Des bénévoles au service du patrimoine écrit

De l'*Oxford English Dictionary* aux *Testaments de Poilus*

Emmanuelle de Champs
CY Cergy Paris Université

Résumé : Les définitions actuelles du « crowdsourcing » ne reflètent qu'imparfaitement la façon dont ce processus peut être appliqué en-dehors des circuits économiques et pour la création d'une valeur immatérielle et non monétaire : l'amélioration des connaissances et l'accès au patrimoine littéraire et historique. Cet article s'interroge sur la spécificité de la démarche collaborative en ligne dans les institutions patrimoniales. Il compare l'appel aux bénévoles anonymes lancé par les éditeurs du *Oxford English Dictionary* à la fin de l'époque victorienne aux initiatives actuelles, plus particulièrement le projet *Testaments de Poilus* mené depuis 2017 aux Archives nationales et départementales.

Mots-clés : Crowdsourcing, bénévoles, manuscrits, TEI, transcription

Abstract : Current definitions of *crowdsourcing* only imperfectly address the specificity of the cultural heritage sector, especially when it is used to create immaterial cultural value such as engagement with historical or literary archival material. This chapter compares and contrasts the collaborative approach used by the first editors of the *Oxford English Dictionary* in the late Victorian era with twenty-first century initiatives in cultural institutions, more specifically the "Testaments de Poilus" initiative carried out at three French archive institutions since 2017.

Keywords : Crowdsourcing, volunteers, manuscripts, TEI, transcription

Depuis la création du néologisme en 2006 dans le magazine *Wired*, plusieurs définitions du *crowdsourcing* ont été proposées. Les plus génériques embrassent un grand nombre d'usages, comme celle donnée par Lebraty en 2009 qui définit le *crowdsourcing* comme l'« externalisation par une organisation, *via* un site Web, d'une activité auprès d'un grand nombre d'individus dont l'identité est le plus souvent anonyme [1] ». L'étymologie du terme renvoie au fait de passer commande auprès d'un prestataire

1. Lebraty, Jean-Fabrice (2009), « Externalisation ouverte et pérennité. Une nouvelle étape dans la vie des organisations », *Revue française de gestion*, 2009/2 (n° 192), p. 151-165. Accessible sur `https://www.cairn.info/revue-francaise-de-gestion-2009-2-page-151.htm?contenu=resume`, consulté le 29 juillet 2020.

extérieur, comme le rappelle la racine « *sourcing* », substantif issu du vocabulaire du management, apparu dans la langue anglaise durant les années 1960 (*Oxford English Dictionary*). C'est là le parallèle que fait Jeff Howe dans l'article paru dans *Wired* en 2006 : « Dans des secteurs aussi divers que l'industrie pharmaceutique ou la télévision, des entreprises innovantes découvrent comment capter le talent latent de la foule. [...] Ce n'est pas de l'externalisation (*outsourcing*), c'est du *crowdsourcing*[2] ». Cet emploi est bien expliqué dans la définition récente donnée par l'*Oxford English Dictionary* (où le terme est entré en 2013) qui retrace la genèse du mot et replace la sollicitation au cœur de la démarche. Le nom *crowdsourcing* est défini comme « la pratique qui consiste à obtenir des informations ou des services en sollicitant un grand nombre de contributeurs, le plus souvent *via* Internet et sans offrir de compensation » (*Oxford English Dictionary*). L'emploi du terme de la langue anglaise est pertinent, car il met en lumière le fait de passer commande d'une tâche ou d'un contenu auprès d'un groupe donné ; il ne s'agit pas de capter une source qui jaillirait spontanément. Ces précisions permettent ainsi de préciser la démarche, contrairement à l'expression « production participative » qui ne dit pas grand-chose du rapport entre demandeurs et contributeurs, ceux qui assignent la tâche et ceux qui la remplissent.

Ces définitions ne reflètent pourtant qu'imparfaitement la façon dont ce processus peut être appliqué en dehors des circuits économiques et pour la création d'une valeur immatérielle et non monétaire : l'amélioration des connaissances et l'accès au patrimoine écrit. Comment comprendre le fait de s'adresser à des bénévoles, le plus souvent inconnus, au service du patrimoine écrit ? Après un éclairage historique sur la démarche mise en œuvre par le *Oxford English Dictionary* dans la seconde moitié du XIX^{e} siècle, cet article aborde la place du *crowdsourcing* appliqué au patrimoine écrit au sein des « sciences participatives ». Enfin, il revient sur des données recueillies dans le cadre du projet Testaments de Poilus (Fondation des Sciences du Patrimoine) depuis 2016[3].

