Le numérique participatif au service de la République des Lettres

L'exemple des archives

Édouard Bouyé

Conservateur général du patrimoine,
directeur des archives départementales de la Côte-d'Or

Résumé : Loin des clichés, l'inflation de la production documentaire a fait des archivistes les animateurs d'équipes nombreuses ; et la consultation des archives est l'une des pratiques culturelles les plus populaires. Cette rencontre d'un public motivé et d'une documentation numérique considérable a suscité le développement d'entreprises de *crowdsourcing* permises par la pratique des documents d'archives ainsi que par les connaissances toponymiques et onomastiques des usagers : édition de textes, annotation collaborative de l'état civil ou de fonds iconographiques, établissement de dictionnaires historiques, etc. Le rôle de l'archiviste est de rendre possible ce travail sur les sources, de l'encadrer, de le faire connaître (par la médiation numérique). Il faut susciter la création d'un réseau, d'une communauté virtuelle (en s'appuyant sur des communautés réelles) et donner de la reconnaissance à son travail. Le public attend de l'archiviste, qui doit se déprendre de la posture du sachant distillant son savoir, qu'il soit néanmoins le garant de la fiabilité de l'information produite, même s'il n'a pas les moyens concrets de la valider systématiquement *a priori*. Diverses questions doivent être examinées : celle de la propriété intellectuelle de l'information produite, celle des rapports entre *crowdsourcing* et *crowdfunding*, celle des raisons et des conditions du succès des entreprises de *crowdsourcing* dans les archives, où chaque participant contribue à l'enrichissement du patrimoine commun.

Mots-clés : Patrimoine, généalogie, communauté, travail collaboratif, indexation

Abstract : Far from the usual stereotypes, the inflation of the documentary production has turned the archivists into leaders of big teams ; and one of the most popular cultural activities is seeking access to archives. The motivation of people in front of a considerable amount of digital documentation, their ability to handle documents and their knowledge of toponymy and name research allowed the development of many crowdsourcing projects : text-editing, collaborative annotation of civil registrars or of iconographic collections, producing historic dictionaries, etc. The archivist has to make this work on sources possible, to oversee it, to make it known (though digital mediation). He has to give rise to the creation of a network, of an online community (based on real ones) and to obtain the recognition of their work. He must stop taking the posture of the one who knows and nevertheless be the one who guarantees the reliability of the information that is produced, even if he can't verify it systematically *a priori*. Various issues must be examined : intellectual property of produced information, links between *crowdsourcing* and *crownfunding*, reasons

and conditions for the success of the crowdsourcing archive projects, in which each member participates to the development of public heritage.

Keywords : Heritage, genealogy, community, collaborative work, indexation

---

> Les gens ont envie de tout prendre, de tout piller, de tout bouffer, de tout manipuler. Voir, déchiffrer, apprendre ne les affecte pas. Le seul affect massif, c'est celui de la manipulation. Les organisateurs (et les artistes et les intellectuels) sont effrayés par cette velléité incontrôlable, car ils n'escomptent jamais que l'apprentissage des masses au spectacle de la culture
> (Baudrillard, *Le Pacte de lucidité ou l'intelligence du mal*, 2004, p. 106).

L'image de l'archiviste compulsant, solitaire, des vieux papiers abscons est le *topos* commode, cocasse et clochemerlesque où le comique naît du grossissement des traits, qui veut que le notaire soit cupide, le croque-mort amateur de gauloiseries, et l'artiste forcément homosexuel. Mais, quoiqu'Anatole France nous présente son Sylvestre Bonnard, chartiste célibataire et érudit, chaussé de « pantoufles » et dans une « robe de chambre », dès la première ligne du roman éponyme [1], l'image n'a jamais complètement correspondu à la réalité du métier, depuis toujours situé au cœur du tissu de relations humaines et savantes de la République des Lettres. Ce modèle ancien a produit de beaux fruits de connaissance. Ce sont les savants du XVI^e^ siècle qui se font communiquer par leurs correspondants des inscriptions ou des copies ; ce sont les Académies et les sociétés savantes qui, depuis le XVIII^e^ siècle, font la synthèse des informations dont elles sont à la fois sources et réceptacles ; c'est le même Sylvestre Bonnard qui va jusqu'en Sicile, en 1869, pour retrouver un manuscrit ; c'est l'exemple éloquent du *Dictionnaire d'Oxford* (voir l'article d'Emmanuelle de Champs), qui manifeste le tropisme anglo-saxon pour la science collaborative, confirmée par l'origine exclusivement anglo-saxonne des participants non-français au colloque sur le crowdsourcing organisé à Angers en octobre 2017 [2]. Pour autant, la France a depuis longtemps l'habitude de faire collaborer le public à la constitution de son patrimoine (écomusées, préinventaire, etc.). C'est le cas, depuis presque un siècle, pour les archives privées, dont la collecte présuppose que l'archiviste soit intégré à de multiples réseaux amicaux, professionnels, familiaux – des réseaux sociaux, en somme, dans une acception large. L'éclatant succès de la Grande Collecte de la Grande Guerre, entre 2013 et 2018, nous parle naturellement du rôle du premier conflit mondial dans la mémoire collective ; il montre aussi que le réseau des archives en France jouit de la confiance de bon nombre de nos compatriotes. La Fabrique du patrimoine lyonnais est un projet du même genre, plus ancré encore dans les quartiers. Mais il est vrai que la science participative a une tradition encore plus implantée en France dans le domaine des sciences naturelles (botanique, astronomie) comme le montre François Houllier [3], et c'est bien naturellement que les herborisateurs sont devenus des herbonautes.

1. France, Anatole (1881) *Le crime de Sylvestre Bonnard*, Paris, Calmann Lévy.

2. Colloque organisé par Cécile Meynard, Florence Alibert, Élisabeth Greslou, Thomas Lebarbé et Valérie Neveu du 18 au 20 octobre 2017, à l'université d'Angers.

