatteuses encore : fin 2012, 30 % des départements se seront mis au web 2.0, soit une hausse attendue de 50 % en un an. À ce rythme, tous les départements proposeront cette fonctionnalité (qui est le plus souvent comprise dans les appels d'offre des quelque 20 départements qui n'ont pas encore d'archives numérisés en ligne) en 2015 ; l'auteur du présent rapport d'étape souhaiterait vivement que ses estimations soient dépassées. Les outils existent, les plâtres sont essuyés, la demande du public va croissant. Le nombre total de signets posés par les internautes dépassera alors à coup sûr toutes les entreprises, commerciales ou associatives, surtout si les projets sont conçus en collaboration et bonne intelligence avec les associations généalogistes, comme c'est le cas, par exemple, dans l'Aube. Cela suppose que les associations partenaires partagent avec le service public le même objectif d'entraide démonétisée et qu'elles déploient toujours davantage leur faculté de diversifier leurs activités et de

¹ Message électronique à l'auteur de Sandrine Aufray, chef de projet « Mémoire des hommes », du 9 décembre 2011: « Le module d'indexation collaborative consistera à saisir et rendre interrogeables les informations complémentaires contenues dans les fiches numérisées selon des champs prédéfinis. En fonction du succès rencontré, nous envisagerons d'étendre l'indexation collaborative à d'autres fonds, en particulier aux registres de l'Infanterie de ligne et de la Garde impériale sous la Révolution et le Premier empire. La période couverte est principalement celle des conflits contemporains (Première et Seconde Guerre mondiale, Guerre d'Indochine, Guerre d'Algérie et des combats du Maroc et de la Tunisie, Guerre de Corée) pour lesquels des bases nominatives sont disponibles, mais également le XVIII^e siècle depuis la mise en ligne des fonds relatifs à la Compagnie des Indes en mai 2011. »

trouver les ressources dont elles ont besoin ; c'est dans le domaine de la formation (paléographie, archivistique, sciences auxiliaires de l'histoire, voire latin), en particulier, que la demande est la plus forte, et que services d'archives et associations de généalogistes peuvent proposer une offre didactique conjointe et adaptée.

Le web 2.0, c'est le pari que le meilleur connaisseur d'un document d'archives, c'est celui qui désire le déchiffrer, et qui acquiert, grâce à ses appétences, les compétences nécessaires et la familiarité avec toponymes et patronymes, conditions indispensables pour une transcription sans erreur. Comment le sous-traitant travaillant dans un pays lointain pourra-t-il rivaliser avec un généalogiste motivé et enthousiaste devenu, à force de fréquenter les archives, le meilleur connaisseur des noms de sa petite patrie ou de la terre de ses ancêtres ?

Le web 2.0, c'est le pari de la complémentarité entre le service public et le public lui-même : le premier apporte expertise archivistique, encodage et mise à disposition numérique des fonds, acquisition et développement d'un outil collaboratif et régulation du travail ; le second donne temps, motivation, compétence et désintéressement, dans le plaisir du partage. Le service public des archives, adoptant une posture nouvelle pour un enrichissement mutuel du « top » et du « down », organise ainsi un « gain informationnel » sans « déperdition conceptuelle », contrairement à ce qui est déploré parfois dans d'autres formes de web 2.0¹. La description normalisée et l'autorité des auteurs des instruments de recherche sont enrichies par les « tags » des internautes.

Les archives publiques, vont au-delà de l'objectif, affiché par le ministère de la Culture et de la Communication, de « la culture pour chacun ». Chacun est l'acteur de l'enrichissement du patrimoine commun ; le travail de chacun profite à tous. La culture de chacun enrichit la culture de tous, et réciproquement, dans un cercle vertueux de partage démonétisé de la connaissance. Fer de lance du web culturel en Europe, les archives communales et départementales françaises sont en train de coproduire avec leurs publics un web patrimonial et culturel 2.0 sans équivalent, à la fois savant et populaire.

¹ BROUDOUX (Evelyne), « Indexation collaborative : entre gain informationnel et déperdition conceptuelle, dans *Traitements et pratiques documentaires. Vers un changement de paradigme ?*, Paris, 2008, disponible sur http://archivesic.ccsd.cnrs.fr/docs/00/36/46/26/PDF/Doc-Soc_Broudoux_V5.pdf

Édouard BOUYÉ
Directeur des Archives départementales du Cantal
ebouye@cg15.fr

Le web collaboratif dans les services d'archives publics : un pari sur l'intelligence et la motivation des publics

La « folksonomie » désigne la possibilité offerte à l'internaute d'enrichir de ses annotations les instruments de recherche et les documents mis en ligne. Les opérateurs du marché de la mise en ligne des archives proposent le plus souvent cette fonctionnalité ; reste à l'archiviste la tâche de choisir les fonds et les modalités scientifiques et techniques du travail collaboratif. Vingt-et-un services d'archives se sont déjà lancés ; seize projets sont annoncés pour 2012. Les permaliens sont une alternative moins contraignante ; ils offrent plus de liberté, mais les annotations ne sont pas intégrées aux IRE. Dans le web 2.0, chacun est l'acteur de l'enrichissement du patrimoine commun ; le travail de chacun profite à tous. La culture de chacun enrichit la culture de tous, et réciproquement, dans un cercle vertueux de partage démonétisé de la connaissance. C'est un pari, à la fois réussi et prometteur, sur la complémentarité entre le service public des archives avec les appétences et les compétences du public. Fer de lance du web culturel européen, les archives en France sont en train de coproduire avec leurs publics un web patrimonial et culturel 2.0 sans équivalent, à la fois savant et populaire.