1 L'histoire de l'*Oxford English Dictionary*, un premier exemple de *crowdsourcing* ?

L'histoire même de la préparation de l'*Oxford English Dictionary* est souvent présentée comme l'un des premiers exemples de *crowdsourcing* dans le domaine de la philologie et de la lexicographie[4]. La démarche mérite d'être rappelée en détail. En avril 1879, la Société philologique de Londres fait imprimer un prospectus intitulé *Appel au public anglophone désireux de lire des ouvrages pour extraire des citations destinées au nouveau dictionnaire de langue anglaise de la Société philologique*. Quatre pages denses

2. Howe, Jeff (2006) « The rise of crowdsourcing », *Wired*, juin 2006. Accessible sur `https://www.wired.com/2006/06/crowds/`, consulté le 29 juillet 2020.

3. Le projet « Testaments de Poilus » a obtenu des subventions de la Fondation des Sciences du Patrimoine en 2016, 2018 et 2020. Il associe deux laboratoires de l'Université de Cergy-Pontoise (Agora et Etis), les Archives nationales, les archives départementales des Yvelines (78), les archives départementales du Val d'Oise (95), l'École des Chartes et l'Université Paris 8. Des bénévoles participent à la collecte, à la transcription et à l'encodage TEI des testaments des Poilus de la première Guerre mondiale morts pour la France. Voir `https://testaments-de-poilus.huma-num.fr/\#!/`, consulté le 28 juillet 2020. Ce travail a bénéficié du soutien de l'École universitaire de recherche Paris Seine Humanités, Création, Patrimoine, Investissement d'Avenir ANR-17-EURE-0021 – Fondation des sciences du patrimoine.

4. Ridge, Mia (2014) « Introduction », *Crowdsourcing Our Cultural Heritage,* Ridge, Mia, éd., Oxford, Routledge, p. 1-17. Voir en particulier p. 5.

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

expliquent la genèse du projet : plus de vingt ans auparavant, les membres de la société ont lancé le chantier d'un nouveau dictionnaire appelé à remplacer celui de Johnson comme ouvrage de référence sur la langue anglaise, et dont l'ambition se veut tout à fait nouvelle. L'objectif est double : d'une part, il s'agit de viser l'exhaustivité en recensant *tous* les mots employés dans la langue anglaise ; d'autre part, de présenter l'histoire des termes et de leurs variations depuis 1250. Il s'agit essentiellement d'un dictionnaire philologique descriptif. Collecter les mots, recenser les moindres variations sémantiques et orthographiques, les illustrer à l'aide d'exemples tirés des textes d'auteurs variés, permettra de réaliser un panorama historique témoignant des évolutions de la langue pendant plus de sept siècles. Cela repose avant tout sur la collecte d'occurrences dans un énorme corpus, ce que les premiers éditeurs décrivent comme une tentative de « lancer un immense filet sur la mer de la littérature anglaise [5] ». Dans le prospectus, après avoir expliqué les ambitions du projet et décrit les pans du corpus qui restent à dépouiller, le texte se clôt sur un appel ciblé :

> La liste ci-dessous, préparée par M. Murray [président de la société philologique et directeur du projet] contient les titres dont il souhaiterait qu'on se charge immédiatement. La société fournira aux lecteurs des fiches portant la date d'impression, l'auteur et le titre de chaque ouvrage afin de leur épargner autant qu'il est possible un fastidieux travail. Dans l'espoir que cette collecte puisse être achevée si possible en moins de trois ans, *la société recherche mille lecteurs* et lance avec confiance cet appel pour que la préparation du dictionnaire débute avec des matériaux exhaustifs. La liste complète des sources et des lecteurs sera publiée à la suite du dictionnaire. L'appel est ouvert à tous, surtout pour les ouvrages modernes – c'est ainsi que les étudiants de M. Murray ont été en mesure de lui fournir en un mois plus de 5000 citations de qualité. [...] Les personnes qui souhaitent proposer leur aide doivent s'adresser à M. Murray, Mill Hill, Middlesex, N.W [6].

Cet appel est largement relayé dans la presse, au Royaume-Uni comme en Amérique du Nord [7].

Il s'agit en fait du second appel à volontaires, le premier ayant été lancé en 1857, toujours à l'initiative de la même Société philologique de Londres, sur proposition d'Herbert Coleridge. En 1857, il avait d'abord circulé parmi les membres de la Société philologique, qui avaient à leur tour sollicité des proches. Il avait été ouvert plus largement dans la presse en 1858, la méthodologie de la collecte ayant été améliorée après le retour des premières contributions : une fiche standardisée avait été imprimée et envoyée aux volontaires. En 1860, près de 90 personnes s'étaient mises au travail, dont 25 avaient terminé la collecte dans les volumes qui leur étaient assignés. Espérant recevoir 100 000 fiches, Coleridge voulait croire que ce travail serait achevé en

5. Gilliver, Peter (2016) *The Making of the Oxford English Dictionary*, Oxford, Oxford University Press, p. 19.

6. Murray, James H., 1879, « An appeal to the English-speaking and English-reading public », Londres, Philological Society. Accessible sur `http://public.oed.com/history-of-the-oed/archived-documents/april-1879-appeal/april-1879-appeal/`, consulté le 28 juillet 2020. Citation p. 4.