3. Houllier, François (dir.) (2016) *Les Sciences participatives en France (février 2016). État des lieux, bonnes pratiques & recommandations*. Rapport élaboré à la demande des ministres en charge de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche. Accessible sur `http://www.sciences-participatives.com/Rapport`, consulté le 25 juillet 2020.

Published in : Cécile Meynard et Thomas Lebarbé (dir.), *Le Crowdsourcing : Partager, enrichir et publier des sources patrimoniales*, Editions des archives contemporaines, ISBN : 9782813003850, doi : 10.17184/eac.9782813003850

Dans le domaine des archives, les pratiques collaboratives sont venues de la généalogie. Loisir aristocratique ou bourgeois (gentilhomme) jusque dans les années 1960, elle est devenue une passion populaire depuis les années 1970. Les généalogistes, actifs, impatients et demandeurs, ont contribué à faire évoluer les services d'archives ; dans le contexte de la décentralisation du début des années 1980, ils ont fait puissamment levier sur les départements pour accélérer la modernisation des bâtiments d'archives et la mise en ligne des sources généalogiques, au premier rang desquelles se trouvent naturellement les registres paroissiaux puis d'état civil. C'est ce qui a aussi permis la modernisation progressive du modèle ancien : les dépouillements d'état civil, souvent mis en ligne sur les sites des associations de généalogistes [4], peuvent aussi être mis en ligne sur un site d'archives publiques, sous forme de base de données ou en format texte. Les usagers internautes peuvent encore contribuer à l'affinement d'un répertoire numérique, dont le dépouillement des originaux en salle de lecture permet de faire un inventaire plus détaillé ; ils peuvent également numériser en salle de lecture des séries d'archives que le service met ensuite en ligne (*via* des opérations ponctuelles ou au long cours, spontanées ou encadrées par une convention) ; ils peuvent, enfin, signaler par courriel les erreurs d'inventaire ou les anomalies dans les images numérisées. Ce faisant, les archivistes sortent du modèle traditionnel et unidirectionnel *top-down*, dans lequel le professionnel est le pourvoyeur d'informations et de science. Cela implique, pour l'archiviste, un changement de posture : les archives cessent d'être arcanes ; le professionnel n'a plus l'exclusivité, il y a un partage du droit moral sur inventaires et images, et une obligation de réactivité pour intégrer aux contenus en ligne les additions, corrections et suggestions des usagers-internautes.

1 Des archives numériques, mais aussi humaines

Depuis la fin du XX^e^ siècle, les archives en France ont mené à bien de très vastes campagnes de numérisation du patrimoine écrit, mis en ligne avec régularité depuis 2003-2004. Les chercheurs se sont faits plus rares dans les salles de lecture, qu'ils ont transformées en studios photographiques pour mieux analyser chez eux des numérisations réalisées au kilomètre. Jamais les services d'archives n'ont eu plus de lecteurs effectifs, mais ils connaissent, dans le même temps, une fréquentation « physique » comparable à celle de l'époque où ils accueillaient seulement quelques lecteurs : universitaires, chanoines, notables et érudits. Le besoin s'est donc fait ressentir de recréer une communauté, en partant du public des expositions et des cours de paléographie, et en s'appuyant, quand ils existaient, sur une association des amis des archives, une société savante ou un cercle de généalogie. Il faut, en somme, créer des espaces et des moments permettant à une communauté physique de se parler, de partager trouvailles et curiosités, d'échanger, en conservant l'esprit et les pratiques des sociétés savantes. Cela permet d'éviter l'erreur qui a conduit à évincer de l'archéologie les archéologues amateurs au moment de sa professionnalisation dans les années 1990, transformant parfois des bénévoles motivés en détectoristes destructeurs de sites archéologiques. Le dialogue et la collaboration avec les généalogistes, avec les associations d'histoire

4. Leur vente à des sociétés commerciales de généalogie permet d'équilibrer des finances mises à mal par la désertion des adhérents, lesquels trouvent sur les sites payants – mais parfois moins chers qu'une adhésion à un cercle généalogique ! – une information plus rapide, ou, sur les sites des services d'archives, une information gratuite et complète.

Published in : Cécile Meynard et Thomas Lebarbé (dir.), *Le Crowdsourcing : Partager, enrichir et publier des sources patrimoniales*, Editions des archives contemporaines, ISBN : 9782813003850, doi : 10.17184/eac.9782813003850

locale, avec les Amis des archives, permet l'éclosion, dans un climat cordial et savant à la fois, d'un écosystème où les archives publiques sont garantes de la pertinence et de l'intégrité des ressources mises à disposition, mais où les bénévoles contribuent à son amélioration et à son enrichissement. C'est ainsi qu'en Côte-d'Or s'est créée, en 2015, une « association des Amis des archives de la Côte-d'Or et des deux Bourgognes » (appellation qui manifeste la profusion des sources de l'histoire de la Bourgogne et de la Franche-Comté dans le dépôt dijonnais), à partir des participants aux cours de paléographie et de latin. Cette association a développé ses activités, proposant aussi des conférences « Sources et méthodes », des conférences « Jeunes chercheurs sur la Bourgogne antique et médiévale », des cours de sigillographie, des cafés-archives, des « rabâcheries » de patois bourguignon permettant aux chercheurs d'échanger sur leurs trouvailles et leurs recherches, et créant des occasions de médiation, numérique ou non. Cette communauté signale aussi à l'archiviste les fonds en déshérence et les difficultés de communication des archives dans certains lieux. Elle fournit également le vivier de bêta-testeurs lorsque le site Internet évolue. Son savoir-faire et ses compétences augmentent dans le domaine et de la lecture et de la compréhension des textes anciens, formant le noyau des contributeurs potentiellement les plus actifs d'un projet d'indexation collaborative. Ces initiatives et ces projets sont naturellement relayés par les médias sociaux : le numérique et l'humain s'enrichissent mutuellement en un cercle vertueux.