7. Gilliver, P., *The Making of the Oxford English Dictionary,* op. cit., p. 114.

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

deux ans. Mais malgré le succès rencontré initialement par l'appel de 1858, les contributions des volontaires s'étaient progressivement taries au cours des années suivantes. Mal conservées, les fiches s'étaient ensuite abîmées. Faute de financement, le nombre d'assistants s'était progressivement réduit et les contributions n'avaient été que peu utilisées (elles représentent environ 10 % du volume final des citations dans la version publiée). C'est James Murray qui, reprenant et perfectionnant la démarche en 1879, a permis d'atteindre et de dépasser les objectifs fixés lors du premier appel. En 1882, plus de 900 000 citations ont déjà été reçues (*ibid.*, p. 117). Quarante ans plus tard, le premier volume de l'*Oxford English Dictionary* paru en 1887 fait usage de six millions de fiches envoyées par des bénévoles. Le dernier volume de la première édition est achevé peu après la Première Guerre mondiale.

L'histoire de l'*Oxford English Dictionary* en fait une entreprise inégalée dans le monde des humanités et, à bien des égards, un modèle. En effet, on retrouve dans la démarche d'Herbert Coleridge les éléments qui sont encore aujourd'hui à l'origine du succès des projets de *crowdsourcing* dans le domaine des humanités numériques :

1. Une masse de données à traiter qui requiert un travail considérable et justifie l'appel à un public ouvert ;
2. Une mission précise, identique pour tous les contributeurs, encadrée par un protocole normé et soutenu par une infrastructure logistique solide. Au plus fort du projet, la poste d'Oxford fera construire une boîte aux lettres spéciale pour recueillir le courrier adressé à Murray, en raison du volume des livraisons quotidiennes ;
3. Un travail qui mobilise les compétences, les intérêts et les ressources propres des volontaires ;
4. Des intermédiaires chargés d'établir la jonction entre les responsables du projet et les contributeurs volontaires. L'*Oxford English Dictionary* fait ainsi appel à des « éditeurs délégués » (*sub-editors*), bénévoles eux aussi, chargés de lire et d'organiser les citations pour une lettre donnée ;
5. Un financement solide : il n'est pas étranger au succès de l'entreprise que l'arrivée de James Murray en tant que responsable éditorial coïncide avec la signature d'un contrat avec Oxford University Press, garantissant ainsi la viabilité financière du projet et sa pérennisation.

Le succès de Murray, outre ses qualités de lexicographe et son investissement sans pareil dans le projet, tient notamment à ce qu'il prend très au sérieux le travail des volontaires. Entre avril 1879 et janvier 1880, il fera imprimer trois versions de l'appel, précisant à chaque fois les attentes, recensant les erreurs commises et les questions qu'on lui pose souvent. Par exemple : faut-il envoyer toutes les citations qui illustrent l'usage d'un mot fréquent, comme l'article « *the* » ? Ou bien se concentrer uniquement sur les mots rares ? Faut-il procéder méthodiquement en extrayant *tous* les mots d'un ouvrage donné, ou bien glaner ici et là (Gilliver, P., *ibid.*, p. 116) ? Soucieux de donner aux volontaires une idée de ce qu'il attend, Murray fait imprimer des spécimens de fiches. Il entre personnellement en correspondance avec un certain nombre

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

de volontaires, souvent pour discuter des points d'étymologie ou des hypothèses lexicographiques (*ibid.*, p. 128). Au cours des années qui suivent, il fait paraître dans la presse des appels plus spécifiques, demandant par exemple de l'aide pour établir à quelle date le mot *tulipe* apparaît dans un texte de langue anglaise (*ibid.*, p. 128).

L'étude du rapport entre le nombre de contributions et celui des contributeurs au cours des premières années du projet est instructive. Selon le *Times* du 31 janvier 1882, le premier appel d'avril 1876 n'avait suscité que 176 réponses. L'année suivante, suite aux trois réimpressions de l'appel, le nombre monte à 754. Le journaliste ajoute : « Pour illustrer l'enthousiasme et le travail de certains des "lecteurs", on peut signaler que sur les 361 670 fiches soumises à l'éditeur, 11 000 provenaient d'un seul "lecteur", 10 350 d'un autre, et 19 200 d'un troisième [8] ». Dans la préface du premier fascicule, parue en 1888, James Murray évaluera le nombre de citations recueillies à 3,5 millions et le nombre de contributeurs à 1300 [9]. Ce constat permet de mettre en perspective les résultats de projets contemporains de *crowdsourcing* en ligne appliqué aux corpus littéraires : par exemple, le Bentham Project qui fait transcrire par des bénévoles les manuscrits du philosophe Jeremy Bentham *via* la plateforme « Transcribe Bentham [10] » voit l'essentiel des transcriptions effectuées par un nombre très réduit de *supertranscribers* [11]. Le même article du *Times* rapporte un autre constat surprenant : si de nombreux contributeurs sont bien des universitaires, comme escompté, une partie non négligeable de ce contingent est composée d'Américains et non de Britanniques. Parmi les Britanniques, beaucoup ne sont pas universitaires. L'article aurait pu ajouter qu'un certain nombre d'entre eux sont des femmes (dans une proportion hélas difficile à établir car la préface du premier fascicule qui contient les noms des contributeurs ne fournit que leur initiale).

Comme en témoigne cet article du *Times*, Murray mobilise volontiers la presse écrite de l'époque qui relaie périodiquement ses appels au début des années 1880. Il fait connaître l'équipe du projet, faisant réaliser des séries de photos sur le travail du « Scriptorium », le bureau d'Oxford où sont reçues et triées les fiches, et y recevant également les visiteurs [12]. En 1882, Murray profite de la tribune que lui offre un journaliste pour

> rappeler à ses contributeurs que certaines des contributions les plus valables sont celles qui concernent les fiches occasionnelles sur des mots rares. Il exprime le souhait que personne ne se laisse aller à la paresse de compter sur un autre lecteur pour les relever si lui-même les néglige. Il rappelle qu'il ne peut lui-même lire un éditorial dans un quotidien sans repérer un mot qui mérite d'être extrait, et il ne se souvient pas avoir déjà reçu des doublons de ces termes-là. (*The Times,* 1882)

8. *The Times* (1882) « The Philological Society's Dictionary », 31 janvier 1882.

9. Murray, J. (1888) « Preface to Volume I A and B », *Oxford English Dictionary, vol. 1, A-Ant*, Oxford, Oxford University Press, p. .

10. `http://transcribe-bentham.ucl.ac.uk/`, consulté le 25 novembre 2019.

11. Causer, Tim, et Terras, Melissa (2014) « "Many hands make light work. Many hands together make merry work". Transcribe Bentham and crowdsourcing manuscript collections », *Crowdsourcing our Cultural Heritage*, Ridge, Mia, éd., Oxford, Routledge, p. 57-88. Citation p. 57.

12. Gilliver, P., *The Making of the Oxford English Dictionary, op. cit.,* p. 131.

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 contrôle des fiches reçues est un point essentiel. Lorsqu'il reprend le projet, Murray se plaint de la qualité parfois très médiocre des citations dont il a hérité, se lamentant en 1899 qu'il « aurait mieux valu tout brûler et tout recommencer » (Gilliver, *op. cit.*, p. 261). Analysant les citations qu'il reçoit, il remarque que les contributeurs les plus prolifiques sont également ceux qui donnent à l'équipe le moins de travail de relecture et de correction (*ibid.*, p. 131). La relecture est effectuée sous sa direction par son équipe. Il s'agit surtout de vérifier que la citation choisie est complète et répond aux normes du projet : c'est là que sont écartées les citations illisibles ou incomplètes. Lorsque c'est possible, ils procèdent à des vérifications dans les originaux (notamment à partir du moment où Murray demande aux volontaires de *ne pas* découper l'original pour le coller sur la fiche – surtout lorsqu'il s'agit de livres rares ou anciens). Il faut également évaluer le pouvoir illustratif de chaque citation. Enfin, malgré l'ampleur de la collecte, certains sens attestés sont toujours en attente de citations correspondantes – Murray commande donc des lectures ciblées à la Bodleian Library, au British Museum, ou encore dans des bibliothèques privées.

Murray met ainsi en place une méthode rigoureuse pour le tri des citations (y employant chacun de ses onze enfants dès qu'il a appris à lire, contre une petite rémunération). Les éditeurs délégués organisent les citations selon les différents sens de chaque terme, avant de laisser la décision finale aux éditeurs professionnels (*ibid.*, p. 263). C'est à ce niveau de compétence que Murray regrette l'appel à des bénévoles : sur une quarantaine d'éditeurs délégués, déplore-t-il, seuls huit ont permis à l'équipe de gagner du temps. En fait, l'essentiel du travail a dû être refait par ses collaborateurs (*ibid.*, p. 264).