2 Quelle posture pour le service public ?

Au début du XXI^{e} siècle, il n'était pas rare d'entendre des collègues répéter que la mission du service public se bornait à mettre à la disposition du public les documents physiques, et que la numérisation était une mesure de conservation préventive et une concession, par pure grâce, à l'insatiable curiosité des internautes. Chaque archiviste de France est convaincu (et le législateur a achevé de convaincre les hésitants en 2016) que cette mise en ligne n'est pas surérogatoire, mais indispensable. On trouve en revanche le même genre de jugement à l'égard de l'indexation collaborative : aux services d'archives la mise à disposition, aux chercheurs les recherches et les index. Le président de la fédération française de généalogie, Jean-François Pellan, n'est pas loin de partager ce point de vue.

> La généalogie associative est prise entre le marteau et l'enclume, depuis quelque temps. L'enclume, ce sont les mises en ligne par les services d'archives. Elles sont nombreuses. Tout le monde s'en réjouit, mais ce n'est pas sans conséquence pour les associations qui voient le nombre de leurs adhérents diminuer. Pour ne rien arranger, certains directeurs d'archives se sont mis en tête de faire de l'indexation collaborative pour contrer les sociétés commerciales. Le marteau, ce sont les sociétés qui mettent en ligne, soit gratuitement des indexations, des images ou encore des images avec des indexations mais de façon payante. Certaines associations sont parfois en cheville avec ces sociétés pour gagner quelques sous [5].

5. Pellan, Jean-François (2017) « Compte rendu de l'Assemblée générale 2016 (exercice 2015) de la fédération française de généalogie », dans *La Vie généalogique,* n° 45, novembre 2017, p. 24-34. Citation p. 27.

Published in : Cécile Meynard et Thomas Lebarbé (dir.), *Le Crowdsourcing : Partager, enrichir et publier des sources patrimoniales*, Editions des archives contemporaines, ISBN : 9782813003850, doi : 10.17184/eac.9782813003850

Il faut dire que les projets d'indexation collaborative des sources généalogiques font concurrence à l'activité principale des associations généalogiques et créent un manque à gagner pour celles qui en font commerce. Ces projets font également pièce aux entreprises commerciales qui achètent à bas prix (voire, suivant le choix des exécutifs locaux, à coût nul) les images numériques que les archives sont de toute façon obligées de leur céder, les font indexer à l'île Maurice, en Inde ou à Madagascar, puis les revendent. Cette « ubérisation » du travail généalogique peut servir d'aiguillon aux projets participatifs, pourvu que les archivistes veuillent bien changer de posture et passer du *top-down* au *bottom-up*. C'est ce à quoi invite l'étude annuelle du rapport 2017 du Conseil d'État, intitulée *Puissance publique et plateformes numériques : accompagner l'ubérisation*.

Cela implique un pari, une prise de risque, un lâcher-prise, un changement de posture : le savant, le sachant, le surplombant doit apprendre à devenir un partenaire. Les premières entreprises menées par les archives départementales entraînent les suivantes (fin 2017, un tiers des services d'archives départementales était engagé dans un projet collaboratif), tout comme Gallica a fait évoluer le réseau des bibliothèques (voir l'article de Mathieu Andro). Les sites Internet ainsi enrichis deviennent des patchworks où tout n'est pas uniformément renseigné ou indexé, où tout n'est pas parfait comme dans un instrument de recherche ayant reçu le visa du service interministériel des Archives de France : il faut l'accepter, en se disant que le numérique permet d'améliorer continûment la qualité de l'information, possibilité que n'offrait pas l'impression en 300 exemplaires des inventaires d'archives, ce qui était encore la pratique habituelle à la fin du XX^{e} siècle.

3 Le choix de l'outil

Lorsque l'on veut se lancer dans un projet collaboratif se présentent trois solutions. Utiliser un outil existant (Youtube, Wikisource, Flickr ou encore Zooniverse, pour ne citer que ceux qui ont fait l'objet d'une présentation lors du colloque d'Angers précité) permet, pour un faible coût, d'avoir beaucoup d'audience, mais pose le problème du rapatriement des informations créées dans le système propre du porteur de projet.

On peut aussi créer des outils nouveaux *ad hoc*, dans le cadre d'un projet de type LABEX, ou d'un PPP (partenariat public-privé). Ces montages assurent un financement ample, quoique limité dans le temps, une synergie de compétences scientifiques et techniques appréciable. Mais il arrive que le projet réinvente à grands frais un outil qui existe déjà. Le problème principal est celui de la pérennisation du résultat après la fin du projet, s'il n'est pas adossé à une institution de conservation. Il n'est pas rare de déplorer la déshérence de bases de données, d'applications ou de *Games With A Purpose* (GWAP) pourtant riches et inventifs.

La troisième solution, intermédiaire, est celle que privilégient les services d'archives. Dans le marché de niche qu'a suscité la mise en ligne massive des archives, quatre ou cinq entreprises se détachent. Elles ont développé, ou ont l'intention de développer, des fonctionnalités collaboratives. L'internaute, qui a l'habitude de « son » outil, peut enrichir en ligne les contenus, et son travail est appelé à être disponible pour les internautes suivants. Deux points d'attention : l'ergonomie des fonctionnalités colla-

Published in : Cécile Meynard et Thomas Lebarbé (dir.), *Le Crowdsourcing : Partager, enrichir et publier des sources patrimoniales*, Editions des archives contemporaines, ISBN : 9782813003850, doi : 10.17184/eac.9782813003850

boratives et la réversibilité qui permet de récupérer les contenus en cas, par exemple, de changement d'outil ou de migration de données, que l'on ait choisi initialement un logiciel libre ou un logiciel propriétaire.