Aujourd'hui encore, l'*Oxford English Dictionary* continue à faire appel à des contributeurs bénévoles, d'abord par le biais de son site Internet où les internautes sont invités à signaler des occurrences antérieures à celles qui figurent dans le dictionnaire quand ils en trouvent [13], puis en partenariat avec d'autres projets de *crowdsourcing*, comme Shakespeare's world, hébergé par Zooniverse, qui propose aux internautes de transcrire des textes anglais du XVI^{e} siècle afin d'alimenter la collecte des premières occurrences ou des variantes orthographiques [14].

2 Le modèle des sciences participatives

L'histoire de l'*Oxford English Dictionary* permet d'inscrire le *crowdsourcing* dans la longue durée des pratiques des sciences participatives, tout en mettant en lumière la place occupée par les humanités, au sens large. Ce point est insuffisamment rappelé lorsqu'on évoque les projets anciens où les sciences naturelles et physiques sont principalement mises en avant. Tout comme en sciences, les académies lancent dès le XVIII^{e} siècle des concours pour résoudre des questions philosophiques, littéraires ou historiques [15]. Il existe aussi des amateurs qualifiés et passionnés dans les sciences humaines et sociales ; pourtant, seuls 3 % des projets de sciences participatives se

13. `https://public.oed.com/appeals/`, consulté le 28 juillet 2020.

14. `https://www.shakespearesworld.org` consulté le 28 juillet 2020.

15. Caradonna, Jeremy L. (2009) « Prendre part au siècle des Lumières. Le concours académique et la culture intellectuelle au XVIII^{e} siècle », *Annales. Histoire, Sciences Sociales* 64 (3), p. 633-662.

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

situent aujourd'hui dans le domaine des « arts et sciences humaines » – sciences de l'éducation et sciences sociales non comprises [16].

Dans le cadre de l'étude menée à la demande des ministres en charge de l'Éducation nationale et de l'Enseignement supérieur et de la recherche, François Houiller définit les sciences participatives comme « une forme de production de connaissances scientifiques auxquelles des acteurs non-scientifiques-professionnels – qu'il s'agisse d'individus ou de groupes – participent de façon active et délibérée [17] ». Le rapport esquisse plusieurs typologies qui permettent de replacer le *crowdsourcing* au sein d'un éventail de pratiques collaboratives appliquées à la recherche scientifique. Il insiste sur la nécessité de dépasser une vision du *crowdsourcing* qui se résumerait à la sous-traitance de la collecte d'informations et assimilerait les bénévoles à des « capteurs de données », selon une typologie établie par Muki Haklay en 2015 (*ibid.*, p. 14). Ainsi, le rapport de François Houiller fait du *crowdsourcing* l'une des modalités à part entière des « sciences citoyennes » : en faisant participer les bénévoles non seulement à la collecte, mais aussi à l'analyse des données, les institutions sont en mesure de « produire des connaissances et des indicateurs », mais aussi d'« éduquer les citoyens aux méthodes scientifiques » (*ibid.*, p. 14). Si les sciences humaines et sociales ne sont hélas pas citées au nombre des domaines prioritaires pour la mise en œuvre de cette démarche, les contributions rassemblées dans ce volume reconnaissent largement la place qu'elles doivent y occuper.

Le rapport invite ainsi à ne pas distinguer ni opposer les deux missions remplies simultanément par la démarche des sciences participatives : collecter et éduquer. Coconstruction du savoir avec des bénévoles et médiation en direction d'un public large sont des aspects que les projets de *crowdsourcing* dans les humanités numériques se doivent de prendre au sérieux, qu'ils émanent d'institutions de conservation du patrimoine ou bien d'équipes de recherche universitaires. C'est également en articulant ces deux objectifs qu'on pourra mieux définir les enjeux éthiques que soulève nécessairement l'appel à des bénévoles. Ces projets mobilisent effectivement des personnes pour un travail sans leur offrir de contrepartie financière, mais être au clair avec la démarche participative et ses implications en termes de médiation doit permettre de tenir à distance les accusations de travail dissimulé tout en appréhendant sans naïveté les enjeux éthiques et économiques [18].

3 La motivation des contributeurs. Étude préalable au projet *Testaments de Poilus*

Le rapport Houiller présente également la synthèse d'une enquête sur les publics des projets de sciences participatives, remarquant que « chaque dispositif participatif français regroupe entre 3 et 60 000 personnes selon les témoignages recueillis ». Le « public » auquel s'adressent ces projets est distingué en quatre groupes : amateurs

16. Houiller, François (2016) « Les sciences participatives en France. État des lieux, bonnes pratiques & recommandations ». Accessible sur `http://www.sciences-participatives.com/Rapport`, consulté le 28 juillet. Citation p. 22

17. Houiller, François (2016) « Les sciences participatives en France », art. cité. Citation p. 6.

18. Bouquillion, Philippe, et Matthews, Jacob T. (2010) *Le Web collaboratif. Mutation des industries de la culture et de la communication*, Grenoble, Presses Universitaires de Grenoble.

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 passionnés ; personnes directement concernées par les résultats de l'étude (patients, habitants...) ; élèves et familles ; professionnels (toutes les personnes qui s'intéressent à titre professionnel à la ressource ou au projet). Cette typologie s'appuie sur les résultats d'une enquête auprès de contributeurs bénévoles (Houiller, art. cité, p. 22-23). La synthèse des réponses révèle la prépondérance de l'intérêt et de la curiosité dans les motivations des bénévoles – il ne faut pas négliger non plus la catégorie un peu floue des « professionnels » qui s'intéressent à la ressource ou au résultat de l'analyse. Le rapport pointe aussi le risque de lassitude ou d'essoufflement des bénévoles, surtout quand les sollicitations sont nombreuses. D'où la nécessité pour les porteurs de projets d'entretenir et de nourrir leur curiosité et leur motivation en communiquant régulièrement sur les objectifs et sur les résultats.

Dans le cadre du projet Testaments de Poilus lancé en 2017, nous avons réalisé en juillet 2017 une série d'entretiens auprès de sept bénévoles qui participent depuis plusieurs années aux projets collaboratifs lancés par les archives départementales des Yvelines (partenaires du projet). Trois de ces projets menés par les AD 78 portent particulièrement sur la guerre de 1914-1918 :

1. L'indexation collaborative des fiches matricules des soldats de la Grande Guerre (Adoptez un Poilu [19]) ;
2. Le dépouillement des archives notariales pour y découvrir les testaments des Poilus morts pour la France [20] ;
3. La contribution au Wiki de la Grande Guerre, où des bénévoles proposent et rédigent des articles et les mettent en forme pour la publication en ligne [21].

Malgré le caractère restreint de cet échantillon, le principe d'entretiens individuels approfondis (qui ont souvent duré plus d'une heure) a permis de mieux connaître ces bénévoles et leurs motivations. Confirmant les constats récents [22], notre enquête a permis de constater que les bénévoles sont presque exclusivement de jeunes retraitées (un homme et une personne en activité professionnelle, sur notre échantillon). En grande majorité, elles ont découvert les archives départementales à l'occasion de recherches généalogiques et ont répondu à des invitations qui leur permettaient d'aller plus loin, une fois leurs recherches familiales en bonne voie. Plusieurs font partie de cercles de généalogistes et d'associations d'histoire locale. Ce lien avec le secteur associatif met en lumière deux aspects importants. D'une part, l'engagement associatif est fondé sur l'entraide, une qualité particulièrement sensible dans les associations de généalogistes. Plusieurs personnes nous ont décrit la vitalité des échanges en ligne ou bien lors de réunions, où les membres consacrent du temps à partager leur connaissance des fonds avec d'autres. D'autre part, la sociabilité associée aux différents projets est

19. `https://archives.yvelines.fr/article.php?laref=2585\&titre=adoptez-un-poilu-`, consulté le 28 juillet 2020.

20. `https://archives.yvelines.fr/article.php?laref=1574\&titre=testaments-de-poilus-saison-2`, consulté le 28 juillet 2020.

21. `https://archives.yvelines.fr/article.php?laref=848\&titre=le-wiki-de-la-grande-guerre`, consulté le 28 juillet.

22. Causer, Tim, et Wallace, Valerie (2012) « Building a volunteer community : results and findings from Transcribe Bentham », *Digital Humanities Quarterly* 6 (2). Accessible sur `http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html`, consulté le 28 juillet 2020. Andro, Mathieu (2017) *Bibliothèques numériques et crowdsourcing*, Paris, ISTE.

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

pour beaucoup d'entre eux l'un des attraits de l'engagement bénévole, même si le travail se fait en ligne depuis chez soi.