4 Panorama en 2017

C'est la généalogie qui se taille la part du lion dans le paysage des archives collaboratives [6]. Historiquement, ce sont les registres paroissiaux et d'état civil (18 services) qui ont ouvert la voie, mais l'essentiel des projets porte désormais sur les registres matriculaires du recensement militaire (37 services), notamment pour participer au Grand Mémorial des combattants de la Première Guerre mondiale. Les listes nominatives de recensement (12 services) complètent logiquement cet ensemble généalogique, enrichi ponctuellement de sources plus rares (déclarations de grossesse, registres d'inhumation). S'y ajoutent des sources qui dépassent et enrichissent la généalogie : tables de successions et absences, minutes notariales (y compris les testaments de Poilus classés au rang des minutes), matrice cadastrale, cahiers de doléances.

Vincent Doom et Stéphane Rodriguez-Spolti ont dit la difficulté de trouver des collaborateurs nombreux pour un projet non généalogique (voir l'article de Stéphane Rodriguez-Spolti). La piste la plus prometteuse semble être celle de l'iconographie, comme le montre le projet *PhotosNormandie* (voir l'article de Patrice Peccatte). L'identification, voire l'indexation de photographies, documents parfois muets, est à la fois un jeu et un défi. On peut y adjoindre, par exemple, les plans cadastraux et les menus, plus facilement indexables de façon sérielle et sans connaissance précise.

L'indexation peut aller jusqu'à l'édition scientifique, comme c'est le cas, dans les Alpes-de-Haute-Provence, pour les visites pastorales de l'époque moderne (wikisource). Enfin, certains services, telles les archives départementales de la Vendée, mettent la pratique collaborative au cœur de leur projet de service, en l'intégrant au processus même de traitement des archives.

5 Profil des indexeurs

Lorsque les projets sont libres et que les indexeurs ne doivent ni ne peuvent s'inscrire, il est difficile de connaître le profil de la communauté. Il est certain que le degré d'expertise requis dépend de la technicité des tâches proposées à l'indexeur, qui peuvent aller de la pose de tags simples à l'encodage en TEI.

Mais l'idée générale est d'additionner les groupes (les « bulles », comme les appelle Yolaine Coutentin, voir son article) : « amis des archives », généalogistes (surtout lorsqu'ils sont loin du « pays », et que la nostalgie les pousse à contribuer à l'écriture de son histoire), étudiants, historiens locaux, chercheurs professionnels. La communauté se constitue en tous les cas autour d'une commune curiosité et du même goût de l'intérêt général ; le travail de chacun profite à tous ; le public est acteur de l'amélioration du service public.

6. Boudarel, Sophie (2017) « L'art de bien indexer », dans *Revue française de généalogie*, n° 232, octobre-novembre 2017, p. 52-54.

Published in : Cécile Meynard et Thomas Lebarbé (dir.), *Le Crowdsourcing : Partager, enrichir et publier des sources patrimoniales*, Editions des archives contemporaines, ISBN : 9782813003850, doi : 10.17184/eac.9782813003850

Le cas des étudiants mérite d'être traité à part. Leur contribution est souvent efficace, par exemple pour la correction de l'OCR du *Dictionnaire universel* (dans le cadre du projet de Ioana Galleron et Geoffrey Williams [7]). Il s'agit certes d'un public captif, dont la contribution aura forcément quelque chose d'un peu artificiel. Soulignons d'emblée que ce public permet de comprendre les appétences et les compétences d'une classe d'âge. Il n'est pas rare non plus d'y trouver des passionnés, graines d'indexeurs. Mais il faut attendre des étudiants plutôt des compétences technologiques qu'une connaissance scientifique approfondie des fonds. Associer au projet des étudiants en archivistique (licences professionnelles, masters 1 et 2, archivistes paléographes) permet aussi de tester le produit, de l'améliorer, voire de le promouvoir. Et il est logique d'associer ces futurs professionnels à ces projets d'avenir.

Plusieurs contributeurs posent, au moins implicitement, une question éthique à propos de cette « servuction », mot-valise proposé par Mathieu Andro, composé de « services » et de « production » (voir son article). Est-il légitime de « profiter » ainsi d'une main-d'œuvre gratuite ? La figure sympathique mais troublante du docteur William Chester Minor, exceptionnel contributeur du *Dictionnaire d'Oxford,* enfermé comme psychotique, est emblématique de ce questionnement (voir l'article d'Emmanuelle de Champ). Addiction [8] ; relation amoureuse avec Gallica [9] ; fanatisme (voir l'article de Suzan Van Dijk) ; obsession du collectionneur ; temps « captif » du prisonnier, du handicapé, de l'étudiant, du retraité qui se sent inutile : est-il moral d'exploiter ce qui s'apparente parfois à des TOC ; n'est-ce pas là du travail dissimulé que les pouvoirs publics se procurent à bon compte ? Si l'on se persuade qu'il s'agit de temps pris aux jeux vidéo (GWAP mis à part) ou à la télévision (qui représente, en temps, 2000 Wikipedia, comme le souligne Mathieu Andro dans son article), le scrupule fait place à la satisfaction de proposer de participer à un projet commun d'intérêt général qui forme de surcroît une communauté humaine. Le *crowdsourcing*, qui donne le sentiment d'être utile, permet aussi de renouer du lien social pour des publics un peu en marge, ou qui se considèrent eux-mêmes comme tels. Comme le résume l'une des participantes au projet CORRECT, de la Bibliothèque nationale de France, il s'agit, en définitive, de « mots croisés utiles [10] ». L'*homo ludens* redevient l'*homo faber* : en jouant, il fabrique et enrichit le patrimoine commun.

6 Rôle du service public

À supposer que l'on ait au préalable jugé le service public légitime à lancer un projet d'indexation collaborative, il convient de se poser d'emblée la question du rôle de ce service public dans la validation du travail de la communauté pour assurer une qualité optimale des données produites. S'il s'agit de *sourcing*, d'édition collaborative de textes littéraires, d'encodage très structuré (par exemple en TEI) de données com-

7. On peut trouver une présentation de ce projet sur le site `http://www.licorn-research.fr/Basnage.html`, consulté le 25 juillet 2020.