Connaître la relation qu'entretiennent les bénévoles à l'outil informatique est crucial pour des projets qui reposent sur la collaboration en ligne. Parmi celles que nous avons interrogées, toutes sauf une utilisent aisément l'informatique au quotidien. Plusieurs se sont formées dans le cadre professionnel à la fin de leur carrière. Elles considèrent Internet et les ordinateurs comme des moyens et se décrivent comme peu compétentes, bien qu'elles s'en servent tous les jours et sous-estiment sans doute leurs compétences (boîte email, navigateurs, Word, Excel, plus rarement logiciels spécialisés comme heredis [23]). Ces personnes travaillent principalement chez elles à partir d'ordinateurs portables ou fixes (peu de smartphones ou de tablettes). Elles ont une grande défiance vis-à-vis des réseaux sociaux, s'informent principalement par email (listes de diffusion) ou bien en se rendant sur les sites des organismes qui les intéressent (en ligne et sur place). Pourtant, malgré le peu de confiance qu'elles semblent manifester dans leurs capacités numériques, ces personnes se sont révélées, au cours des projets collaboratifs en ligne menés par les archives départementales des Yvelines, capables d'effectuer des tâches qui réclament plus de compétences que celles qui sont sollicitées dans un usage basique et quotidien d'Internet : l'indexation réclame une bonne maîtrise des champs et des outils de visualisation des images, et la contribution au wiki implique de maîtriser les premiers rudiments du balisage xml. Sur notre échantillon, trois s'étaient mises sans difficulté au balisage simplifié à l'usage des wiki, en appréciant le fait de pouvoir en cas de besoin solliciter sur place les responsables du projet aux archives départementales.

Nous avons également interrogé les bénévoles plus précisément sur la spécificité du projet Testaments de Poilus, qui leur propose non seulement de transcrire les testaments des Poilus morts pour la France, mais aussi de réaliser le balisage en xml/TEI. Aller au-delà de la transcription et mobiliser les bénévoles pour le balisage est l'un des défis du projet. C'est une étape cruciale qui permettra de réaliser une édition électronique qui soit de bout en bout collaborative. Lors des entretiens aux archives départementales des Yvelines, nous leur avons présenté le principe de la plate-forme (il était alors trop tôt pour soumettre un prototype), en leur demandant s'il leur paraissait possible de réaliser l'encodage en plus de la transcription. La réponse a été positive dans plus de la moitié des cas. Pour tous ces bénévoles, la motivation n'est pas prioritairement l'apprentissage d'une nouvelle compétence, mais surtout la volonté d'apporter une contribution exploitable par l'équipe du projet et dont ils comprennent la valeur ajoutée.

Les entretiens ont illustré à la fois l'importance de la passion ou du moins de l'intérêt pour le sujet (en ce qui concerne les projets des archives départementales des Yvelines, les bénévoles qui contribuent le plus à l'indexation collaborative sont également ceux qui sont prêts à rédiger des articles pour le wiki, où ils sont totalement libres de choisir un sujet et de le traiter comme ils l'entendent). Ils ont aussi montré l'attachement qu'ont des bénévoles à une institution avec laquelle ils ont noué des liens particuliers. Presque tous avaient fréquenté les salles de lecture des archives départementales des

23. `https://www.heredis.com/`, 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

Yvelines avant de s'investir dans les projets collaboratifs en ligne. Ils avaient alors tissé des relations de confiance avec plusieurs personnes de l'institution. L'institution était donc présente dans leurs discours à la fois en tant que commanditaire de la tâche et que point de contact en cas de problème. Ce n'est pas, loin s'en faut, une caractéristique essentielle du *crowdsourcing* : beaucoup de projets ont « lancé un grand filet sur la mer des internautes » – pour reprendre une image déjà utilisée – sans s'appuyer sur des relations préexistantes avec les bénévoles (voire partent de la défiance à l'égard des institutions, comme le projet de transcription collaborative des notes de frais des députés à la Chambre des communes britannique lancé par le *Guardian* en 2009 [24]). Pourtant, il semble important de ne pas négliger cet aspect dans le montage de projets de *crowdsourcing* pour l'exploitation et la diffusion de sources patrimoniales. Aux institutions et aux projets d'aller à la rencontre de leurs bénévoles – c'est d'ailleurs une démarche qui est de plus en plus de répandue, par exemple en organisant des ateliers, des conférences liées au projet, des rencontres. Les projets dont le corpus permet de s'appuyer sur un tissu associatif vivant ont là un vivier de contributeurs à ne pas négliger. Il existe donc un niveau intermédiaire entre la foule et l'individu qui est précisément celui où beaucoup de projets peuvent se situer.

Les apports de ces entretiens ont été directement mobilisés au moment de la conception de la plate-forme des Testaments de Poilus lancée le 30 janvier 2018. L'importance de l'encodage en xml/TEI est mise en avant et expliquée dès la page d'accueil [25]. Il était en effet indispensable pour la réussite du projet que les bénévoles aient conscience de la différence de qualité entre une transcription simple et une transcription balisée. En outre, le tutoriel réalisé par Florence Clavaud propose une démarche rigoureuse de transcription et de balisage qui permet, en cinq étapes, de passer d'une transcription au kilomètre à une transcription balisée et enrichie [26]. Sans cacher les aspects techniques (le code xml/TEI est visible à toutes les étapes et peut même être saisi directement par les transcripteurs), le projet les met en avant et en explique la logique intellectuelle et scientifique. Lire pour transcrire en ligne et encoder, c'est manipuler un document (à la fois matériel et son image numérique), mais aussi mieux le comprendre et l'enrichir. Quelques semaines après la mise en ligne de la plate-forme (un bilan complet a été réalisé à l'été 2018), la capacité des transcripteurs à s'approprier un système de balisage assez complexe est tout à fait satisfaisante. Le lien avec les institutions (les archives nationales et les archives départementales des Yvelines) a été marqué lors de la demi-journée de lancement qui s'est tenue sur le site de Paris des Archives nationales. Les ateliers ont permis aux bénévoles de tester le tutoriel et à l'équipe d'avoir un premier retour sur le dispositif, permettant ainsi des corrections immédiates. Aller à la rencontre des bénévoles en préalable à la conception de la plate-forme nous a ainsi permis de mieux cibler le public potentiel et ses attentes.