8. Mathieu Lafourcade racontait au colloque de 2017 sur le *crowdsourcing* que, lorsque les serveurs de jeuxdemots.org sont en panne, des internautes lui proposent de l'argent pour contribuer à les remettre en marche…

9. Tweets relevés par Pauline Moirez, conservatrice à la BnF, lors d'une intervention dans le cadre du colloque de 2017 (non publié).

10. Expression relevée également par Pauline Moirez.

Published in : Cécile Meynard et Thomas Lebarbé (dir.), *Le Crowdsourcing : Partager, enrichir et publier des sources patrimoniales*, Editions des archives contemporaines, ISBN : 9782813003850, doi : 10.17184/eac.9782813003850

plexes (tels les registres médiévaux présentés par Marika Ceunen et Agata Dierick, voir l'article de Marika Ceunen), le service public expert doit valider le travail une fois qu'il est achevé. S'il s'agit au contraire de *tagging*, de pose d'index simples et rapides, on peut imaginer une régulation par la communauté. C'est le principe que nous avions choisi dans le Cantal en 2010 pour l'indexation des registres paroissiaux et d'état civil [11] : n'importe quel internaute peut poser un index, n'importe quel internaute peut envoyer un message d'erreur au Webmestre, mais aucun internaute ne peut modifier ni supprimer un index. C'est le Webmestre qui tranche et qui arbitre ; en 2013, nous avions calculé que, pour 700 signets posés chaque jour, seuls 5 messages d'erreur quotidiens parvenaient au Webmestre, soit 35 par semaine, soit 7 arbitrages à rendre par jour ouvrable, soit une vingtaine de minutes de travail quotidien.

Il existe naturellement d'autres systèmes de régulation aussi faciles à gérer par le Webmestre. Le double ou triple examen des mêmes archives à indexer par des indexeurs qui ne se connaissent pas permet de valider le travail : seuls les index identiques sont automatiquement validés et publiés.

Puisque les modalités de validation varient en fonction de la difficulté du travail à accomplir, on peut anticiper le temps que cette validation ou modération prendra en équivalent temps plein (ETP) lors de la mise en route du projet (actions de médiation numérique et de pédagogie pour les indexeurs novices) comme dans le rythme de croisière. Si les index doivent être validés systématiquement, il convient de faire en sorte que le délai de validation et donc de mise en ligne soit le plus court possible, sauf à décourager les indexeurs, surpris puis agacés de ne pas voir en ligne le fruit de leur travail. Enfin, et par-delà l'évident intérêt d'un projet collaboratif en termes de partage, de découverte, etc., il faut comparer le temps que prendrait une indexation effectuée entièrement par le service public, avec le temps passé à la médiation numérique, à la relecture, à la correction et à l'encodage du travail des internautes ; en d'autres termes, pour que le jeu en vaille la chandelle, il faut que l'investissement en temps soit « rentabilisé » par la collaboration des internautes ; sinon, l'exercice perd, c'est le moins que l'on puisse dire, de son intérêt.

7 Où placer le curseur ? Les questions à se poser

Avant de lancer le projet, il faut placer le curseur en fonction des objectifs du projet, de la complexité du fonds et de l'outil, de la masse de travail totale et des moyens humains que l'on peut y consacrer. Voici quelques-unes des questions à se poser pour trouver le point d'équilibre adapté au projet :

11. Voir `https://archives.cantal.fr/?id=recherche_guidee_etat_civil`, consulté le 26 juillet 2020.

Published in : Cécile Meynard et Thomas Lebarbé (dir.), *Le Crowdsourcing : Partager, enrichir et publier des sources patrimoniales*, Editions des archives contemporaines, ISBN : 9782813003850, doi : 10.17184/eac.9782813003850

Tableau 1

Collaboration/Participation	**Contribution**
Tagging	*Sourcing*
Annotation	Structuration
Travail anonyme ; pas de reconnaissance pour les indexeurs	Travail sourcé ; reconnaissance pour les contributeurs
Plein texte	Données structurées
Au fil de l'eau	Par lot
Fonds simple	Fonds complexe
Tâche simple et/ou élémentaire	Tâche complexe
Productivité forte	Productivité faible
L'engagement peut être faible	L'engagement peut être fort
Outil simple et intuitif	Outil complexe doté de didacticiels ; technicité qui peut faire peur ou décourager
Liberté	Nécessité d'une inscription, voire de la réussite aux tests
Prend peu de temps	Chronophage
Masse/Foule/*Crowd*	Communauté
Grand public	Experts
Microtasks	*Macrotasks, voire metatasks*
Recherche d'amélioration	Recherche de perfection
Pas de droit d'auteur	Droit d'auteur si le travail va au-delà de la transcription
Risque de malveillance	Possibilité d'identifier et d'exclure les malveillants

8 Les facteurs de réussite

Adopter la bonne posture, mettre à la disposition du public des instruments techniques : ce sont là des conditions nécessaires, mais pas suffisantes à la réussite d'un projet. Le mérite de notre colloque est aussi d'avoir présenté des expériences en demi-teintes, ou des voies finalement abandonnées.

Les enquêtes d'opinion montrent une forte motivation de la société française pour le bénévolat [12]. Il est intéressant, pour définir un projet collaboratif, de savoir que la tranche d'âge la plus active est celle des plus de 65 ans ; que les diplômés du supérieur sont au premier rang ; que les trois premières motivations sont, dans l'ordre décroissant, « Être utile à la société et agir pour les autres », « Agir de façon concrète »

12. Recherches & Solidarités. *La France bénévole*, 14e édition, juin 2017. Accessible sur `https://recherches-solidarites.org/benevolat/`, consulté le 27 juillet 2020. Publication de la *10e enquête annuelle du Baromètre d'opinion des bénévoles (BOB)/IFOP*. Réalisée en ligne entre le 1r mars et le 7 avril 2017. Panel de 3062 bénévoles de 18 ans et plus, actifs dans une association. Échantillon représentatif des bénévoles en France, traité selon la méthode des quotas.

Published in : Cécile Meynard et Thomas Lebarbé (dir.), *Le Crowdsourcing : Partager, enrichir et publier des sources patrimoniales*, Editions des archives contemporaines, ISBN : 9782813003850, doi : 10.17184/eac.9782813003850

et « Lier des relations avec les autres ». C'est donc dans ce cadre général que prend nécessairement place le volet français d'un projet collaboratif.

La difficulté propre aux archives et à la science, c'est une image trop aride et institutionnelle, qu'il s'agit donc de nuancer, de transformer. Des principes simples doivent ensuite être suivis.

Tout projet doit avoir un ou plusieurs animateurs (*community managers*), qui relancent l'intérêt, interagissent avec les contributeurs, pratiquent la médiation numérique. Il s'agit d'une animation d'un caractère presque commercial. John Murray, qui fut l'âme du *Dictionnaire d'Oxford*, est la figure pionnière de l'infatigable animateur de projet de *crowdsourcing* (voir l'article d'Emmanuelle de Champ).

Il faut faciliter la vie des indexeurs, si l'on veut qu'ils viennent en foule et dépassent le cadre d'un comité d'experts. Foin des interfaces incompréhensibles, hérissées d'acronymes et comportant des dizaines de champs obligatoires ! Quel que soit le système choisi pour la validation et la mise en ligne, il faut qu'il soit réactif.

Il convient d'additionner les viviers potentiels et les communautés existantes ou à susciter. À côté de la « foule », chaque projet attire aussi des « super-indexeurs », qui forment une petite minorité du total des contributeurs, mais fournissent une grosse majorité des index, dans un effet de longue traîne. Ces indexeurs sont, en généalogie, très représentés chez les jeunes retraités, sexagénaires et septuagénaires qui font d'ailleurs vivre les associations en général, et les associations généalogiques en particulier. Il faut trouver le moyen de les reconnaître, ce qui est simple si les contributeurs sont obligatoirement identifiés, mais plus difficile si l'indexation est libre – en ce cas, une information périodique sur le nombre de signets posés, voire une courbe cumulative des signets posés depuis le début du projet, permettent de situer le rameur solitaire au milieu de l'océan des indexeurs. On peut réfléchir à des marques de reconnaissance, sous forme de jeux, de récompenses ou, tout simplement, en publiant un palmarès, en affichant la performance ou en aménageant un *Hall of Fame*. On peut aussi proposer aux contributeurs de se réunir physiquement, une ou deux fois par an, pour des échanges de bonne pratique, des visites un peu exceptionnelles, à grand renfort de produits locaux (ou non), pour entretenir une atmosphère à la fois cordiale et concentrée favorisant le partage de la connaissance et la découverte.

Il faut prévoir du temps et de l'énergie pour la communication, ne pas être avare de médiation numérique ni de communication institutionnelle (envisagée comme un facteur de réussite et non comme une fin en soi en termes d'image) ; cela permet de donner une image attirante au projet, pour donner envie d'y participer. Cet effort doit être soutenu notamment durant le palier que l'on observe généralement dans la chronologie des projets : après une phase initiale dopée par l'attrait de la nouveauté, survient un moment où les indexeurs de masse ne sont pas foule, où ils peuvent se décourager par lassitude, et où les indexeurs potentiels, utilisant l'interface de recherche, constatent qu'ils n'obtiennent que peu de résultats, ce qui ne les encourage pas à venir en foule. Plus tard, dès qu'une masse critique d'index fournit à la plupart des recherches des résultats pertinents, encourageant les nouveaux arrivants à contribuer à leur tour à un projet qu'ils trouvent d'envergure, on peut considérer que ledit projet

Published in : Cécile Meynard et Thomas Lebarbé (dir.), *Le Crowdsourcing : Partager, enrichir et publier des sources patrimoniales*, Editions des archives contemporaines, ISBN : 9782813003850, doi : 10.17184/eac.9782813003850

a trouvé sa vitesse de croisière et que le bouche-à-oreille pourra faire le reste. Mais, dans la phase critique, les porteurs du projet ne doivent pas ménager leurs efforts pour en parler et en faire parler.

9 Quelle graphie choisir ?

Il faut d'emblée décider si la transcription donne le mot comme il devrait être, ou s'il convient de transcrire fidèlement le mot comme il est. Cela vaut pour des textes anciens : un moteur de recherche trouvera plus aisément « Marguerite » que « Margueritte », « Beaune » que « Beaulne », etc. Et pourtant, je pense qu'il convient de demander à l'indexeur de transcrire ce qu'il lit, car l'inviter à l'interprétation peut donner des résultats erratiques, également éloignés de l'orthographe réelle et de l'orthographe rêvée. Rien n'empêche de doubler la transcription littérale d'une transcription dans l'orthographe actuelle. Cela vaut aussi pour des textes plus récents, tels ces testaments de Poilus écrits en langue phonétique, qui nous renseignent sur leur degré d'alphabétisation et, le cas échéant, sur la prégnance des parlers locaux dans la langue effectivement parlée.

10 Jeremy Bentham, ou le plaisir comme levier

Il n'est pas indifférent que l'un des projets phares de *crowdsourcing* britannique soit consacré à la transcription de l'œuvre de Jeremy Bentham. Cet auteur a étudié le plaisir comme levier de l'agir humain. Or sont apparus, au fil des communications du colloque d'Angers, des mots appartenant au registre lexical du plaisir, ou à des registres circonvoisins : « *enjoyable task* » (Mia Ridge) ; « plaisir » (Mathieu Lafourcade) ; « curiosité », « excitation », « joie » (Bénédicte Grailles) ; « mots d'amour » (Pauline Moirez) ; il n'est pas jusqu'au nom de la plate-forme PHuN (Patrimoine et Humanités Numériques [13]) qui ne joue avec le *fun* et l'idée d'une activité attrayante et agréable. Les GWAP sont quant à eux directement dans le registre du jeu.