24. `https://www.theguardian.com/gnm-press-office/crowdsourcing-mps-expenses`, consulté le 28 juillet 2020.

25. Florence Clavaud, « Qu'est-ce qu'une édition numérique ? » et « Qu'est-ce que TEI ? », `https://testaments-de-poilus.huma-num.fr/\#!/content/17`, consulté le 28 juillet 2020.

26. `https://testaments-de-poilus.huma-num.fr/\#!/training/11/exercise/presentation`, consulté le 28 juillet 2020. Le tutoriel n'est accessible qu'aux utilisateurs ayant créé un compte sur la plateforme.

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

4 Conclusion

La première édition de l'*Oxford English Dictionary* a fait l'admiration des lexicographes du monde entier, mais le travail des volontaires n'a pas particulièrement retenu l'attention de la communauté scientifique de l'époque. En 1884, présentant le premier volume édité par Murray (*A-ANT*) à l'académie des Inscriptions et Belles-Lettres, le philosophe Paul Meyer (alors directeur de l'École des Chartes) soulignait surtout l'ampleur du travail éditorial ainsi que les contributions des érudits les plus notables – tous membres de la Société philologique de Londres, oubliant le grand nombre d'inconnus, certains sans aucun autre titre que leur passion pour les mots, qui avaient contribué à la réussite du projet [27].

Pourtant, l'épopée de la mise au travail de centaines de contributeurs anonymes a séduit l'imagination des contemporains. En 1915, un journaliste américain publiait l'anecdote suivante : intrigué par le nombre de contributions envoyées par un certain D^r^ Minor et souhaitant le remercier pour la qualité et l'originalité de ses fiches, James Murray avait émis en 1897 le vœu de le rencontrer. Contacté par courrier, le D^r^ Minor avait accepté avec enthousiasme. Après un court trajet en train, puis en calèche, Murray se rendit compte que l'on approchait d'un hôpital. En arrivant, il se présenta au directeur : « Monsieur, je présume que j'ai l'honneur de saluer le lexicographe qui nous envoie depuis tant d'années de si précieuses contributions ». « Vous faites erreur, répondit le directeur, le D^r^ Minor est interné dans l'aile psychiatrique de l'établissement » (l'anecdote fournit le point de départ d'un roman [28]). Pourtant, après avoir rencontré le D^r^ Minor, James Murray se lia d'amitié avec ce lexicographe amateur américain, par ailleurs victime de crises de paranoïa qui avaient conduit à son hospitalisation forcée.

Si ce récit de leur rencontre est hélas apocryphe [29], il n'en n'est pas moins exact que le D^r^ Minor, patient psychotique de l'asile de Broadmoor, a rédigé entre 5 000 et 8 000 fiches pour l'*Oxford English Dictionary*. Disposant de temps et d'une fortune familiale, il avait les moyens de se procurer les ouvrages anciens qu'il compulsait pour remplir ses fiches, au point que l'administration de l'hôpital avait mis une chambre supplémentaire à sa disposition pour héberger sa bibliothèque. Si l'histoire de la conception de l'*Oxford English Dictionary* peut être considérée comme le récit des origines du *crowdsourcing* en sciences humaines, elle fournit également un bon exemple des fantasmes auxquels l'appel aux volontaires peut donner lieu, ainsi que, espérons-le, un bon antidote.

27. Meyer, Paul, 1884, « Dictionnaire historique de la langue anglaise par M. James Murray », *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres*, 28 (1), p. 119-122. Accessible sur `https://doi.org/10.3406/crai.1884.68966`, consulté le 28 juillet 2020.

28. Voir Winchester, Simon (1998) *The Surgeon of Crowthorne : a tale of murder, madness and the love of words*, Londres, Viking.

29. Pour une étude historiquement documentée des relations entre Murray et Minor, voir Stevens, Mark (2013) *Broadmoor Revealed : Victorian Crime and the Lunatic Asylum*, Barnsley, Pen & Sword.

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