Or l'activité d'indexation ou d'édition collaborative est exigeante ; répétitive, aride parfois, elle peut être techniquement difficile ; elle n'évoque guère une foule en quête de plaisir.

On peut essayer d'élucider cette apparence contradiction au prisme du *felicific calculus* de Bentham. La qualité du plaisir dépend selon lui de sept critères : durée, intensité, certitude, proximité, étendue, fécondité et pureté.

Le plaisir de l'indexation est long et durable. Le champ est en effet immense, et il est multiple. Dans le Cantal, un internaute m'avait confié qu'il indexait chaque soir, durant une heure, l'état civil de sa commune, qui n'avait jusque-là intéressé presque personne. Comme un navigateur abordant une *terra (quasi) incognita*, ou en tous

13. Prototype de plate-forme de transcription collaborative de manuscrits d'auteurs, conçue par Anne Vikhrova (docteure de l'université Grenoble-Alpes), qui, depuis, a été repensé par l'équipe ELAN de l'université Grenoble-Alpes dans le cadre de la plateforme TACT (`https://tact.demarre-shs.fr/`). On y trouvera en particulier l'espace de transcription des manuscrits de *Mon évasion* de Benoîte Groult (`https://tact.demarre-shs.fr/project/5`), projet mené conjointement par Cécile Meynard (université d'Angers) et Élisabeth Greslou (université Grenoble-Alpes).

Published in : Cécile Meynard et Thomas Lebarbé (dir.), *Le Crowdsourcing : Partager, enrichir et publier des sources patrimoniales*, Editions des archives contemporaines, ISBN : 9782813003850, doi : 10.17184/eac.9782813003850

les cas vierge d'indexation, il avait entrepris d'en reconnaître et d'en décrire tous les recoins.

Quand elle concerne des ancêtres et/ou des lieux connus, l'indexation fait vibrer : on a le sentiment de faire revivre les êtres et les temps passés, d'entrer en familiarité avec eux. Il n'est pas rare que, à l'occasion de séances de paléographie, lorsque l'on rencontre un personnage ayant vécu au XVI^{e} ou au XVII^{e} siècle, l'un des paléographes dise au groupe : « Je le connais, celui-là ; c'était d'ailleurs un rude gaillard/un fieffé polisson/le père de 18 enfants/le reconstructeur de la ville voisine/un protestant acharné/un homme d'affaires rusé », etc. La dimension territoriale est capitale, non seulement pour la motivation des indexeurs, mais aussi pour leur compétence, grâce à leur connaissance des noms de lieux et de personnes (y compris dans leurs variantes anciennes) ; ce qui apporte aussi le plaisir d'un travail bien fait. Mais l'intensité du plaisir peut aussi être liée au domaine concerné par la nature des documents proposés à l'indexation : gastronomie (menus) ; Seconde Guerre mondiale (photographies de l'armée américaine en Normandie) ; toponymie (indexation des plans cadastraux) ; collections (cartes postales, chromolithographies) ; fascination des manuscrits précieux, etc.

Pour donner de la certitude au plaisir de l'indexation, il faut que le système de validation soit le plus réactif possible : immédiateté de la publication, rapidité de la correction, résolution rapide des bugs en cas de dysfonctionnement.

L'illustration la plus éloquente de la proximité du plaisir de l'indexation est naturellement l'*immediate replay* des GWAP présentés par Mathieu Lafourcade (voir son article) : dès que le joueur a donné sa réponse, il se voit proposer une autre séquence de jeu. Cette proximité diminue évidemment à mesure que se complexifie la besogne proposée à l'indexeur. Si l'on peut poser des tags sur l'état civil dans l'autobus ou le taxi, il est plus difficile d'y transcrire Bentham efficacement. Les tâches simples sont les plus addictives. On obtient alors l'« effet Pringles » : « *Once you pop you can't stop*[14] », mais sans risque d'indigestion ni d'obésité.

Le travail d'indexation est certes solitaire, mais il n'a d'intérêt que parce qu'une foule y contribue, et que tous peuvent profiter du travail de chacun. C'est le travail de la communauté qui donne au plaisir son étendue ; plus la foule est nombreuse et active, plus le travail avance, et plus augmente le plaisir de n'y pas contribuer de façon solitaire et ingrate.

Le travail produit par ces « mots croisés utiles » est naturellement cumulatif. Ce n'est pas comme une grille de sudoku ou de mots fléchés, dont l'achèvement apporte une satisfaction passagère et, éventuellement, contribue à augmenter la technicité de celui qui vient de la réussir : le travail d'indexation est utile pour soi (on a posé des signets que l'on peut aisément retrouver le lendemain), mais aussi pour les autres (puisque d'autres que moi peuvent retrouver grâce à mes propres index des archives perdues

14. Réflexion de Mia Ridge, exprimée par exemple sur sa page Web le 15 août 2015 dans ses conseils aux porteurs de projets muséaux en crowdsourcing (Guest post « Tips for digital participation, engagement and crowdsourcing in museums » for London Museums Group, `http://www.miaridge.com/guest-post-tips-for-digital-participation-engagement-and-crowdsourcing-in-museums-for-london-!museums-group/`, consulté le 28 juillet 2020).

Published in : Cécile Meynard et Thomas Lebarbé (dir.), *Le Crowdsourcing : Partager, enrichir et publier des sources patrimoniales*, Editions des archives contemporaines, ISBN : 9782813003850, doi : 10.17184/eac.9782813003850

dans l'océan du Web – et la réciproque est vraie). Retrouvant un jour, dans les registres paroissiaux du Cantal, un homme nommé « abjuration » à la fin du XVII^{e} siècle, je m'étais aperçu que l'indexeur avait posé des signets (qui ne prévoyaient que les deux champs « NOM » et « PRÉNOM ») sur les passages du registre des baptêmes où le curé avait consigné les abjurations consécutives à la révocation de l'édit de Nantes en 1685. Et nous avions décidé de conserver ces signets, considérant qu'il s'agissait d'une appropriation inattendue de l'outil mis à disposition. C'est ce caractère cumulatif et cette appropriation du patrimoine qui donnent sa fécondité au travail d'indexation. Cette idée de fécondité sous-tend le nom donné par les archives départementales des Yvelines, « Adoptez un Poilu », à leur projet d'indexation [15].

Enfin, le travail d'indexation apporte, pourvu que l'outil mis à disposition soit efficace, un plaisir sans mélange, la pureté dont Bentham fait le septième et dernier critère du plaisir dans son *Introduction aux principes de la morale et de la législation*. Il n'apporte pas de souffrance ultérieure. Mieux : le travail des autres démultiplie le plaisir de chacun. Et si le travail d'indexation peut parfois, comme on l'a vu, être addictif, il s'agit d'une addiction sans plus d'effet néfaste que tout autre travail prolongé devant un ordinateur. L'analyse de la temporalité des activités d'indexation montre un pic dans les mois froids, et une baisse à partir du printemps : signe que les archivonautes sont aussi adeptes, au temps chaud, du jardinage, du grand air et du patrimoine de plein vent.

Il y a, dans le grand escalier du palais des archives de la Côte-d'Or, le buste austère de l'un de mes prédécesseurs, demeuré en fonction entre 1832 et 1903. Ni sa figure ni celle de la momie de Bentham à Londres n'inspirent réellement l'idée de plaisir… Et pourtant, quel puissant moteur que le plaisir d'archives !

11 Quel statut juridique pour les résultats du *crowdsourcing* ?

Les simples tags, on l'a vu, ne sont grevés d'aucun droit d'auteur, non plus que les simples transcriptions. Le cas est différent pour les transcriptions scientifiques assorties d'annotations, ainsi, naturellement, que pour les traductions.

Tous ces travaux entrent en France, en revanche, dans le régime de la donnée publique, et donc de l'*open data*, telle que l'entend le *Code des relations du public avec l'administration*. Autrement dit, le fruit du travail collectif est susceptible de faire l'objet de réutilisation commerciale par une société privée, sans que le service public puisse arguer du mode de constitution (libre, désintéressé, démonétisé, communautaire, etc.) de cette information pour en refuser la fourniture pour réutilisation. Mais il faut informer en amont les indexeurs potentiels de cette éventualité, car les réticences du public à l'égard du *business* de la généalogie commerciale ne sont pas nulles, comme l'ont montré les débats, pétitions, controverses depuis 2009, moment de l'irruption du profit dans un monde jusqu'alors exclusivement gouverné par les lois de la recherche et du loisir désintéressés. Dans notre colloque, Manon Istasse et Vincent Jaillot ont souligné que les participants aux projets collaboratifs avaient parfois la peur d'être

15. `https://archives.yvelines.fr/article.php?laref=2594\&\%3Btitre=adoptez-un-poilu-`, consulté le 28 juillet 2020. Voir l'article d'Emmanuelle de Champs.

Published in : Cécile Meynard et Thomas Lebarbé (dir.), *Le Crowdsourcing : Partager, enrichir et publier des sources patrimoniales*, Editions des archives contemporaines, ISBN : 9782813003850, doi : 10.17184/eac.9782813003850

volés, surtout (mais pas exclusivement) lorsque la réutilisation est commerciale (voir leur article).

La réutilisation du travail peut néanmoins être considérée comme une chance, puisqu'elle démultiplie la visibilité du travail collectif et qu'elle l'agrège parfois, et parfois habilement, à d'autres jeux de données. Néanmoins, il me semble que le promoteur d'un projet de *crowdsourcing* a le devoir d'informer au préalable que, en vertu de la loi, le travail des internautes est susceptible de réutilisation, même commerciale, par n'importe qui.

12 Une République des Lettres désormais au cœur de l'écosystème numérique

Les jeux de mots utiles à l'édification de la République des Lettres prennent la place (et le temps) des jeux inutiles et des loisirs passifs. Chacun enrichit le patrimoine commun et contribue à un objectif partagé de connaissance. Les projets collaboratifs, participatifs et contributifs opèrent un va-et-vient entre le monde virtuel et le monde réel : le travail numérique a une réalité et les communautés virtuelles savent se réunir réellement à l'occasion.

Cela implique une redéfinition des métiers du patrimoine et du savoir. Si le rôle traditionnel d'expertise en dernier ressort qu'a traditionnellement le conservateur d'archives est attendu du public, sa posture change. Convaincu de la complémentarité du *crowdsourcing* avec les missions de service public (chercheurs, universitaires, experts, responsables de services patrimoniaux), il est à l'écoute des besoins et des goûts du public. Il organise et rend possibles les travaux collectifs, grâce à un outil technique à la fois pertinent et ergonomique. Tel un chef d'orchestre (formule de Yolaine Coutentin lors du colloque), il accepte de se départir de son instrument propre, de cesser d'être un virtuose solitaire, ou du moins de ne pas jouer de tous les instruments : il modère, encourage, reconnaît les mérites, corrige les erreurs, met à disposition les résultats du travail.

Utilisant les ressorts du *felicific calculus* de Bentham, il propose au public un loisir cultivé et désintéressé, *otium* qui s'oppose au *negotium* monétisé de la vie des affaires. Le numérique permet de démultiplier l'efficacité et la visibilité de la construction, par d'utiles citoyens antiquaires, d'un commun patrimoine littéraire, scientifique ou historique.

Published in : Cécile Meynard et Thomas Lebarbé (dir.), *Le Crowdsourcing : Partager, enrichir et publier des sources patrimoniales*, Editions des archives contemporaines, ISBN : 9782813003850, doi : 10.17184/eac.9782813003850